



Proceedings

August 25 - 27, 2025

**Churchill College
University of Cambridge
Cambridge, UK**

Researchers whose paper appears in these proceedings are allowed to translate, copy or upload only their own contribution to any social networking service or database.

WCE-2025 ISBN: 978-1-913572-80-8
i-Society 2025 ISBN: 978-1-913572-81-5

2025 WCE and i-Society

All rights reserved. This work may not be translated, copied or uploaded to any social networking service or database in whole without the written permission of the publisher. The use of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary right.

Copyright © WCE and i-Society 2025

WCE | i-Society

August 25 - 27, 2025

Churchill College
University of Cambridge
Storey's Way
Cambridge
CB3 0DS
United Kingdom

Message from the Steering Committee Chair

Welcome to the WCE and i-Society 2025. The 12th World Congress on Education (WCE- 2025) and 18th International Conference on Information Society (i-Society 2025) are co-located conferences.

The WCE and i-Society 2025 initially received 513 papers from 82 countries of which 102 papers were accepted after the first review. And 26 full papers were finally accepted with 5 abstracts, 4 extended abstracts, 1 workshop, 15 speakers' proposals, 1 session speaker's proposal and 3 posters from 26 countries. The double-blind paper review method was adopted to evaluate each of the conference's submissions and their specific research contribution to the knowledge domain within the subject matter. Please note that all authors will be invited to extend their papers for journal publications and to contribute to book chapters.

Many people have worked extremely hard to make WCE and i-Society 2025 possible. We would like to thank all, who have helped in making WCE and i-Society 2025 a success. The Steering Committee and reviewers each deserve credit for their excellent job. We would also like to acknowledge our appreciation to the following organisations for their sponsorship and support: Inclusive Education and Lifelong Learning, Mental Health Awareness, Young Researchers ThinkTank, Desing Thinkers, Music Creativity, Inclusiveness, and Infonomics Society. The long-term goal of WCE and i-Society is to build a reputation and respectable conferences for the international community.

On behalf of the WCE and i-Society 2025 Executive members, we would like to thank all the authors and participants who have contributed to both conferences. Your future participation in the WCE and i-Society as authors, speakers, panellists, workshop organisers and volunteers will be well-appreciated.

We wish you a pleasant stay in Cambridge, UK.

Professor Charles A. Shoniregun
WCE and i-Society 2025 General Chair

Table of Contents

Message from the Steering Committee Chair	3
Table of Contents	4
WCE-2025 Executive Committees	11
i-Society 2025 Executive Committees	12
Speakers	14
Speaker 1	15
Title: Future of Cyber Resilience	17
Speaker: Thiago Vieira	
DOI: 10.20533/WCE.2025.0001	
Speaker 2	18
Title: The Importance of Building Community	20
Speaker: Deketa Tranae	
DOI: 10.20533/WCE.2025.0002	
Speaker 3	21
Title: Bridging the Knowledge-Action Gap in Education for Sustainable Development: A Student-Led Partnership Model from Hong Kong	23
Authors: Janet KY Chan, Peter FM Lau, TY Hui, Zoe Liao, Sunny Chiu	
Speaker: Janet KY Chan	
DOI: 10.20533/WCE.2025.0003	
Speaker 4	28
Title: Generative AI to Assist Research Journaling in the Academe	30
Speaker: Lyberius Ennio Taruc	
DOI: 10.20533/iSociety.2025.0001	
Speaker 5	31
Title: The Collaborative Instructional Design System (New-Gen): An Ai-Assisted Multidimensional Aspect of Instructional Design Towards VUCA World	33
Speaker: Ismail Zain	
DOI: 10.20533/WCE.2025.0004	
Speaker 6	34
Title: The Impact of the Innovative Differentiated Didactic Approach to Teaching the Arts (D.D.A.T.A.) on Enhancing Skills in Students with Intellectual Disabilities	36
Speaker: Ioannis Makris	
DOI: 10.20533/WCE.2025.0005	

Speaker 7	43
Title: A Future of Hope: Ensuring Educator Retention	45
Speaker: Alysha Collins	
DOI: 10.20533/WCE.2025.0006	
Speaker 8	46
Title: Empowering Futures: Cultivating Lifelong Learning in Arab Youth	48
Speaker: David Lazaro	
DOI: 10.20533/WCE.2025.0007	
Speaker 9	49
Title: Bridging the Gap: Elevating Higher Education Assessment through Professional Standards	51
Speaker: Samar El-Farra	
DOI: 10.20533/WCE.2025.0008	
Speaker 10	52
Title: Modeling Multidimensional Competence: An LCDM-Based Evaluation of 21st-Century Skills	54
Speaker: Vamsidhar Yerramsetty	
DOI: 10.20533/WCE.2025.0009	
Speaker 11	55
Title: Enhancing Educational Insights and Engagement through Learning Analytics: Exploring Digital Reading Packages	56
Speaker: Judith Hanke	
DOI: 10.20533/WCE.2025.0010	
Speaker 12	58
Title: Promoting Scientific and Environmental Literacy through Non-Formal Education: The Case of the CienTec Park at the University of São Paulo	60
Speaker: Suzana Ursi	
DOI: 10.20533/WCE.2025.0011	
Speaker 13	65
Title: The State of Foundational Learning in Ghana: Gender Equality, Quality of Teaching, Accessibility, and Relevance	67
Speaker: Festus Buetey Longmatey	
DOI: 10.20533/WCE.2025.0012	
Workshop	68
Title: Brave and Bold: Fostering Psychological Safety	70
Organiser: Shari Collins	
DOI: 10.20533/WCE.2025.0013	

Sessions	71
Session 1: Session 1: Education and Digital Society	72
Title: Development of an Interactive Mobile Application to Support Environmental Education for Young Children	73
Authors: Yeseo Kim, Hae-Deok Song, Ah-Jeong Hong	
DOI: 10.20533/WCE.2025.0014	
Title: Applying Active Learning to a Seminar on the Basics of Quantum Computing	74
Authors: Pedro J. Roig, Salvador Alcaraz, Katja Gilly, Cristina Bernad, Carlos Juiz	
DOI: 10.20533/iSociety.2025.0002	
Title: Minimalist Approach: Analysis of “Enigmatic Ruins” Video Game Developed by Game Design Students	80
Authors: Oluwarotimi Abayomi Randle, David Ajayi	
DOI: 10.20533/iSociety.2025.0003	
Title: Enhancing Project-Based Learning in Engineering Education with Expert-Guided Retrieval-Augmented Generation (RAG)-Based Chatbot	85
Authors: Y.W. Kiang, A.T.L. Lee, V.W.L. Tam, K.H. Lam, V.C.S. Lee, S.M. Yiu	
DOI: 10.20533/WCE.2025.0015	
Session 2: Learning and Teaching Methodologies	87
Title: Saudi Secondary Education Tracks: Enhancing Educational Quality and Alignment with International Practices	88
Author: Khalid Almalhy	
DOI: 10.20533/WCE.2025.0016	
Title: Entagogy: Developing a Systems-Theoretical Framework for Autopoietic Co-Construction Between Learners and AI in Posthumanist Education	89
Authors: Alexander Harris, Stephanie Holt	
DOI: 10.20533/WCE.2025.0017	
Title: Using Psychological factors to influence STEM education	99
Authors: Richard Kyere Asomah, Francis Kwesi Nsakwa Gabriel-Wetley	
DOI: 10.20533/WCE.2025.0018	
Title: Teaching to Learning – A Paradigm Shift	100
Author: Amlan Datta	
DOI: 10.20533/WCE.2025.0019	
Title: Development of Mobile Language Learning Applications Using Interactive Multimedia and Gamification for Early Childhood Education	101
Authors: Yu-Kyung Kim, Yeseo Kim, Hae-Deok Song, Ah-Jeong Hong	
DOI: 10.20533/WCE.2025.0020	
Session 3: Pedagogy and AI	102
Title: The Compulsion to Create in Artists’ Narratives: A Psycho-Pedagogical Perspective	103
Author: Agnieszka Janiszewska	
DOI: 10.20533/WCE.2025.0021	

Title: Engaging Young Children with Mathematics – Is Agency a Requirement? Authors: Audrey Cooke, Maria Johansson, Eva Norén DOI: 10.20533/WCE.2025.0022	104
Title: AI's Impact on Higher Education: Transforming Research, Teaching, and Learning Author: Alyse Jordan DOI: 10.20533/WCE.2025.0023	111
Title: AI Digital Textbooks in Korean Classrooms: The Codle Platform in Action Author: Justin Eom DOI: 10.20533/WCE.2025.0024	112
Session 4: Inclusive Education	113
Title: Vocational Challenges in Teaching Students with Learning Difficulties: A Comprehensive Literature Review Author: Manal Aldaihani DOI: 10.20533/WCE.2025.0025	114
Title: Key Functions for the Development of a Health Promotion Programme in the Primary School Author: Veronique Regnier DOI: 10.20533/WCE.2025.0026	115
Title: Korean STEAM Education: The Unique Role of Emotion Author: Nicole E. Anderson DOI: 10.20533/WCE.2025.0027	116
Title: Assessment of Road Accessibility Barriers of People with Disabilities in Nigeria Authors: Aondoseer A. Atoo, Folake O. Akintayo DOI: 10.20533/iSociety.2025.0004	119
Session 5: Cyber Security	121
Title: Secure Software Brownfield Engineering – Physical & Logical Partition Identification Authors: Jadish Singh, Aspen Olmsted DOI: 10.20533/iSociety.2025.0005	122
Title: Cyber Immune Model for Combating DDOS Attacks Authors: Abereowo Oladimeji, Akinwomi Akintoba, Kuboye Bamidele, Alese Boniface K. DOI: 10.20533/iSociety.2025.0006	127
Title: Cybercrime Detection and Analysis Model for Evidence Collection Authors: O. Y. Ogunlola, I. D. Bolarinwa, B. K. Alese DOI: 10.20533/iSociety.2025.0007	132
Title: Enhancing Individual Cyber Resilience: A Holistic Approach to Online Safety and Security Authors: Arwa Alsaiari, Steven Furnell, Helena Webb DOI: 10.20533/iSociety.2025.0008	136

Session 6: ICT in Education	142
Title: The Agency and Resilience of Teacher Educators in the Age of AI: A Case Study of Teacher Educators in Taiwan	143
Author: Hung Jui-Hsuan	
DOI: 10.20533/WCE.2025.0028	
Title: Designing a CTF-style Activity as a Self-study Exam in a Computing Course	145
Authors: Pedro J. Roig, Salvador Alcaraz, Katja Gilly, Cristina Bernad, Carlos Juiz	
DOI: 10.20533/WCE.2025.0029	
Title: University Orientation and Conversational Technologies: A Study on the User Experience of an Embodied Conversational Agent	151
Author: Joy Ciliani	
DOI: 10.20533/WCE.2025.0030	
Session 7: e-Society	157
Title: A Mobile Decision Support System for Early Prediction of Mental Health Disorders	158
Authors: Divine Prince Daniel, Enesi Femi Aminu, Ayobami Ekundayo	
DOI: 10.20533/iSociety.2025.0009	
Title: Development of a Colon Cancer Classification System	166
Authors: V.I. Ainoko, A.O. Odeniyi, O.Y. Ogunlola, A. Innocent, B.K. Alese	
DOI: 10.20533/iSociety.2025.0010	
Title: Unmasking Fake Identities on Instagram Using Machine Learning	170
Authors: Olaniyi Abiodun Ayeni, Ethel Femi-Gbenoba	
DOI: 10.20533/iSociety.2025.0011	
Session 8: Internet Applications and Technology	177
Title: Impact of E-Commerce Adoption on the Performance of Hotel and Accommodation Businesses in Ondo State, Nigeria	178
Authors: Bolaji Comfort Adeniyi, Iyabode Abisola Adelugba	
DOI: 10.20533/iSociety.2025.0012	
Title: A Secure E-Shopping Platform for University Students: Case Study of FUTA	179
Authors: Afolabi Oladeinde, O.Y Ogunlola, Olumide Sosanya, M.T. Kinga	
DOI: 10.20533/iSociety.2025.0013	
Title: IoT-Based Energy Consumption Monitoring System	182
Authors: Olaniyi A. Ayeni, Aryee A. Feyisetan, Gbenga S. Oluokun	
DOI: 10.20533/iSociety.2025.0014	
Session 9: e-Society	185
Title: Breast Cancer Detection in Mammography Using a Stacked Ensemble of Three Pretrained Models	186
Authors: A.D. Adegoke, A.O. Ogunde, B.O. Olorunfemi	
DOI: 10.20533/iSociety.2025.0015	

Title: Development of a smart based medication, dispenser and prescription system Authors: Ololade Oladimeji, Enesi Femi Aminu, Ayobami Ekundayo DOI: 10.20533/iSociety.2025.0016	198
Title: A Machine Learning Framework for Hemoglobin-Based Severity Classification in Sickle Cell Disease Names: Oluwale Olajide, Aderonke Sakpere, Adesesan Adeyemo, Godwin Ogbale, Toluwase A. Olowookere, Samson Arekete, Benjamin Aribisala DOI: 10.20533/iSociety.2025.0017	207
Session 10: Cyber Security and Secured Transactions	213
Title: Access Control System for Financial Institutions Using RBAC. Authors: Olatoun Deborah Adeyanju, Olaniyi Abiodun Ayeni DOI: 10.20533/iSociety.2025.0018	214
Title: CryptoTrust Framework: Asymmetric Encryption-Integrated Zero Trust Architecture Authors: A.O. Oronti, B.K. Alese, O. Olabode, O.A. Akinsowon DOI: 10.20533/iSociety.2025.0019	217
Title: Mitigating SSRF Attacks Through a Zero Trust Security Model Authors: Olaibi David, Yetunde O. Ogunlola, B.K. Alese DOI: 10.20533/iSociety.2025.0020	221
Session 11: Intelligent Control	226
Title: Development of an Intelligent Traffic Management System to Address Traffic Signal Visibility Occlusion in Ibadan, Nigeria Authors: Agbede Caleb Oluwale, Akintayo Folake Olubunmi, Agbede Oluwale Akinyele DOI: 10.20533/iSociety.2025.0021	227
Title: Least Significant Bit (LSB) Audio Steganography with AES 256 Encryption Authors: Olaniyi Abiodun Ayeni, Adekunle Obajimi Isaac DOI: 10.20533/iSociety.2025.0022	237
Title: CNN-Based Detection of Image Manipulation Authors: D. O. Babatunde, O.Y. Ogunlola, B.K. Alese DOI: 10.20533/iSociety.2025.0023	244
Session 12: e-Society	248
Title: Fusion Strategies of Convolutional Neural Networks and Vision Transformers: A Comparative Study on Binary-Class Cancer Classification Authors: Aanuoluwa Oyebola Adio, Oluwaseun Favour Amore, Toluwase Ayobami Olowookere DOI: 10.20533/iSociety.2025.0024	249
Title: A Decision Support Framework for Predicting Soil Nutrients Based on Machine Learning Models and Sensors Technology Authors: Taoheed Abioye Abdurraheem, Enesi Femi Aminu, Ayobami Ekundayo DOI: 10.20533/iSociety.2025.0025	255

Title: A Multi-Criteria Evaluation of Urban Transport System Performance in Makurdi Metropolis 262
Authors: Emmanuel O. Nwafor, Folake O. Akintayo, Inyeneobong E. Edem
DOI: 10.20533/iSociety.2025.0026

Session 13: Artificial Intelligence 264

Title: A Lightweight and Explainable Hybrid AI Approach for Face Morphing Attack Detection in 265
High Security Environments
Authors: Oluwafemi A. Sanya, Amarachi F. Peters, Opani M. Aweh, Abiodun Oguntimilehin,
Tope-Oke Adebisola, Oluwatoyin B. Abiola, Okebule Toyin
DOI: 10.20533/iSociety.2025.0027

Title: Advancing Data Security in Cloud Computing Through Secure File Sharing 274
Authors: Olaniyi Abiodun Ayeni, Adejumo Ibitola Elizabeth, Wende Olaoluwa Mayowa
DOI: 10.20533/iSociety.2025.0028

Title: Leveraging on Honeypot to Detect Cyber Attack on Critical Infrastructure 282
Title: Olaniyi Abiodun Ayeni, Onibokun Oluwapelumi Oluwabori, Oluokun Samuel Olugbenga
DOI: 10.20533/iSociety.2025.0029

WCE-2025 Executive Committees

General Chair

Charles A. Shoniregun, Infonomics Society, UK and ROI

Steering Committee Chair

Charles A. Shoniregun, Infonomics Society, UK and ROI

Steering Committees

Cynthia Northington-Purdie, William Paterson University, USA
Barba Aldis Patton, University of Houston-Victoria, USA
Emmanuel Hooper, Harvard University, USA

Publicity Chair

Galina Akmayeva, Infonomics Society, UK and Ireland

Conference Coordinator

Margaret Smith, Infonomics Society, UK

Research Student Forum Chair

Marta Cabral, Columbia University, USA

International Chair

Tintswalo V. Manyike, University of South Africa, South Africa

Program Committees

Regis Chireshe, College of Education, University of South Africa, South Africa
Isao Miyaji, Okayama University of Science, Japan
Olaniyi Alaba Sofowora, Obafemi Awolowo University, Nigeria
Marta Cabral, Teachers College, Columbia University, USA
Lorayne Robertson, University of Ontario Institute of Technology (UOIT), Canada
Christopher Kong, Hua Chin University of Tasmania, Australia
David M. Chapinski, Rutgers, The State University of New Jersey, USA
Elizabeth Magaret, St. Ann's College for Women, India
Mikhail A. Rodionov, Penza Pedagogical State University, Russia
Ahmed Ibrahim Usman, Federal College of Education, Kano, Nigeria
Barba Aldis Patton, University of Houston-Victoria, USA
Arend E. Carl, Stellenbosch University, South Africa
Florence Nkechi Okeke, Imo State University, Nigeria
Zhanna Dedovets, The University of The West Indies, Trinidad and Tobago
Kamla Suleiman Salim Al Aamri, Sultan Qaboos University, Sultanate of Oman
Abiodun Popoola, University of Ado-Ekiti, Nigeria
Julie A. Holmes, Louisiana Tech University, USA
Luyanda Dube, University of South Africa, South Africa
Maria Dorota Stec, University of Silesia in Katowice, Poland
Veli Jiyane, University of South Africa, South Africa
Faten Saad Mahmoud Abdel-Hameed, University of Bahrain, Bahrain
Ben John Siyakwazi, Great Zimbabwe University, Zimbabwe
Paul Walsh, Cork Institutie of Technology, Ireland

i-Society 2025 Executive Committees

General Chair

Charles Shoniregun, Infonomics Society, UK and Ireland

Steering Committees Steering Committees

Chan Yeob Yeun, KUSTAR, UAE
Daniel Fischer, Technische Universität Ilmenau, Germany
Frank Wang, University of Kent, United Kingdom
George Ghinea, Brunel University, United Kingdom
Ion Tutanescu, University of Pitesti, Romania
Mohamed Jamal Zemerly, Khalifa University of Science, Technology and Research, UAE
Paul Hofmann, Senseforce, Austria
Safwan El Assad, Ecole polytechnique de l'université de Nantes, France

Technical Programme Committees

Ina Taralova, Laboratoire des Sciences du Numérique de Nantes LS2N, France
Fredrick Japhet Mtenzi, Aga Khan University, Dar es Salaam, Tanzania
Daniel Fischer, Technische Universität Ilmenau, Germany
Takeshi Yamakawa, Fuzzy Logic Systems Institute, Fukuoka, Japan
Geoffrey Skinner, The University of Newcastle, Australia
Gustavo Rossi, Lfia – Universidad Nacional De La Plata, Argentina
Galina Akmayeva, Infonomics Society, UK and ROI
Mudasser Wyne, National University, United States
Rory McGreal, Athabasca University, Canada
Kang Zhang, The University of Texas at Dallas, United States
Matjaz B. Juric, University of Ljubljana, Slovenia
Bernard Moulin, Université Laval, Canada
Bernd Markscheffel, Technische Universität Ilmenau, Germany
Dana Petcu, Western University of Timisoara, Romania
Maslin Masrom, Universiti Teknologi Malaysia International Campus, Malaysia
C. Suresh Gnana Dhas, Vel Tech Multi Tech, India
Peter Milligan, Queen's University Belfast, Northern Ireland, United Kingdom
Roberto Pereira, University of Campinas (UNICAMP), Brazil
Xia Xie, Huazhong University of Science and Technology, China
Christian Werner, University of Lübeck, Germany
Jeannette Chin, Anglia Ruskin University, United Kingdom
Ricardo Rabelo, Universidade Federal de Santa Catarina, Brasil
Camarinha-Matos Luis M., New University of Lisbon, Portugal
Vyacheslav Grebenyuk, Kharkiv National University of Radioelectronics (KNURE), Ukraine
Princely Ifinedo, Cape Breton University, Canada
Samir Tata, Telecom SudParis, Institut Mines-Telecom, France
Robert Joseph Skovira, Robert Morris University, United States
Abdullah M. Alnajim, Qassim University, Saudi Arabia
Nicola Fanizzi, Università di Bari, Italy
Michal Krátký, Technical University of Ostrava, Czech Republic
Ralph Deters, University of Saskatchewan, Canada
Epaminondas Kapetanios, The University of Westminster, United Kingdom
Soraya Kouadri Mostéfaoui, The Open University, United Kingdom
Andreas Kühn, Bern University of Applied Sciences, Switzerland
Terry Fanning, Imperial College London, United Kingdom
Jiannong Cao, Hong Kong Polytechnic University, Hong Kong
Lundy Lewis, Southern New Hampshire University, United States
Hanan Lutfiyya, The University of Western Ontario, Canada
Chrisa Tsinaraki, Technical University of Crete, Greece

Niv Ahituv, Tel Aviv University, Israel
Dario Maggiorini, Università degli studi di Milano, Italy
Andrew Kusiak, The University of Iowa, United States
Ouri Wolfson, University of Illinois at Chicago, United States
Costin Badica, University of Craiova, Romania
George Spanoudakis, City University, United Kingdom
Ali Farooq, University of Turku, Finland
Wong Seng Yue, Taylor's University, Malaysia
Marcus Specht, Open University Of The Netherlands, Netherlands
Robert Bestak, Czech Technical University in Prague, Czech Republic
Mustafa Atay, Winston-Salem State University, United States
Kateryna Synytsya, IRTC ITS, Ukraine
Antoanela Naaji, Vasile Goldis Western University Arad, Romania

Speakers

Speaker 1

Speaker 1: Thiago Vieira



Thiago Vieira is an Executive Director of Cybertech Acceleration programme in Portugal. He has his Master's degree in European Union Business Law, and is currently pursuing a PhD in AI. Thiago is a digital forensic expert appointed by the Porto Court of Appeal in Portugal and the Minas Gerais Courty of Justice in Brazil. Thiago Vieira is a former developer and has sixteen years of experience in tech. He is also an educational leader on awareness-raising projects related to cybersecurity, soft skills, networking, and the risks of artificial intelligence inside schools.

Future of Cyber Resilience

Thiago Vieira
Cybertech
Portugal

Abstract

As the digital world becomes ever more integral to education, communication, and daily life, the need to cultivate cyber resilience from an early age has become a global imperative. Cyber resilience, defined as the ability to anticipate, withstand, recover from, and adapt to cyber threats, can no longer be confined to corporate or governmental settings. Children and young people must be equipped with not only digital skills, but also the critical mindset and emotional readiness to navigate an increasingly complex online environment.

The future of cyber resilience begins in the classroom. By introducing real-life digital scenarios through project-based learning, simulations, and age-appropriate challenges, we can shift from theoretical instruction to practical resilience training. This proactive approach empowers students to recognize risks, make informed decisions, and recover from setbacks, both online and offline.

This presentation explores innovative pedagogical strategies, interdisciplinary models, and policy frameworks that aim to embed cyber resilience into the heart of education systems. Preparing resilient digital citizens is not only an educational goal, it is a societal necessity.

Speaker 2

Speaker 2: Deketa Tranae



Deketa Tranae is the Founder and CEO of Pure Necessity By Deketa Tranae® - a lifestyle service that specializes in teaching others how to define and build a legacy both personally and professionally. Deketa Tranae is a faith shifter who dares professionals to tap into their values, skills, and network to be positioned for a career building breakthrough. As an extension of her human resources background and as a certified John Maxwell speaker and consultant, she also leverages some of these same elements to mentor, coach, and consult company and organization leaders (and their support teams) to their highest potential.

The Importance of Building Community

Deketa Tranae
Pure Necessity By Deketa Tranae®
USA

Abstract

In "The Importance of Building Community," I share strategies for cultivating meaningful personal and professional relationships that lead to lasting impact. I encourage audiences to actively engage with others, embrace discomfort as a catalyst for growth, and stay consistent in their efforts. Grounded in my belief in faith, patience, and trusting the process, my talk highlights how these principles can create resilient communities and empower individuals to succeed. Whether navigating workplace dynamics or personal challenges, I aim to inspire others to build supportive environments where everyone can thrive.

Speaker 3

Speaker 3: Dr Janet Chan



Dr Janet Chan is a Senior Lecturer at the School of Biological Sciences and a member of the Institute for Climate and Carbon Neutrality at the University of Hong Kong (HKU). She serves as the Program Coordinator and Chief Examiner for HKU Master of Science in Environmental Management program. Dr Chan's research and teaching focus on environmental and sustainability education, waste management, environmental pollution, environmental health, and research methods. Her commitment to research and teaching has earned her various accolades, including her appointment as a panel member for the Research Assessment Exercise (RAE) 2026 by the Hong Kong University Grants Committee. She has also received the Advanced Materials Innovation Award (2025) from the International Association of Advanced Materials (Sweden), University's Outstanding Teaching Award (2020), Faculty of Science's Teaching Excellence Award (2019), and was recognized as an Inaugural Member of the HKU Teaching Academy (2022). Additionally, Dr Chan holds the Senior Fellowship of Advanced HE since 2022 and the Professional Membership of the Hong Kong Institute of Qualified Environmental Professionals (HKIQEP) since 2017. As Co-founder and Program Leader of the university-wide initiative "HKU Climate Ambassador," Dr. Chan has led the program to win multiple local and international awards. These include being a finalist for the International Green Gown Award (2025), PIONEER Award (2025) and the QS Reimagine Sustainability Education Action Award (2024); securing Advance HE Global Impact Grant (2024), and HKU Teaching Development Grant Award (2024); winning the Gold Award at the HKTEA Design Thinking CoP Competition (2024); and receiving the UGC Teaching Development and Language Enhancement Grant (TDLEG) for 2022-25 (focused on experiential learning and AI). Under Dr Chan's leadership, the HKU Climate Ambassador program will expand to four universities in Hong Kong, supported by the UGC TDLEG. To further extend her research and teaching impact, Dr. Chan actively serves as an advisor of various environmental professional societies. She is frequently consulted by government bodies, environmental organizations, and social NGOs for her expertise and advice. Prior to her academic career at HKU, Dr Chan served at the headquarters of an international social NGO in Geneva, Switzerland. In this role, she spearheaded the development of programs focused on the Millennium Development Goals and held the esteemed position of the organization's global focal point on climate change. Her work was dedicated to promoting youth empowerment and creating sustainable change.

Bridging the Knowledge-Action Gap in Education for Sustainable Development: A Student-Led Partnership Model from Hong Kong

Janet KY Chan^{*1,2}, Peter FM Lau³, TY Hui⁴, Zoe Liao^{1,2}, Sunny Chiu^{1,2}

¹*School of Biological Sciences, Faculty of Science, The University of Hong Kong, Hong Kong SAR, China*

²*Master of Science in Environmental Management, Faculty of Science, The University of Hong Kong, Hong Kong SAR, China*

³*Teaching and Learning Innovation Centre (TALIC), The University of Hong Kong, Hong Kong SAR, China*

⁴*Science Unit, Lingnan University, Hong Kong SAR, China*

Abstract

*Education for Sustainable Development (ESD) is crucial for addressing global challenges like climate change and social inequity [1]. However, traditional ESD often emphasizes theory over practice, creating gaps between knowledge and action [2]. This study investigates how the Students as Partners (SaP) model—where students collaborate with faculty and practitioners on curriculum and solutions [3]—can bridge this divide. Focusing on HKU's Climate Ambassador Program, we evaluate a three-tiered SaP framework: (1) student-faculty co-creation of ESD curricula, (2) industry-academic applied learning, and (3) student-led SDG-aligned initiatives. Mixed-methods analysis (pre-post questionnaires: *n* = 60; interviews: *n* = 17) revealed significant behavioral improvements (38% less single-use plastics, 27% more plant-based diets) and increased institutional-change agency. Qualitatively, 89% of participants adopted "sustainability advocate" identities, demonstrating SaP's role in fostering interdisciplinary collaboration and fulfilling UNESCO's action-oriented ESD vision [1,4,5,6]. The findings highlight SaP's potential to transform sustainability education by merging theory with real-world impact.*

1. Introduction

Under the accelerating threats of climate change, resource depletion, and social inequity, Education for Sustainable Development (ESD) has emerged as a critical pedagogical framework to empower learners as agents of systemic change [1]. UNESCO's ESD for 2030 roadmap positions ESD as a transformative tool to achieve all 17 Sustainable Development Goals (SDGs), emphasizing its role in fostering "cognitive, socio-emotional, and behavioral dimensions of learning" for sustainability [7]. While institutions increasingly adopt ESD principles, traditional pedagogies often reduce sustainability

education to siloed environmental science modules, neglecting the interconnected social, economic, and ethical dimensions central to ESD [4]. This disconnect perpetuates a critical gap: learners acquire sustainability knowledge but lack the agency, interdisciplinary skills, or collaborative frameworks to drive tangible action [8,9].

This study addresses this challenge by examining how participatory, student-led models can operationalize ESD's transformative potential. Focusing on the University of Hong Kong's (HKU) Climate Ambassador Program, the research evaluates the Students as Partners (SaP) approach—a collaborative pedagogy where students co-create curricula and solutions with faculty and practitioners [3]—as a mechanism to bridge ESD's knowledge-action gap. By analyzing shifts in participants' sustainability competencies, leadership identity, and pro-environmental behaviors, the study contributes to global efforts to reorient education systems toward justice-centered, action-driven ESD frameworks [10].

2. Literature Review

2.1 ESD: A Framework for Holistic Sustainability

ESD transcends narrower constructs like Environmental Education (EE) or Climate Change Education (CCE) by integrating three pillars of sustainability: environmental integrity, social equity, and economic viability [4]. While EE focuses on ecological literacy and CCE on climate-specific knowledge, ESD emphasizes systems thinking—the ability to analyze interconnections between global challenges like poverty, inequality, and resource consumption [11]. For example, an ESD curriculum might explore how urban food waste (environmental) exacerbates hunger (social) while inefficient supply chains (economic) contribute to both. UNESCO further positions ESD as a driver of "green and blue transitions," equipping learners to critically engage

with sustainability dilemmas, from renewable energy adoption to ethical AI development [12].

2.2 Barriers to Effective ESD Implementation

Despite its recognized importance, ESD faces systemic implementation challenges. A UNESCO audit of 100 national curricula found only 53% reference climate change, with fewer addressing ESD's social or economic dimensions [1]. Even when adopted, ESD programs often prioritize didactic knowledge transfer over participatory action, resulting in a persistent knowledge-action gap [13]. For instance, students may understand carbon footprint concepts but lack opportunities to design campus decarbonization strategies. This gap reflects a broader pedagogical mismatch: ESD's emphasis on action competence [14]—the ability to critically assess and address sustainability issues—requires learner-centered approaches rarely employed in traditional curricula [15].

2.3 Students as Partners (SaP): A Transformative Pedagogy for ESD

The Students as Partners (SaP) model, defined as "a collaborative, reciprocal process where students and staff contribute equally to curricular or pedagogical design" [3], offers a promising pathway to address ESD's implementation gaps. While SaP originated in general education research, its principles align with ESD's goals:

- **Agency Development:** By co-creating sustainability projects, students transition from passive learners to active problem-solvers, building the leadership identity critical for ESD [16].
- **Interdisciplinary Collaboration:** SaP's emphasis on peer/mentor partnerships mirrors ESD's call for cross-sectoral solutions.
- **Real-World Impact:** SaP's action-oriented projects operationalize ESD's mandate to link learning with community-level change [1].

Empirical studies demonstrate SaP's efficacy in sustainability contexts. For example, Manor et al. [17] found that students in partnership models exhibited 32% higher self-efficacy in addressing environmental challenges compared to traditional cohorts. Similarly, Lau & Zheng [18] documented how SaP-driven climate initiatives in Southeast Asian universities increased student-led sustainability innovations by 41%.

3. Methodology

3.1 Program structure and procedure

This study adopts a mixed-methods approach to examine the impact of the Climate Ambassador Program at the University of Hong Kong (HKU) on

undergraduate students' climate-related attitudes, knowledge, and behaviors. Sixty undergraduate students were recruited from all ten faculties across HKU at the beginning of the study as three different batches, to ensure a diverse and interdisciplinary cohort reflective of the university's academic breadth. These students participated as climate ambassadors, engaging in a structured program designed to enhance their understanding of climate change and initiate climate action within their communities. The participating students were organized into small groups of three to four members each. Each group was paired with an experienced mentor specializing in environmental or climate-related fields to provide guidance and instructions throughout the program. As a culminating requirement, every group developed and presented an original climate action initiative or implementation plan by the program's conclusion.

Through workshops, local and overseas field trips, service-learning, and team-building activities, student participants build their capacity in solving complex challenges during the program. Training topics cover Climate Change101, Waste Management, Low Carbon Living, Biodiversity Conservation, and Environmental Education. The program provides "horizontal" learning by empowering students to create their own climate-related initiatives in partnership with peers, mentors, teachers and stakeholders. Equipped with knowledge and skills acquired through the program, students have executed over 15 sustainable actions ranging from waste reduction, sustainability education, water pollution, electronic waste, food waste, office waste, climate change, biodiversity, to sustainable living.

3.2 Three levels of the Students as Partners Model

The Climate Ambassador Program operationalizes the Student as Partners (SaP) framework through in total three distinct but interconnected levels of engagement:

- **Level 1 Co-Creation of Program Vision:** At the foundational level, two Master of Science in Environmental Management students partnered with a faculty leader (Professor Dr. Janet K.Y. Chan) as co-founders of the initiative. This partnership model positioned students as equal stakeholders in program design, with the Professor serving as both mentor and collaborator.
- **Level 2 Industry-Academic Partnership:** The second level engaged environmental professionals and industry experts as mentors who worked alongside student groups in a partner capacity. Unlike conventional mentorship models that emphasize unidirectional knowledge transfer, this approach

created interactive learning relationships where both students and practitioners contributed expertise. Industry partners and mentors provided real-world perspectives while students offered innovative, reality-based thinking.

• **Level 3 Student-Led Climate Action:** At the action implementation level, student groups developed and executed original climate action plans. This last tier of partnership empowered participants to transition from learners to other roles, such as leaders, creators and advocates, applying acquired knowledge to design real-life based sustainability initiatives. The process cultivated student agency, enriched students' identities while demonstrating the program's translational impact beyond academia.

3.3 Quantitative Data Collection

The quantitative component utilized a pre-post questionnaire design to assess changes in participants' understanding in various aspects of environmental and climate-related topics. The same questionnaire was administered to the students before and after the program, allowing for within-subject comparison and measurement of individual development over time. The instrument comprised four sections: (1) environmental attitudes, (2) understanding of climate change, (3) engagement in low-carbon behaviors, and (4) climate change mitigation. This structure aligns with established frameworks in climate education research, which emphasize the importance of measuring not only cognitive outcomes but also behavioral intentions and eventually, climate actions. The repeated-measures design enhances the reliability of the findings by controlling baseline differences among participants. The findings are intended to indicate the implications of the questionnaire result difference.

3.4 Quantitative Data Collection

To complement the quantitative data, in-depth individual interviews were conducted with a purposive sample of student participants following program completion. The individual interview protocol consisted of five thematic sections: (1) motivation for joining the project, project awareness, and experience, (2) perceptions of climate education outcomes, attitudes, and behaviors, (3) student agency, (4) identity development-including self-identity and interactions with mentors and peers-and (5) reflection and intentions for future climate action. This approach draws on best practices in environmental education research, which highlight the value of qualitative inquiry for capturing slight or major shifts in motivation, identity, and student agency that may not be fully reflected in quantitative measures.

3.5 Data Analysis

Quantitative data from the pre- and post-program questionnaires will be analyzed using a variety of statistical methods to identify significant changes in the four measured domains. The analysis calculates score differences for each participant to track individual changes over time, using within-subject analysis to ensure reliability. Principal Component Analysis (PCA) visualizes overall response patterns across different batches and traits. Non-parametric statistical tests, including Kruskal-Wallis tests and Cumulative Link Models, compare results between batches and faculties. Correlation analyses using Kendall's tau coefficient explore relationships between different traits.

Qualitative interview data will be transcribed and subjected to thematic analysis, enabling the identification of recurring patterns and emergent themes related to students' experiences, learning processes, and personal development as climate ambassadors. Integrating these two data streams provides more comprehensive and multifaceted understandings of the program's impact.

4. Progress and Preliminary Findings

4.1 Status of Data Collection

Data collection for the research on HKU Climate Ambassador Program has been successfully completed. After careful data cleaning and validation, the final dataset includes independent pre- and post-program responses from 43 undergraduate students across all ten faculties at the University of Hong Kong. Participants were recruited from multiple cohorts, including Batch 1, Batch 2, and the Summer cohort. Specifically, 12 students were drawn exclusively from Batch 1, 5 students participated in both Batch 1 and 2, another 5 students were from Batch 2 only, and 21 students were part of the Summer cohort. In addition to the questionnaire responses, 17 individual interviews were conducted with students from 3 different recruitment batches, with a purposive subset of participants to collect deeper qualitative insights regarding their perceptions, learning experiences, and development throughout the program.

4.2 Preliminary Findings

The current results from quantitative analysis highlight several promising outcomes from the environmental education program. The progress so far indicates promising trends across several key domains. Notably, significant improvements were recorded in participants' engagement with low-carbon behaviors and climate change mitigation actions following program participation. Students reported greater efforts in reducing single-use plastic consumption, adopting plant-based diets, minimizing

carbon footprints, and demonstrating increased confidence in influencing others toward sustainable practices. These positive shifts suggest the program effectively encouraged actionable, pro-environmental habits among participants. In contrast, changes in environmental attitudes were more modest at the overall scale; however, at the level of individual questionnaire items, positive shifts were observed. In particular, students exhibited stronger agreement with statements recognizing the severe consequences of ecological collapse if action is not taken, and showed a decreased acceptance of anthropocentric views, as reflected in greater recognition of the intrinsic rights of plants and animals to exist and a reduced belief in the dominance of humans over nature.

The preliminary findings from quantitative analysis so far suggest that the HKU Climate Ambassador Program has been effective in cultivating climate-conscious behaviors and strengthening the climate action capacities of participating students. It demonstrates that the program successfully advanced key effective and behavioral outcomes crucial for empowering future climate leaders. These early insights offer valuable guidance for the ongoing refinement and scaling of the program and provide a strong foundation for further analysis in the next phase of the research. Future research directions will include a qualitative analysis of in-depth interviews with student participants. This approach will provide richer insights into their personal experiences, motivations, and perceived impacts of the program, complementing the quantitative findings. By exploring individual perspectives, we aim to uncover deeper themes, such as barriers to behavior change or the emotional and cognitive drivers behind shifts in environmental attitudes.

5. References

- [1] UNESCO. (2021). UNESCO declares environmental education must be a core curriculum component by 2025. Retrieved from <https://www.unesco.org>
- [2] Kollmuss, A., & Agyeman, J. (2002). Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, 8(3), 239-260.
- [3] Cook-Sather, A., Bovill, C., & Felten, P. (2014). *Engaging students as partners in learning and teaching: A guide for faculty*. John Wiley & Sons.
- [4] UNESCO. (2017). Education for sustainable development goals: Learning objectives. Retrieved from <https://unesdoc.unesco.org>
- [5] Kolenatý, M., Kroufek, R., & Činčera, J. (2022). What Triggers Climate Action: The Impact of a Climate Change Education Program on Students' Climate Literacy and Their Willingness to Act. *Sustainability*, 14(16), 10365.
- [6] Lubicz-Nawrocka, T. M. (2018). Students as partners in learning and teaching: The benefits of co-creation of the curriculum. *International Journal for Students as Partners*, 2(1), 47-63.
- [7] UNESCO. (2021). Getting every school climate-ready: How countries are integrating climate change issues in education. Retrieved from <https://unesdoc.unesco.org>
- [8] Bamberg, S., & Möser, G. (2007). Twenty years after Hines, Hungerford, and Tomera: A new meta-analysis of psycho-social determinants of pro-environmental behaviour. *Journal of Environmental Psychology*, 27(1), 14-25.
- [9] Kollmuss, A., & Agyeman, J. (2002). (As above)
- [10] Stevenson, R. B. (2022). Approaches to Education for Sustainability. *Oxford Research Encyclopedia of Education*.
- [11] Sterling, S. (2010). Learning for resilience, or the resilient learner? Towards a necessary reconciliation in a paradigm of sustainable education. *Environmental Education Research*, 16(5-6), 511-528.
- [12] UNESCO. (2023). Ecological sciences for sustainable development. Retrieved from <https://en.unesco.org>
- [13] Kolenatý, M., et al. (2022). (As above)
- [14] Mogensen, F., & Schnack, K. (2010). The action competence approach and the 'new' discourses of education for sustainable development, competence and quality criteria. *Environmental Education Research*, 16(1), 59-74.
- [15] Stevenson, R. B. (2022). (As above)
- [16] Lubicz-Nawrocka, T. M. (2018). (As above)
- [17] Manor, C., Block-Schulman, S., Flannery, K., & Felten, P. (2010). Foundations of Student-Faculty Partnerships in the Scholarship of Teaching and Learning. In C. Werder & M. M. Otis (Eds.), *Engaging Student Voices in the Study of Teaching and Learning* (pp. 3-15). Stylus.
- [18] Lau, P., & Zheng, J. (2025). Exploring the Interplay Among Student Identity Development,

University Resources, and Social Inclusion in Higher Education: Analyzing Students as Partners Project in a Hong Kong University. *Social Sciences*, 14(3), 119.

Speaker 4

Speaker 4: Lyberius Ennio F. Taruc



Lyberius Ennio F. Taruc is a professional computer engineer. His areas of specialization are SDN-NFV, Cloud Computing, Virtualization and Containerization, Edge Computing using RaspberryPi, Machine Learning and Linux Operating Systems. Lyberius is currently pursuing a Master's degree at the Polytechnic University of the Philippines. He is a full-time manager and cloud infrastructure engineer at the Network Technical Group (NTG), Globe Telecom, Inc.

Generative AI to Assist Research Journaling in the Academe

Lyberius Ennio Taruc
FEU-Institute of Technology
Manila, Philippines

Abstract

As the use of AI becomes more evident across different sectors, it is important to acknowledge that the academic community is also facing challenges in its ethical use. This talk explores the use of generative AI tools to support research journaling practices within academic settings. It highlights how AI can assist researchers (especially students) in organizing thoughts, maintaining consistency in documentation, and enhancing reflective writing throughout the research process. This talk aims to demonstrate how generative AI can become a collaborative partner in knowledge creation—streamlining workflow without compromising academic integrity—while avoiding the pitfalls on research, such as plagiarism and AI hallucinations.

Speaker 5

Speaker 5: Dr Ismail Md. Zain



Dr Ismail Md. Zain is a Former “Special Grade Senior Excellent Lecturer”, Malaysia Teacher’s Institute of Education, Malaysia. Currently a Director of Dynamic Global Vision Consultancy and founder of The Collaborative Instructional Design System - CIDS New-Gen. He obtained a B.A. (Hons) Humanities with Education 1979, M.Ed. 1987, and PhD. in Instructional Design, 1994 from University Sains Malaysia. He is also a former committee member of the Malaysia Smart School implementation Team, and a member of the special think tank team for the Malaysia Institute of Education – Malaysia Ministry of Education. He wrote several books published locally for teachers and students, and guest speaker for local TV Talk shows. He is actively involved in presenting papers as a keynote speaker at several international conferences and editorial team member in several international journals. He was awarded “Wall of Fame” from the Centre for Education Technology and Multimedia from the same university, 2017. Currently, he is committed to enhancing teachers’ professionalism in visualizing the 21st-century education needs.

The Collaborative Instructional Design System (New-Gen): An Ai-Assisted Multidimensional Aspect of Instructional Design Towards VUCA World

Ismail Zain
Dynamic Global Vision
Malaysia

Abstract

The transformation of the education system is evolving fast. Besides knowledge, other competencies with multiple dimensions are needed in developing a holistic personality to meet the VUCA world (volatility, uncertainty, complexity, and ambiguity). Teachers are responsible for designing their instruction to prepare learners for more rapid economic and social change in their future careers. However, in the era of digitalization, AI-assisted features are capable of identifying the competencies required. Thus, this system aims to provide an integrated and effective approach for designing instructions as a solution to the above global issues. Applying the Integral ASIE Instructional Design Model as an innovative online planning tool in the system with AI-assisted encompasses the competencies formulated in the planning elements adapted from the Four-Dimensional Education Framework (Centre for Curriculum Redesign, 2015, 2024). It has fundamentally encompassed the philosophical attributes of metaphysics, epistemology, axiology, ethics, and logic where the instructional design principles and theories as the background of the system are applied. The model served as a procedural flow of designing instructions in a simple, systematic, and professional manner, creating high-quality instructional planning activity. Besides providing important features for 21st-century learning needs with AI-assisted, planning items are customizable where teachers are given autonomy to fulfill their needs and requirements of the curriculum, integrating external links and resources for virtual or hybrid classroom interactions while the planning activities are shareable. It is work-effective, time-effective, and cost-effective in motivating teachers to prepare their instructional planning activities. Thus, this system has a significant impact on teachers and administrators as best classroom practices in the process of redesigning and reengineering academic directions, visualizing the required education landscape, and preparing learners' future career paths in facing the VUCA world.

Speaker 6

Speaker 6: Dr Ioannis Makris



Dr Ioannis Makris is a distinguished composer, conductor, and educator with a specialization in Special Education. He is a full professor of music in Greek State Special Schools, serving under the Directorate of Secondary Education of West Athens. He is also a scientific associate in Counseling Psychology at the Higher School of Pedagogical and Technical Education in Athens. Dr Makris is the author of the pedagogical method "The Differentiated Didactic Approach to Teaching the Arts (D.D.A.T.A.) in Special Education," as well as several educational books widely used in both Higher and Secondary Education in Greece. He holds a Ph.D. in Psychology (2003) and a Post-Ph.D. (2006) from the Ecole Pratique des Hautes Etudes. He also holds Master's degrees in Music History (D.E.A./University Paris IV-Sorbonne), Special Education, and Educational Management (Med/University of Nicosia). His fundamental studies in musicology were completed at Paris IV-Sorbonne University, and he holds degrees in Classical Guitar and Byzantine Music. He has also pursued extensive professional musical training in countries including the USA, UK, Russia, Austria, Germany, Spain, the Czech Republic, Bulgaria, and Singapore. Dr Makris has been an associate faculty member in Higher Education in Greece since 1997, and since 2005, he has served as a scientific associate in the Pedagogical Department of the Higher School of Pedagogical and Technical Education, specializing in Psychology. He has been a Judge for the Medea Awards from 2016 to 2025 in Brussels and a Reviewer for international scientific journals and conferences in countries such as Portugal and Latvia. Furthermore, he has actively participated in numerous European programs, including Erasmus+ KA1, KA2, KA3, Comenius Regio, and Grundtvig.

The Differentiated Didactic Approach to Teaching the Arts (D.D.A.T.A.): A Holistic Intervention for Skill Development in Students with Intellectual Disabilities and Autism Spectrum Disorders

Ioannis Makris

*Directorate of Secondary Education C Athens-Special State School of Peristeri
Greece*

Abstract

This study investigates the effectiveness of the Innovative Differentiated Didactic Approach to Teaching the Arts (D.D.A.T.A.) as a pedagogical tool for enhancing various skills in students with intellectual disabilities. Traditional arts education often faces challenges in effectively engaging and supporting this diverse student population. D.D.A.T.A. addresses these challenges by offering a flexible, individualized, and multi-sensory framework that adapts to the unique learning styles and cognitive abilities of students with intellectual disabilities. The research employs a mixed-methods approach, combining quantitative data from skill assessments (e.g., fine motor skills, communication, social interaction, emotional expression) with qualitative data derived from observations, teacher interviews, and student work samples. Participants include students with varying degrees of intellectual disabilities across different educational settings. Preliminary findings suggest that the implementation of D.D.A.T.A. leads to significant improvements in targeted skill areas, fostering enhanced creativity, self-expression, and overall engagement in learning. Furthermore, the approach appears to promote a more inclusive classroom environment, empowering students to participate actively and develop a sense of accomplishment. This study contributes to the growing body of literature on inclusive education and provides valuable insights into effective arts-based interventions for students with intellectual disabilities, advocating for the wider adoption of differentiated instructional strategies in special education curricula.

1. Introduction

The education of students with intellectual disabilities (ID) and Autism Spectrum Disorders (ASD) presents significant challenges due to their diverse learning needs and cognitive profiles. Traditional didactic approaches often limit participation and skill development, particularly in arts education, which, despite its potential for promoting expression and creativity, frequently lacks individualized and accessible methods for this

population [11]. The Differentiated Didactic Approach to Teaching the Arts (D.D.A.T.A.) offers an innovative, sustainable, and multi-sensory framework designed to address these challenges. D.D.A.T.A. adapts to the unique learning styles and cognitive abilities of students with ID, enhancing accessibility and effectiveness through its core techniques: visualization and verbalization [16]. This approach fosters active participation and empowerment, contributing to enhanced creativity, self-expression, and overall engagement in learning by leveraging inherent artistic inclinations. It cultivates an environment where students feel capable, valued, and develop a strong sense of accomplishment. This study quantitatively and qualitatively investigates D.D.A.T.A.'s effectiveness as a pedagogical tool for enhancing various skills in students with intellectual disabilities. Furthermore, it examines how D.D.A.T.A. promotes a more inclusive classroom environment, empowering active student participation and fostering a positive attitude towards learning. This research contributes significantly to the literature on inclusive education and effective arts-based interventions for students with special educational needs.

2. Literature Review

2.1. Challenges in the Education of Individuals with Intellectual Disabilities and ASD

Students with intellectual disabilities (ID) and Autism Spectrum Disorders (ASD) face various educational challenges, including cognitive, communication, social, self-regulation, and sensory difficulties [26], [8]. A primary goal of special education is to foster a broad spectrum of skills, such as cognitive (memory, attention, language), motor, emotional, communicative, and social abilities [9], [10], [11], [6], [21]. While skill teaching often involves imitation, indirect methods can also be effective [7]; for instance, rhythm instruction can facilitate counting skills [16].

2.2. The Differentiated Didactic Approach to Teaching the Arts (D.D.A.T.A.).

D.D.A.T.A. emerged as an innovative pedagogical approach to address the gap between ID/ASD students' needs and conventional teaching. Developed between 2012-2024 through extensive research and clinical application, D.D.A.T.A. leverages the multi-sensory and expressive nature of the arts for learning and skill development [13], [16], [17], [18].

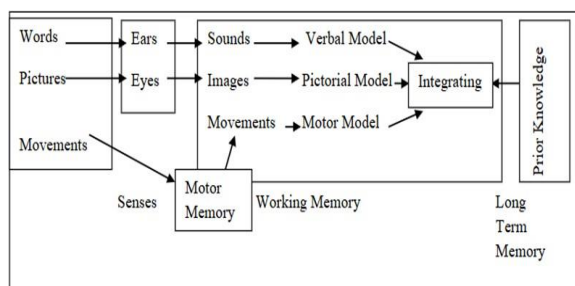
Its core techniques, visualization and verbalization, have yielded impressive results [14], characterized as "new" due to their unique application in scientific literature [16]. These techniques were successfully integrated into European Erasmus+ KA2 programs (T.E.L.L. Through Music, M.E.L.O.S., and CAP program "Cap sur l'école inclusive en Europe"), building upon the "Makris Method" [12]. Their consistent success across these initiatives underscores their practical value and contribution to inclusive education.

2.3. Theoretical Foundations of D.D.A.T.A.

D.D.A.T.A. is grounded in multiple theoretical approaches, providing a comprehensive framework for the education of students with ID/ASD:

(1) Mayer's Cognitive Theory of Multimedia Learning [25]: This theory, initially central for skill development, was expanded in D.D.A.T.A. to include a Motor Model. Clinical practice revealed the necessity of movement, derived from repetition and tactile sensation from musical instruments, alongside Mayer's Verbal and Pictorial Models in Working Memory. This revised model by Ioannis Makris is shown in Figure 1. Repetition and practice enhance tactile sensory stimuli, supporting movement and muscle strengthening for optimal student performance [16].

Figure 1. Revised Cognitive Theory of Multimedia Learning by Ioannis Makris



(2) Applied Behavior Analysis (ABA): D.D.A.T.A. integrates ABA principles like shaping, task analysis, and reinforcement [5]. This facilitates breaking down complex skills into manageable steps, enhancing

accessibility and successful activity completion. Systematic progress tracking and intervention adjustment are also ABA-based.

(3) Information Integration Theory (ITT) by Norman H. Anderson: Recognizing learning as an information integration process, D.D.A.T.A. aligns with ITT's focus on how individuals combine diverse information sources [1], [2]. The development of D.D.A.T.A. was supported by findings indicating that systematic provision of multiple, complementary stimuli (e.g., visual, auditory, tactile, motor) significantly improves ID/ASD students' understanding and performance by accommodating their cognitive particularities.

(4) Reversal Theory [3]: D.D.A.T.A. applies Reversal Theory to understand student motivation, adapting activities to their shifting metamotivational states (e.g., play vs. seriousness). This offers opportunities for enjoyment, challenge, and autonomy, leading to increased engagement and positive emotions.

(5) Positive Psychology (P.E.R.M.A. Model) [27]: D.D.A.T.A. actively contributes to student well-being by incorporating Seligman's P.E.R.M.A. model elements: Positive Emotions (joy from creative expression), Engagement (flow in individualized activities), Relationships (strengthened by group work), Meaning (derived from personal artistic creation), and Achievement (competence and self-confidence from mastering skills) [16], [17], [18], [19].

2.4. The Role of Arts in Special Education and the S.T.E.A.M. Framework

Arts (music, dance, drama, visual arts) are recognized as essential tools in special education, offering numerous benefits [24]. They contribute to: cognitive development (enhancing attention, memory, problem-solving, sequencing); emotional development (expression, regulation, empathy, anxiety reduction); social development (cooperation, interaction, turn-taking, conflict resolution); motor development (fine and gross motor skills, coordination, balance); communication development (verbal and non-verbal expression); and self-esteem/self-confidence [16]. Arts can be segmented to foster specific skill areas [13], [16]. The pleasure derived from artistic engagement, particularly in music where factors like melody, timbre, tonality, and rhythm are additive and independently utilizable, acts as a powerful motivator for participation and learning [20], [22], [16]. D.D.A.T.A. is inherently integrated into the S.T.E.A.M. (Science, Technology, Engineering, Arts, Mathematics) framework, acknowledging Arts as indispensable for innovative

learning [14]. This combination fosters a multi-sensory and empowering environment, promoting not only academic skills but also crucial 21st-century skills like critical thinking, creativity, collaboration, and communication, vital for student success and societal integration.

3. Methodology

3.1. Research Design

This study employed a mixed-methods approach, combining quantitative skill assessment with qualitative investigation of the intervention experience. This allowed for a comprehensive understanding of D.D.A.T.A.'s effectiveness and the mechanisms of change. Quantitative analysis utilized a pre-post intervention design with three distinct measurements.

3.2. Participants

The quantitative study included 117 students (63% boys, 37% girls) aged 12 to 25 years, diagnosed with Intellectual Disability (ID) and/or Autism Spectrum Disorders (ASD). These participants were drawn from various special educational settings (e.g., special schools, day centers), including a sub-sample of 17 middle school students.

Additionally, qualitative analysis involved clinical observations from approximately 40 cases over five years of visualization and verbalization techniques in special music education [16], and from 100 individuals with ID/ASD. This provided complementary qualitative data on the approach's long-term impact and broader applicability across age ranges and clinical contexts.

3.3. D.D.A.T.A. Intervention Procedure

The D.D.A.T.A. intervention spanned two months, structured into two main phases:

(1) Individual Sessions: The first four 45-minute sessions, conducted twice weekly, focused on individualized assessment and basic skill development through targeted artistic activities, applying ABA principles. An educator/therapist specially trained in D.D.A.T.A. led this phase.

(2) Group Work: Following individual sessions, students were integrated into small groups (4-6 individuals) for the remainder of the two months. Group sessions, 60 minutes long and held twice weekly, emphasized social interaction, collaboration, and applying acquired skills in a dynamic, inclusive setting. The educator/therapist facilitated group dynamics, provided support, and differentiated activities as needed.

3.4. Tools and Data Collection Techniques

The D.D.A.T.A. methodology initially assesses a student's ability to distinguish colors, shapes, and their names [23]. If necessary, instructors first teach these concepts using colored and shaped cards [16].

Quantitative Data: Skill development was quantitatively assessed using an observation / assessment rubric specifically designed for this study. This rubric, based on established ID/ASD skill assessment criteria e.g., ADOS-2, Vineland Adaptive Behavior Scales, [10], included rating scales for: Fine Motor Skills, Gross Motor Skills, Verbal Communication, Non-Verbal Communication, Social Skills, Emotional Expression and Regulation, Creativity and Self-expression, and Overall Engagement. Assessments occurred at three time points during the two-month intervention: Baseline (before individual sessions), after Individual Sessions (4th session), and Final Measurement (after group sessions). To ensure objectivity, assessments were conducted by specially trained therapists and educators not directly involved in the intervention's daily implementation.

Qualitative Data: Qualitative data deepened the understanding of D.D.A.T.A.'s impact through: systematic observations of student behavior (participation, interactions, emotional expression); semi-structured interviews with implementing educators/therapists (observed changes, challenges, successes); and collection/analysis of student artistic work samples (e.g. drawings, constructions, recordings) to assess creativity, self-expression, and motor skill development.

3.5. Specific Visualization and Verbalization Techniques

D.D.A.T.A.'s application in music education employs specific visualization and verbalization techniques:

(1) Rhythm (Tempo):

(a) Visualization Phase [14]: Simple objects (glasses, straws, pens, markers) were used to represent rhythmic values (straw=quarter, pen=half, marker=whole, two straws=eighth note). Students learned to count and clap/play rhythms based on object sequences. Different colors (red/yellow) indicated left/right hand use [16].

(b) Verbalization Phase [14]: Rhythmic concepts were verbalized with sounds like "Doum" (center hit) and "Te" (edge hit) for percussion. Geometric shapes (squares for "Doum," triangles for "Te") supplemented verbalization when needed, fostering improvisation on rhythmic forms [14].

(2) Melody [14]: (a) Verbalization: A Digital Audio Workstation (DAW) sampled male (low octave) and female (high octave) voices singing notes. Students played requested notes by matching keys to corresponding sampled sounds [16]. This cognitive process is illustrated in Figure 2. The process was repeated for note learning across various instruments [16].

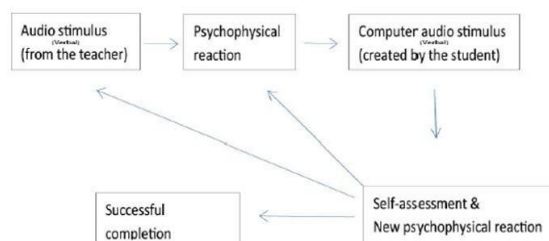


Figure 2. Cognitive process using computer audio stimulus (T.E.L.L. Method)

Visualization: Cardboard cards of varied colors and sizes represented notes and durations (e.g., 4x2 cm for sixteenth, 4x4 cm for eighth, 4x8 cm for quarter, 4x32 cm for whole). Notes were color-coded and corresponding stickers placed on instrument keys. Students performed melodies by observing card sequences [16].

Visualization and Verbalization (Combination): Achieved via DAW's MIDI mapping, allowing simultaneous sampled note listening and verbalization [16].

(3) Chord: Chord visualization was solely via a specific color generated by a visual signal generator [13], [16].

3.6. Material and Data Analysis

Materials for D.D.A.T.A. implementation included: a Digital Audio Workstation (DAW), a "Makris" Pedal Switch Visual Signal Generator, two speakers, cardboard, straws, pens, markers, the musical instrument used, and stickers for instrument key positions [16]. Quantitative analysis used the SPSS statistical package. Paired-samples t-tests compared mean skill scores across three measurement points (Initial vs. Individual, Individual vs. Final), with statistical significance set at $p < 0.05$.

Qualitative analysis applied thematic analysis [4] to observations, interviews, and work samples. This involved coding data, identifying recurring patterns, and generating themes to highlight D.D.A.T.A.'s key experiential aspects and impact.

4. Results

Data analysis revealed statistically significant improvements in all assessed skills during the

D.D.A.T.A. intervention, supporting preliminary findings. Table 1 presents the mean scores (M), standard deviations (SD) for each skill at three measurement points, alongside paired-samples t-test results. As demonstrated, all skills showed highly significant improvement ($p < 0.001$) across each intervention phase.

Table 1. Summary Table of Statistical Results - Skill Development Across Three Measurement Phases (N = 117)

Skills	Initial (M ± SD)	Individual (4th Session) (M ± SD)	Final (Group) (M ± SD)	t (Initial vs Individual)	p (Initial vs Individual)	t (Individual vs Final)	p (Individual vs Final)	Significance
Fine Motor	3.0 ± 1.1	4.5 ± 1.0	7.8 ± 0.9	6.2 ***	<.001	12.5 ***	<.001	***
Gross Motor	3.2 ± 1.0	4.8 ± 1.1	7.5 ± 1.0	6.5 ***	<.001	10.9 ***	<.001	***
Verbal Communication	2.5 ± 0.9	3.8 ± 1.0	6.9 ± 1.1	5.3 ***	<.001	11.8 ***	<.001	***
Non-Verbal Communication	2.8 ± 1.0	4.3 ± 1.0	7.3 ± 0.9	6.0 ***	<.001	12.1 ***	<.001	***
Social	2.3 ± 0.8	3.6 ± 0.9	6.5 ± 1.0	5.5 ***	<.001	11.0 ***	<.001	***
Emotional Expression & Regulation	2.2 ± 0.9	3.5 ± 1.0	6.8 ± 1.1	5.1 ***	<.001	11.5 ***	<.001	***
Creativity & Self-expression	2.9 ± 1.0	4.2 ± 1.0	7.5 ± 0.9	5.8 ***	<.001	12.0 ***	<.001	***
Overall Engagement	3.5 ± 1.0	5.0 ± 0.9	8.0 ± 0.8	6.3 ***	<.001	12.8 ***	<.001	***

Significant improvements were observed in targeted skill areas:

(a) Cognitive Skills (through visualization / verbalization): Clinical application of visualization and verbalization techniques, particularly in rhythm and melody instruction, significantly improved cognitive skills like memory, perception, and attention. Trainees enhanced their ability to distinguish, group, and verbally express colors and shapes, fundamental for musical concept learning [16].

(b) Fine and Gross Motor Skills: Notable improvements were seen in both fine (e.g., manipulating instruments like percussion with "Doom" and "Te" strikes) and gross motor skills, including hand-eye coordination and accuracy in artistic creation [16].

(c) Communication Skills: Significant increases were observed in both verbal (e.g., using more words/phrases for requests and comments) and non-verbal communication (e.g., improved eye contact and gesture use), especially during musical concept verbalization and visualization [16].

(d) Social Skills: Enhanced participation in group activities, improved cooperation, and better turn-taking were evident, notably highlighted during a recent inclusive Christmas concert (2024) held by the special school of Peristerite.

(e) Emotional Expression and Regulation: Students showed improved emotion recognition and expression, with fewer distress outbursts and increased self-regulation, aligning with the P.E.R.M.A. model [16].

(f) Enhanced Creativity and Self-expression: Significant improvements in creativity and self-expression were supported by qualitative data from student work samples (increased originality, diverse material use, personal style development) and educator interviews (increased experimentation, willingness to express thoughts/feelings through arts), directly linked to enhanced self-confidence [16].

(g) Overall Engagement in Learning: This skill showed marked improvement, with observations indicating longer attention spans, increased interest, and spontaneous initiative during lessons. Strong interest in participation was a key finding from clinical application [16].

(h) Promotion of a More Inclusive Environment: Qualitative data indicates D.D.A.T.A. significantly fostered inclusivity. A joint music event within the European Program "Cap sur l'Ecole Inclusive," involving eighteen musician-trainees with ID and a twenty-member choir from a regular school, showcased the trainees' professionalism and yielded significant benefits for the school choir members [16]. Despite the overall positive outcomes, a small percentage of students (approximately 5% or 6 out of 117) showed limited improvement. This underscores the need for further research and individualized interventions for cases with more complex needs or severe intellectual disabilities.

5. Discussion

This study's results strongly affirm the Differentiated Didactic Approach to Teaching the Arts (D.D.A.T.A.) effectiveness in developing diverse skills in students with intellectual disabilities (ID) and Autism Spectrum Disorders (ASD). The statistically significant improvements across all assessed areas (cognitive, motor, communicative, social, emotional, creativity, and overall engagement) highlight the approach's holistic nature.

D.D.A.T.A.'s success stems from integrating its visualization and verbalization techniques. These enrich Working Memory with the crucial Motor Model, as per the revised Cognitive Theory of Multimedia Learning by Ioannis Makris [16]. The combination of specific tactile and visual stimuli (e.g., glasses/straws for rhythm, cardboard cards for notes) with verbalization (e.g., "Doom," "Te," sampled voices) creates a multi-sensory environment. This facilitates information intake and processing for students with cognitive particularities. The Motor

Model's emphasis on repetition and practice further enhances tactile sensory stimuli, supporting movement and muscle strengthening vital for optimal performance [16].

The efficacy of D.D.A.T.A. is also underpinned by its theoretical foundations:

(a) Applied Behavior Analysis (ABA): Provides a framework for systematic teaching, task breakdown, and skill shaping, maximizing success opportunities.

(b) Information Integration Theory (ITT) by Norman H. Anderson: Supports the provision of multiple, complementary stimuli for effective information integration.

(c) Reversal Theory: Explains high student engagement by addressing changing metamotivational states, fostering pleasure and motivation.

(d) P.E.R.M.A. Model of Positive Psychology: Highlights D.D.A.T.A.'s holistic impact on well-being through positive emotions, engagement, relationships, meaning, and accomplishment. The integration of arts within D.D.A.T.A. is crucial. Arts' ability to be broken into discrete components allows for targeted skill development [13], [16]. The inherent pleasure from artistic engagement, especially music with its independently utilizable factors (melody, timbre, tonality, rhythm), powerfully motivates participation and learning [21], [22], [16].

The Inclusive concerts held by the special School of Peristeri exemplifies D.D.A.T.A.'s capacity to promote inclusion and showcase the musical potential and professionalism of individuals with ID, while benefiting their non-disabled peers. This underscores specialized educational practices' role in fostering understanding and appreciation across student populations. Despite encouraging results, this study has limitations. The quantitative analysis's sample size ($N=117$) is adequate, but the absence of a control group limits attributing improvements solely to D.D.A.T.A. The reliance on an observation rubric, while systematic, may introduce subjectivity. Additionally, the small number of students (6 out of 117) showing limited improvement highlights the need for further research into factors influencing treatment response and the development of more individualized strategies for complex needs.

6. Conclusions

The Differentiated Didactic Approach to Teaching the Arts (D.D.A.T.A.) represents a significant advancement in special education, offering a holistic and effective methodology for students with intellectual disabilities and Autism Spectrum Disorders. This study conclusively demonstrates

D.D.A.T.A.'s capacity to foster substantial improvements across a broad spectrum of skills: cognitive, motor, communicative, social, emotional, creativity, and overall engagement—through its innovative visualization and verbalization techniques, underpinned by robust theoretical frameworks. Beyond individual skill development, D.D.A.T.A. cultivates inclusive learning environments that empower students and enhance their participation and well-being. The documented successes, particularly in collaborative settings like the Inclusive concerts, highlight the profound potential of arts-based interventions to bridge educational gaps and promote societal integration. While acknowledging limitations, the overwhelmingly positive outcomes underscore D.D.A.T.A.'s value as a transformative pedagogical tool.

7. References

- [1] Anderson, N. H. (1981). *Foundations of information integration theory*. Academic Press.
- [2] Anderson, N. H. (1991). *Contributions to information integration theory*. Lawrence Erlbaum Associates.
- [3] Apter, M. J. (1989). *Reversal theory: Motivation, emotion and personality*. Routledge.
- [4] Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- [5] Cooper, J. O., Heron, T. E., and Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Pearson Prentice Hall.
- [6] Heward, W. L. (2011). *Exceptional children: An introduction to special education*. Pearson Education.
- [7] Krumboltz, J. D., and Krumboltz, H. B. (1972). *Changing children's behavior*. Prentice-Hall.
- [8] Lord, C., Charman, T., Havitt, E., Howlin, P., Pickles, A., Sallows, G., ... and Gotham, K. (2020). The Simons Simplex Collection: A resource for research into autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 50, 1740-1755.
- [9] Macri, L., and Makris, I. (2014a). Education of children with special needs through music in Greece. In C. X. Li (Ed.), *Culture in education: A collection of articles on Chinese and international culture and education* (pp. 209-216). Yunnan People's Publishing House.
- [10] Macri, L., and Makris, I. (2014b). The development of the "Visual Signal Generator" for individuals with intellectual disabilities. *International Journal of Education and Research*, 2(10), 45-56.
- [11] Macri, L., and Makris, I. (2014c). *Music and visual signal generator: Methodological approaches for individuals with intellectual disability*. Self-published.
- [12] Makris, I. (2005). *Music and autism: The Makris Method*. Self-published.
- [13] Makris, I. (2015). *Music and visual pedagogy: New educational approaches for children and adults with intellectual disabilities*. CreateSpace Independent Publishing Platform.
- [14] Makris, I. (2016a). *New visual and verbal techniques for teaching music to children with intellectual disabilities*. CreateSpace Independent Publishing Platform.
- [15] Makris, I. (2016b). *Teaching music to children with intellectual disability: A practical guide for parents and educators*. CreateSpace Independent Publishing Platform.
- [16] Makris, I. (2018). *The D.D.A.T.A. Method: A differentiated didactic approach to teaching the arts to individuals with intellectual disabilities*. CreateSpace Independent Publishing Platform.
- [17] Makris, I. (2020). The differentiated didactic approach to teaching the arts (DDATA) in special education and positive psychology. *Psychology*, 11(04), 551.
- [18] Makris, I. (2022). Differentiated didactic approach to teaching the arts (D.D.A.T.A.). Στο C. A. Shoniregun (Πρόλογος), *Learning and teaching methodologies volume 1*. Accolade Digital Publishing.
- [19] Makris, I. (2024). STEAM and DDATA: A multisensory approach to music and development of skills among students with intellectual disabilities, *Education and New Developments*, 189-193.
- [20] Makris, I., and Macri, L. (2003). *Music and children with mental retardation*. Self-published.
- [21] Makris, I., and Mullet, E. (2003). Music preferences: A comparison of three methods. *Musicae Scientiae*, 7(1), 77-94.
- [22] Makris, I., Mullet, E., and et al. (2012). Structure of pleasure in music perception and practice in typically developed individuals and individuals with intellectual disabilities. *Musicae Scientiae*, 16(2), 227-245.
- [23] Makris, I., Vilkelienė, G., and et al. (2017). The T.E.L.L. Through Music Method: A European approach for teaching music to individuals with intellectual disabilities. *European Journal of Special Needs Education*, 32(4), 519-532.
- [24] Malchiodi, C. A. (2012). *Handbook of art therapy*. Guilford Press.
- [25] Mayer, R. E. (2008). *Multimedia learning* (2nd ed.). Cambridge University Press.
- [26] Schalock, R. L., Borthwick-Duffy, S. A., Bradley, V. J., Buntinx, W. H. E., Coulter, D. L., Craig, E. M., ... and Yeager, M. H. (2010). *Intellectual disability: Definition, classification, and systems of supports* (11th ed.). American Association on Intellectual and Developmental Disabilities.

[27] Seligman, M. E. P. (2011). Flourish: A visionary new understanding of happiness and well-being. Free Press.

Speaker 7

Speaker 8: Alysha Collins



Dr Alysha Collins is an internationally certified school psychologist and certified administrator who enjoys supporting students, teachers, and their families. Collins obtained her doctorate in educational leadership and an educational specialist degree in school psychology. Her dissertation included studying hope during COVID-19. She is an adjunct instructor of university psychology classes and consults with organizations on mental health, educational research, and well-being in the workplace. Collins likes to apply her experience in trauma, hope, positive psychology, mindfulness, and inclusion to all her work. She is based in London, England.

A Future of Hope: Ensuring Educator Retention

Alysha Collins
Educational Psychologist
Southend Council
UK

Abstract

Are teachers really leaving the profession? Globally, UNESCO has reported that teacher turnover has doubled from 4.62% in 2015 to 9.06% in 2022 and that number is just getting worse. New data according to The Global Report on Teachers shows that 44 million additional teachers will be required to achieve universal primary and secondary education by 2030. These are overwhelming statistics, but there are individual and systemic hope strategies that can create influence throughout an educational setting. Learn facts, engage with others, and leave with meaningful action steps! Here's to hope for a new and exciting future.

Speaker 8

Speaker 9: David Lazaro



David Lazaro is the Director of Edraak, an initiative of Her Majesty Queen Rania of Jordan, launched in 2014 and supporting over 10 million learners across the Middle East and Arabic-speaking world. David leads education reform and workforce development projects, as well as humanitarian interventions supporting educational needs for youth across the MENA region from Amman, Jordan.

Empowering Futures: Cultivating Lifelong Learning in Arab Youth

David Lazaro
Queen Rania Foundation
Jordan

Abstract

The proposed presentation highlights strategies success Edraak has faced advancing a culture of continuing education and professional development among Arab youth, emphasizing its role in enabling long-term economic resilience, personal fulfillment, and social advancement, as the region faces numerous security challenges. It begins by examining current educational gaps and the mismatch between traditional learning models and the fast-evolving demands of the labor market and the “overpromise” offered at times of online education. The presentation highlights the value and social change observed among youth of a mindset of lifelong learning from school education through to adulthood, integrating both formal and non-formal learning opportunities.

Key enablers such as access to digital platforms, updated and revised content reflective of the economic outlook, and youth-centered policies are discussed as critical drivers of sustained engagement in professional growth and viable economic opportunity. The presentation additionally includes notable examples from Edraak’s interventions in Jordan, Syria, Palestine, Morocco and others.

Speaker 9

Speaker 11: Samar Ahmad ElFarra



Samar is a Canadian higher education leader, academic, and healthcare professional with expertise in institutional effectiveness, accreditation, curriculum development, and program assessment. As Director of Institutional Effectiveness and Accreditation at the University of Al Dhaid, UAE, she drives policy design, assessment frameworks, and quality assurance processes, fostering evidence-based decision-making to enhance academic quality and operational excellence.

She has served as a faculty member and program chair in Medical Imaging, developing innovative curricula, securing accreditations, and advancing student-centered learning. An active voice in the medical imaging community, Samar is Vice President of the Radiographers Society of Emirates, a Senior Fellow of the UK's Advance HE, and a published researcher with contributions to international journals and conferences.

Bridging the Gap: Elevating Higher Education Assessment through Professional Standards

Samar El-Farra
University of AlDhaid
United Arab Emirates

Abstract

Entrustable Professional Activities (EPAs) offer a transformative framework for assessing competency across diverse fields in higher education. Originally developed for medical education, EPAs have broad applicability in evaluating students' readiness to perform essential, real-world tasks independently in their respective disciplines. Whether in healthcare, engineering, education, business, or the arts, EPAs provide a practical approach to bridging the gap between academic learning and professional practice. This presentation explores how EPAs can be integrated into curricula across domains to assess learners' progression toward competency. EPAs are defined as critical, observable tasks that professionals must execute with minimal supervision. By identifying domain-specific EPAs and aligning them with program outcomes and professional standards, institutions can create assessment strategies that ensure graduates are prepared for the complexities of their industries.

Key discussion points include designing EPA-based assessment tools, mapping EPAs to industry-specific competencies, and making entrustment decisions that reflect learners' readiness for autonomous practice. Examples from various disciplines will illustrate how EPAs can evaluate skills such as project management in engineering, classroom teaching in education, client relationship management in business, and creative problem-solving in the arts.

By embedding EPAs into assessment strategies, higher education institutions can foster a competency-based approach that emphasizes accountability, transferable skills, and readiness for professional roles. This not only enhances the validity and reliability of assessments but also strengthens the connection between academic achievement and real-world performance, empowering graduates to excel in their chosen professions.

Speaker 10

Speaker 12: Vamsidhar Yerramsetty



Dr Vamsidhar Yerramsetty is a tri-sector education leader with experience across nonprofits, government, and the private sector passionate about improving student learning outcomes especially for disadvantaged students. He has been a teacher, strategy consultant, and ed-tech entrepreneur. Dr Yerramsetty leads cross-functional teams to design user-centric ed-tech products, applying advanced psychometric methods and AI to personalise learning. The highlights include product leadership building unified data systems, open-data portals, adaptive assessments with an impact reaching over 10,000+ schools over time. He won Innovate UK grant to build adaptive assessments. Educated at the University of Oxford's Educational Assessment program and a PhD from University of Minnesota.

Modeling Multidimensional Competence: An LCDM-Based Evaluation of 21st-Century Skills

Vamsidhar Yerramsetty
Bloom Learning
UK

Abstract

Automation and AI are compelling humankind to relook at traditional curriculum knowledge attainment and underscoring the need for complex problem-solving, collaboration, creativity, and emotional intelligence. Schools therefore must both teach and measure these 21st-century skills. Traditional tests, however, collapse attainment into a single score and cannot differentiate the many overlapping abilities that modern tasks demand. Cognitive-diagnostic models (CDMs) solve this problem: they link every item to the discrete attributes it requires and then estimate each learner's mastery profile, yielding fine-grained, actionable feedback. Thus, by combining 21st-century skill assessments with the diagnostic precision of CDMs, educators gain actionable insights to effectively integrate these essential competencies into classroom teaching and learning.

This study used a Log-Linear CDM (generalized DINA with logit link) to analyse a 35-item 21st-century skills assessment completed by 583 Grade 7-8 gifted Indian students. Thirteen attributes were mapped in the Q-matrix, ranging from digital literacy and verbal intelligence to social-emotional and creative competencies. Model fit was strong (RMSEA 0.014–0.152; $M = 0.091$), confirming that the LCDM could reproduce most observed response patterns while simultaneously estimating multiple skill effects per item.

Mastery estimates varied widely: Cognitive skills (~80 %) and disciplinary knowledge (~83 %) were strengths, whereas digital skills (~37 %) and logical reasoning (~39 %) lagged. High polychoric correlations (e.g., 0.94 between problem solving and verbal intelligence) highlighted foundational or redundant constructs, while negative/weak links (e.g., sequential thinking vs. social-emotional skills) exposed distinct cognitive strands. CDM output also flagged items with potential misfit, guiding future revision.

By revealing who possesses which competencies and how those competencies interact, CDMs provide a diagnostic lens that single-score tests lack. These insights enable educators and test designers to target instruction, streamline frameworks, and refine items, ensuring that assessments truly capture the multidimensional skill set learners need in a rapidly changing world.

Speaker 11

Speaker 13: Judith Hanke



Judith Hanke has been a scientific associate at Europa-Universität Flensburg since 2021. Her research focuses on advancing the integration of digital tools in education, including digital diagnostics, formative assessment, learning analytics, and adaptive learning, with the goal of creating more personalized and data-informed learning environments. She also explores primary education through inclusive teaching, reading support, and the development of effective reading strategies.

Enhancing Educational Insights and Engagement through Learning Analytics: Exploring Digital Reading Packages

Judith Hanke
University of Flensburg
Germany

Abstract

Learning analytics (Heilala, 2022; Hernández-de-Menéndez et al., 2022; Kollom et al., 2021) is an evolving interdisciplinary field that leverages data-driven methodologies, computational processes, and automation to optimize educational outcomes. It operates through an iterative cycle involving data collection, measurement, analysis, and targeted intervention to improve learning experiences. Grounded in data science and artificial intelligence, it encompasses four primary types: descriptive, diagnostic, predictive, and prescriptive analytics. These analytics help uncover patterns, forecast trends, and provide strategic feedback, driving personalized learning, curriculum refinement, and early intervention for students facing academic challenges.

This presentation explores the role of learning analytics through a practical case study involving digital reading packages (Hanke & Diehl, 2024). These packages function both as tools for formative assessment (Black & Wiliam, 1998; OECD, 2008; Wylie, 2023) and as sources of valuable learning analytics data. Designed to support reading comprehension, each package includes literary texts and exercises tailored to the learner's proficiency level. Data on student interaction, such as task completion times and button usage (Nokelainen, 2006), provides diagnostic insights into learning behaviors and informs refinements to educational tools.

Results from an intervention study demonstrated that targeted adjustments improved platform efficiency and reduced disruptions. The presentation highlights how these enhancements, together with an analysis of student performance data, can improve educational platforms and support more effective student engagement and skill development.

Speaker 12

Speaker 14: Suzana Ursi



Dr Suzana Ursi is an Associate Professor in the Department of Botany at the University of São Paulo (USP), Brazil, and Director of USP's Science and Technology Park (CienTec). She holds a Ph.D. in Botany from USP, with part of her doctoral research conducted at Stockholm University and Uppsala University, Sweden. Her research focuses on teacher education, teachers' professional knowledge, environmental perception, and strategies to promote plant awareness. She coordinates the research group BotEd - Botany in Education, dedicated to contextualized teaching and teacher professional development, and supervises graduate students in Biology Education.

Promoting Scientific and Environmental Literacy through Non-Formal Education: The Case of the CienTec Park at the University of São Paulo

Suzana Ursi

Universidade de São Paulo, Brazil

Abstract

This paper presents the educational strategies implemented by the Science and Technology Park CienTec, in São Paulo, Brazil. As a non-formal educational space open to the public since 2002, CienTec promotes scientific and environmental literacy through diverse activities, exhibitions, and nature-based experiences. After the pandemic closure, the Park resumed operations with a renewed focus on inclusive science communication, environmental awareness, and community engagement. Undergraduate students act as mediators in interactive educational actions. New initiatives such as citizen science, a native plant trail, and a community garden were implemented. CienTec's practices align with the Sustainable Development Goals and Brazil's national curriculum, reinforcing its social relevance. The experience highlights the role of non-formal education in fostering a scientifically literate and environmentally conscious society, particularly in biodiversity-rich contexts like Brazil.

Keywords: CienTec Park; community engagement; environmental education; PCK; science communication; sustainable development.

1. Introduction

The Science and Technology Park (CienTec) was officially established in 2001 as an outreach unit of the University of São Paulo (USP). Located within the State Park of the Ipiranga Springs (PEFI), a conservation area of Atlantic Forest, CienTec encompasses 141 hectares with significant environmental and historical value (Figure 1). Numerous springs of the Ipiranga Creek, where Brazil's independence was proclaimed, are found within its grounds. CienTec offers both indoor attractions (e.g., Planetarium, Microscopy Lab) and outdoor installations (e.g., Physics Playground, Solar System Walk). It also includes a man-made lake and educational trails that enable immersive experiences with native plants such as pau-brasil and Araucaria, as well as local fauna like howler monkeys and stingless bees, fostering scientific learning for all visitors [1].

Many of its buildings, dating from the 1930s and 1940s, are officially listed as historic landmarks [2] and are currently used for exhibitions and interactive educational activities.

Several educational and environmental conservation initiatives (including dialogues with the surrounding community) have been successfully implemented at the Park, always embedded in a process of ongoing evaluation and improvement. Currently, the majority of visitors during the week are school groups, while weekends and holidays attract a broader and more spontaneous audience. According to visitor feedback, one of CienTec's most distinctive features is the work of its undergraduate mediators from USP, who bring conceptual depth, enthusiasm, and care to the educational experience. The opportunity to engage with green spaces is another major highlight, as is the wide variety of activities available to visitors of all ages.

This paper aims to report and reflect on the educational strategies adopted by CienTec in the post-pandemic period, with a focus on promoting scientific and environmental literacy among its diverse audience.



Figure 1. Aerial image capturing sections of CienTec Park, the State Park of the Ipiranga Springs, and surrounding urban zones of São Paulo

2. Mission, Vision and Social Relevance

We live in a challenging world defined by volatility, uncertainty, complexity, and ambiguity, as

summarized by the acronym VUCA [3]. These transformations affect every sphere of life, from education to the environment. In this context, promoting scientific literacy becomes essential to ensure that citizens understand scientific concepts, the process of scientific knowledge production, and its social implications [4]. In Brazil, this goal is supported by national curricular guidelines, and it is crucial in addressing the country's severe environmental crisis.

Educational practices, including those in non-formal settings, must adapt to this complex reality to foster broader and more socially connected learning. CienTec's mission aligns with this goal by promoting scientific and environmental literacy through community-oriented and dialogic extension activities.

Our Vision: To be a national and international reference in science communication and non-formal education, promoting citizen engagement, environmental awareness, and the development of human resources.

Our Mission: To disseminate science and technology to society through accessible, engaging, and dialogical approaches; to spark young people's interest in scientific fields; to promote environmental consciousness and life quality; and to contribute to formal education by supporting both students and teacher training.

3. Activities

3.1. Reconstruction process: from total closure to expanded activities and audiences

The COVID-19 pandemic deeply impacted CienTec, leading to a complete shutdown and accelerated physical deterioration. The Park only reopened in November 2021, initially offering self-guided tours using QR codes, without group activities or mediation. The year 2022 marked a process of reconstruction, including rebuilding interpersonal dynamics after a period of isolation that affected social and mental well-being. School group visits only resumed fully in the second half of the year, and by 2023 nearly all pre-pandemic activities had returned. This recovery is reflected in the increase in the number of visitors.

Table 1. Number of visitors by year

Year	2019	2020	2021	2022	2023	2024
Visitors	21.819	472	567	18.570	24.926	22.441

An important factor in expanding activities was the integration of new cultural and extension projects, which allowed for more scholarship and internship positions. Despite structured training for undergraduate mediators, further investment is needed. The Pedagogical Content Knowledge (PCK) framework was recently introduced at CienTec to support contextualized, visitor-focused mediation [5]. Research efforts are also underway to evaluate these practices and contribute to the field of non-formal science education [6].

3.2. Regular educational activities

CienTec offers a wide variety of educational activities with distinct themes and interactive dynamics. Highlights include:

- Educational Trails: immersive walks in Atlantic Forest habitat, with highlights such as the Park's Lake (see Figure 2).



Figure 2. View of the lake at CienTec Park

- Bee Walk and Pollination Exhibition: educational trail featuring native stingless bees, hives, and pollination ecology.
- Historical Tour: visiting heritage-listed buildings, including the Zeiss telescope and stained glass of Urania, goddess of astronomy (see Figure 3).



Figure 3. Stained glass of Urania located at CienTec Park

- Planetarium Sessions: immersive audiovisual astronomy experiences with sky simulations.

- Soils Activity: in partnership with the “Soils at School” project, addressing soil properties and conservation.
- Geophysics Space: exploring Earth’s interior, earthquakes, tectonics, and the popular earthquake simulation.
- Mathematics Space: hands-on activities presenting key math concepts in accessible and fun ways.
- Physics in Daily Life: interactive physics exhibits such as the Van de Graaff generator, dynamo bike, granite sphere, human gyroscope, and solidarity seesaw.
- Nighttime Astronomy Events: monthly night tours combining heritage visits, planetarium sessions, and telescope sky viewing (see Figure 4).



Figure 4. Young visitor observing through the telescope during CienTec Park’s nighttime astronomy session

- Microscopy Lab: visitors explore a variety of microorganisms.
- Solar System Walk: an outdoor installation with artistic representations and scientific facts about the Sun and its planets, including Earth (see Figure 5).



Figure 5. Partial view of the Solar System Walk installation at CienTec Park

3.3. Temporary exhibitions

Among the exhibitions hosted by CienTec after the pandemic, those exploring the dialogue between Science and Art were especially impactful.

- “Threads of Life – Art and Botany” (2022–2023): Curated by CienTec’s director, Suzana Ursi, and ceramic artist Débora Amaral, this exhibition was

inaugurated during the University of São Paulo Arts and Culture Week. It featured ceramic artworks combined with natural elements such as vines and repurposed tree trunks. Visitors could also observe vine anatomy under a microscope. The exhibition fostered learning through artistic engagement and inspired both national and international presentations and publications [7][8].

- “Biomes of Brazil” (since 2023): Created by artist Ju Barros, this exhibition presents six glass-painted mosaics celebrating the biodiversity of Brazilian biomes. Each piece highlights native species at risk of extinction and women from local communities. Augmented reality enhances visitor interaction with animated elements. An extension activity on native Brazilian plants complements the visit and addresses the phenomenon of ‘plant invisibility’—the common failure to perceive plants in everyday life [9].

3.4. Projects

CienTec hosts several ongoing projects. Two of them are outlined below.

- “Community Vegetable Garden” This project is a partnership with the local residents’ association “Novos Passos da Ilha - Vila Fachini”, whose neighborhood borders the Park. It represents a case of critical and participatory environmental education [10], transforming a former dumping site into a shared vegetable garden. The garden fosters knowledge exchange between academia and community and promotes social–environmental sustainability. In 2025, cultivation is focused on Non-Conventional Edible Plants, such as *Stachys byzantina* (lamb’s ear) and *Pereskia aculeata* (ora-pro-nóbis), as well as culturally important plants like *Lagenaria siceraria* (calabash), used to produce bowls and musical instruments.
- “Insects in Schools”: This initiative, in collaboration with the Zoology Department at University of São Paulo, provides entomological boxes for loan to schools and use in CienTec activities. It supports learning about insect biodiversity and classification, enhancing science education in line with recent pedagogical recommendations.

4. Impact and Evaluation

4.1. Internal impact

CienTec’s academic impact is evident in the training of undergraduate students who act as mediators, as well as in the educational experiences of school visitors. Student development is progressive, beginning with scholarship holders and advancing to internships. The Park team conducts

continuous training, including sessions led by external experts, on topics such as inclusive education (e.g., support for autistic visitors), non-formal science communication, and specific scientific content.

4.2. External impact

CienTec plays a key role in science and environmental outreach, serving both school and general audiences. Visitors frequently report transformative learning experiences. The Park fosters critical thinking on Science, Technology, Society, and Environment issues, promoting inclusive, anti-racist, and environmentally conscious perspectives. Examples include regular educational trails, citizen science projects such as fauna monitoring, and the community garden. These initiatives also enable knowledge exchange between communities and university staff. CienTec's social media platforms support program promotion and function as important tools for science communication.

4.3. Indicators

A key indicator is the number of visitors, which has surpassed pre-pandemic levels (see Table 1 and Figure 6). However, CienTec also emphasizes qualitative evaluation. Group visit coordinators complete feedback forms, while spontaneous visitors provide informal feedback through social media, email, or phone. Positive feedback highlights the mediators' conceptual rigor, enthusiasm, adaptability to age groups, and the opportunity to engage with green spaces. The Planetarium sessions are frequently praised. Areas identified for improvement include infrastructure conditions and historical building maintenance.



Figure 6. Figure 6. Spontaneous visitors during a special event (opening of the meliponary at CienTec Park)

5. Conclusion

CienTec Park has overcome significant post-pandemic challenges, transitioning from full closure and infrastructure degradation to full activity

restoration and increased public engagement. With this recovery phase complete, the Park now focuses on strengthening its infrastructure and advancing innovative educational, inclusive, and sustainable initiatives. Future directions emphasize dialogical models of science communication and non-formal education that move beyond knowledge transfer and toward social engagement. CienTec Park holds vast potential, and the reflections presented here aim to support ongoing and future planning and action.

6. References

- [1] Ursi, S., Barbosa, P.P., Sano, P.T., and Berchez, F.A.S. (2018). Ensino de Botânica: conhecimento e encantamento na educação científica. *Estudos Avançados*, vol. 32, no. 94, pp. 7–24.
- [2] Mantovani, M.S.M., and Massambani, O. (2024). Parque Cientec: Parque de Ciência e Tecnologia da USP. São Paulo: Edusp.
- [3] Taskan, B. et al. (2022). Clarifying the conceptual map of VUCA: A systematic review. *International Journal of Organizational Analysis*, vol. 30, no. 7, pp. 196–217, 2022.
- [4] Hurd, P.D. (1998). Scientific Literacy: New Minds for a Changing World. *Science Education*, vol. 82, no. 3, pp. 407–416.
- [5] Busch, K.C., Kudumu, M., and Park, S. (2023). Pedagogical content knowledge for informal science educators: Development of the ISE-PCK framework. *Research in Science Education*, vol. 53, no. 2, pp. 253–274.
- [6] Ursi, S., and Gentile, C. (2024). Abordagem da Botânica em Espaços de Educação não-formal: Conhecimento Pedagógico do Conteúdo de mediadores do Parque CienTec. IX ENEBIO, Belo Horizonte, 2024 (in press).
- [7] Ursi, S., Amaral, D., and Cesquim, T. (2023). Exposição 'Tramas da Vida - Arte e Botânica': encantamento e conhecimento no Parque CienTec da USP. In *E a vida continua: vazios e esperanças no Ensino de Ciências e de Biologia*, São Paulo: Pimenta Cultural, pp. 263–272.
- [8] Marinho, L.C. et al. (2023). Brazilian Botanists Flirting with Arts: Valuing the Multicultural Heritage. *American Journal of Botany*, vol. 69, pp. 182–185.
- [9] Wandersee, J.H., and Schussler, E.E. (2001). Toward a theory of plant blindness. *Plant Science Bulletin*, vol. 47, pp. 2–9, 2001.

[10] Silva, R.L.F., and Campina, N.N. (2011). “Concepções de educação ambiental na mídia e em práticas escolares: contribuições de uma tipologia,” *Pesquisa em Educação Ambiental*, vol. 6, no. 1, pp. 29–46.

Acknowledgements

Activities and outreach actions were funded by the Dean’s Office for Culture and Extension of the University of São Paulo (Pró-Reitoria de Cultura e Extensão Universitária).

Speaker 13

Speaker 15: Festus Buetey Longmatey



With over 15 years of experience in program management and administration, Festus Buetey Longmatey is a seasoned Programs Director skilled in handling administrative, operational, and outreach functions. His expertise includes personnel management, financial reporting, and project coordination. Festus holds a Bachelor of Arts in Statistics and Economics from the University of Ghana and is currently pursuing an MBA with certifications in Six Sigma, Project Management, and Digital Marketing from the IIBM Institute of Business Management, India. He has led strategic initiatives as the Head of Programs at the Ghana National Education Campaign Coalition, enhancing program offerings and compliance with donor standards. Festus's role on various national committees, including the National Child Protection Committee and the National Steering Committee for Inclusive Education, has involved developing programs for children's well-being and evaluating policy initiatives. He has also managed projects for Education Out Loud, ensuring timely implementation and fund management. Festus Buetey Longmatey's certifications include Transform Social Protection Leadership and Advocacy and Policy Influencing.

The State of Foundational Learning in Ghana: Gender Equality, Quality of Teaching, Accessibility, and Relevance

Festus Buetey Longmatey
*Ghana National Education Campaign Coalition
Ghana*

Abstract

The study investigates the effectiveness of Ghana's education system in supporting foundational learning, which includes basic literacy, numeracy, and socio-economic skills essential for future learning. These skills are crucial for children's success and encompass early literacy, mathematical understanding, critical thinking, problem-solving, and collaborative abilities. The report aligns with NEST's global vision of integrating academic and transversal skills, such as socio-emotional learning and creative thinking, into education.

The research focuses on lower basic schools in Ghana, examining how the education system creates meaningful learning opportunities that enable students to understand and apply knowledge in real-life contexts. These opportunities involve collaborative learning, problem-solving tasks, resilience-building activities, and practical application of literacy and numeracy, aiming to make learning relevant and engaging.

Guided by NEST's frameworks of the 3Cs (Commitment, Capacity, Cohesion) and the 4Ps (Purpose, Pedagogy, Positioning, Power), the study finds that while there is policy-level commitment to foundational learning, implementation gaps exist due to uneven capacity and fragmented coordination. Pedagogical practices often rely on rote learning, hindering active engagement and creativity. Curriculum reforms and teacher deployment frequently do not align with community realities, and power asymmetries limit local input into reform processes.

The study, employing a mixed-methods approach, involved 317 participants across eight regions and sixteen MMDAs, reflecting Ghana's geographic, socio-economic, and educational diversity. It addressed five key questions, including barriers to gender equality, stakeholder perceptions of teaching quality, factors affecting foundational literacy and 21st-century skills, policy impacts on learning accessibility and relevance, and the system's effectiveness in helping children learn what matters.

Key findings reveal significant barriers to gender equality, inadequate teaching resources, and inconsistencies in policy implementation. Recommendations include enhancing community awareness, providing scholarships for girls, targeted teacher training, clear policy guidance on teacher deployment, and curriculum reforms. The study concludes that comprehensive policy reforms and increased investment are crucial for transforming Ghana's education system at the foundational level to better prepare learners for future success.

Workshop

Workshop Organiser: Shari Collins



Shari Collins is a leading consultant and speaker specializing in creating thriving workplace cultures. With global experience spanning education, training, HR, and organizational development, Shari empowers workplaces and educational institutions to foster positive change. Recognized as "The Most Admired Leader to Watch in 2024" by Insights Success Magazine, Shari serves on the Advisory Board for Harvard Business Review, the Editorial Board for Infonomics Society, and as Diversity Director for HR-Nebraska State SHRM Council. With recent certifications in Psychological Safety, Ageism, and Organizational Happiness, Shari is at the forefront of applying various research-based strategies for modern workplace challenges. She recently advanced her expertise through the Chief Well-Being Officer Certification with the World Happiness Academy. She facilitates professional development for educators across the USA through the National Education Association and is a published author in the Parent Ready! book series. She also facilitates national workplace conversations for Open Sesame's Workplace Culture Club. On a personal note, Shari loves reading and learning, caring for her plants, and spending time with her family and their 2 rescue Yorkie sisters. Connect with Shari to explore how she can help your organization cultivate a culture of belonging, happiness, and well-being.

Workshop: Brave and Bold: Fostering Psychological Safety

Scope: Creating an environment where employees feel safe to express themselves, share ideas, and take risks without fear of retaliation is essential to fostering innovation, collaboration, and overall well-being. This workshop is designed to introduce attendees to the concept of psychological safety - what it is, why it matters, and how it can be cultivated in the workplace or educational system. Participants will explore practical strategies to build trust, encourage open communication, and promote a culture where team members feel respected and empowered to contribute without fear of negative consequences. This session will provide attendees with actionable tools to foster a supportive work environment, contributing to a more positive organizational culture. Regardless of your role in education, the workshop provides valuable insights on how to ensure team members feel safe, heard, and valued in their roles. You won't want to miss it!

Key Takeaways:

- Understanding the concept of psychological safety and its impact on employee performance, well-being, and physical safety
- Practical methods to encourage open communication and innovative decision-making
- Techniques for building trust, fostering inclusivity, and promoting a culture of mutual respect
- Connections to the innovation threshold are included

Organiser: Shari Collins
Shari Collins Consulting, USA

Sessions

Session 1: Education and Digital Society

Title: Development of an Interactive Mobile Application To Support Environmental Education For Young Children

Authors: Yeseo Kim, Hae-Deok Song, Ah-Jeong Hong

Title: Applying Active Learning to a Seminar on the Basics of Quantum Computing

Authors: Pedro J. Roig, Salvador Alcaraz, Katja Gilly, Cristina Bernad, Carlos Juiz

Title: Minimalist Approach: Analysis of “Enigmatic Ruins” Video Game Developed by Game Design Students

Authors: Oluwarotimi Abayomi Randle, David Ajayi

Title: Enhancing Project-Based Learning in Engineering Education with Expert-Guided Retrieval-Augmented Generation (RAG)-Based Chatbot

Authors: Y.W. Kiang, A.T.L. Lee, V.W.L. Tam, K.H. Lam, V.C.S. Lee, S.M. Yiu

Development of an Interactive Mobile Application To Support Environmental Education For Young Children

Yeseo Kim, Hae-Deok Song, Ah-Jeong Hong
Chung-Ang University
South Korea

Abstract

Despite the escalating imperative for robust environmental education to address climate change and ecological challenges, educators face a significant obstacle in the form of insufficient instructional programs and materials.

The research focused on leveraging mobile technology to create an engaging and effective learning tool for environmental awareness among kindergarten-aged children in Korea. This study aimed to develop and evaluate an interactive mobile application designed to enhance environmental education for young children, using the ADDIE (Analysis, Design, Development, Implementation, Evaluation) model.

The needs analysis was conducted through a survey of kindergarten teachers selected by snowball sampling. The results confirmed the teachers' experience of insufficient learning resources and their needs for interactive, game-based materials to enhance young children's engagement. A flowchart was created to establish learning objectives based on the findings of the needs analysis and to guide the development of the prototype. The development stage involved implementing a prototype using Figma in accordance with the flowchart specifications. The implementation and evaluation phase involved pilot testing, leading to the final refinement of the application.

This study presents the development of an innovative mobile application that uses gamification principles to foster ubiquitous, engaging, and socially interactive learning experiences for early childhood environmental education. The mobile platform serves as a valuable resource for kindergarten educators and parents, enabling them to effectively implement environmental education through accessible digital devices. The research findings provide empirically-based recommendations for the design, development, and implementation of interactive mobile learning applications tailored to young learners. These findings contribute to the broader understanding of how to effectively leverage mobile technology to improve educational outcomes in early childhood settings.

Applying Active Learning to a Seminar on the Basics of Quantum Computing

Pedro J. Roig¹, Salvador Alcaraz¹, Katja Gilly¹, Cristina Bernad¹, Carlos Juiz²

¹Miguel Hernández University, Av. Universidad s/n (Ed. Alcudia), Elche, SPAIN

²University of the Balearic Islands, Ctra. Valldemossa km 7.5, Palma de Mallorca, SPAIN

Abstract

Active learning methodologies allow students to take responsibility for their own learning process, thus allowing to not only acquire technical skills but also soft skills. In this paper, a seminar on the basics of quantum computing has been designed, where different active learning methodologies have been applied all along the seminar. The evaluation of this seminar has been done through the weighted average of four activities, where the first one was a challenge-based learning exercise done on an individual basis. Then, the second one was a team-based learning exercise done on a pair basis. Also, the other two tasks were related to a project-based learning activity, which not only had to be delivered in time, but it also had to be presented in a pitch presentation format. The results obtained in the seminar were successful, as the overall average score was located in the higher end of the grading system, whereas the data distributions per activity all met the conditions for the assumption of normality.

Keywords: active learning, challenge-based learning, team-based learning, project-based learning, quantum computing.

1. Introduction

Active learning could be defined as a student-centered approach in order to construct knowledge focused on activities and strategies aimed at fostering higher-order thinking [1]. Alternatively, active learning could be seen as an approach whose target is to emphasize student engagement and participation [2]. Furthermore, active learning is related to pedagogies based upon constructivism [3].

The concept of active learning is widely associated to student-centered learning, where their cooperation and collaboration are expected during the learning process [4]. Some of the key concepts for active learning to succeed are a better alignment between research and teaching practice, along with a supporting infrastructure, as well as staff professional development and learning designs [5].

There have been reported multiple techniques to apply active learning, such as challenge-based learning, team-based learning, or project-based learning [6], where all of them have in common the assignment of a more active role to students in their learning process, whilst teachers are facilitators [7].

The organization for the rest of the paper goes as follows: section 2 describes the method, section 3 shows the results, section 4 exhibits the discussion, and finally, section 5 draws the final conclusions.

2. Method

We organized a seminar for twelve students about the basics of quantum computing. Those students were in the second semester of their initial course of a degree in computer engineering during the academic course 2024-2025. During the first semester, they learned the basics of linear algebra, which are crucial to deal with the notation, as well the basics of computer science, which are key to understand the key concepts, and the basics of physics, which are necessary to have some core background about quantum mechanics. Hence, they already had the prerequisite knowledge necessary to face the key concepts imparted in this seminar.

The seminar was organized in five sessions, where each of them was structured with a different setup. In fact, four different features define each of those sessions, such that a given session has some specific values for those features, which may well be identified as the dimensions defining each session. Those dimensions are defined as the place where the learning experience is delivered, the time when the learning experience occurs, the frameset in which the learning experience is organized, and the direction in which the learning experience is structured.

The dimension related to space has two different values, which are either on-site if the learning experience takes place in a canonical classroom setting, or otherwise, if it takes place online, where the participants could be anywhere. Also, the dimension related to time has two diverse values as well, which

are either synchronous if the learning experience takes place at the same time for everyone, or otherwise, if each participant does it at its own convenience.

The dimension related to frameset has three different values, which are a frontal class, where a teacher imparts the session and students take a passive role, an interactive class, where a teacher imparts the sessions, even though students take a more active role, and a hands-on class, where students take the active role and the teacher acts as a facilitator. Also, the dimension related to direction has three possible values as well, which are a masterclass, where a teacher delivers a lecture to the students, a breakout room, where students interact to each other in separate groups, and presentations, where students deliver presentations to the rest of peers and to the teacher.

To start with, the first session was a traditional masterclass, where extra concepts of linear algebra for complex domains were imparted by the teacher in a traditional manner. The goal in this session was to learn the algebraic concepts and the matricial notation needed to deal with quantum computing [8]. Regarding the dimensions of this session, it was on-site, as it was held in college, and it was synchronous, as students attended the lecture while it was being held. Furthermore, the frameset of this session was frontal class, as the teacher was lecturing the students, such that the teacher had the active role. Besides, the direction of the session was from teacher to students, as it is the case for a masterclass.

Then, the second session was a self-managed workshop, divided in breakout rooms. With this setup, students were put in pairs so as to deal with an online application called Quirk [9], which is a drag-and-drop quantum circuit simulator. The goal in this session was to understand the effect of quantum gates in a quantum circuit. With respect to the dimensions of this session, it was on-site, as it was held in college, even though it is held outside the classroom time. Furthermore, it was asynchronous, as each group of students did it at their own convenience. Moreover, the frameset of this session was hands-on class, as it was a practical class for students to learn to manage the online application, where the teacher was just a facilitator. Besides, the direction of the session was from students to students within the same group, as it is the case for a breakout room.

After that, the third session was a live online lesson, organized as an interactive masterclass. This session was dedicated by the teacher to expose the basics concepts of a Python library called qiskit in order to build up simulations of quantum circuits [10]. The goal in this session was to learn how to design quantum circuits composed of different quantum gates in Python. With regards to the dimensions of this session, it was online, as it was held on the internet, and it was synchronous, as students attended the lecture while it was being held. Additionally, the frameset of this session was interactive class, as the

teacher was lecturing the students, even though many practical examples were proposed to the students to deal with them during the class. Moreover, the direction of the session was from teacher to students, as it is the case for a masterclass class.

Next, the fourth session was actually some homework for the students, divided in breakout rooms, to be done outside the classroom time. This way, each group composed of two students had to build up a design of a small project with quantum circuits in Python with some predefined conditions by means of the qiskit library [11]. The goal in this session was to face a real project involving quantum computing, which had to be simulated in Python. Regarding the dimensions of this session, it was online, as it was held at any location through the internet, and it was asynchronous, as all groups of students attended at their convenience. Besides, the frameset of this session was hands-on class, as it was a practical class for students to design the quantum circuits through Python code, where the teacher's role was just to solve issues. Moreover, the direction of the session was from students to students within the same group, as it is the case for a breakout room.

Eventually, the last session was a carousel of pitch presentations, where each group composed of two students had to deliver a short presentation about the small project they had to undertake. The goal in this session was to be able to summarize what they did and how they did it with respect to the project [12]. With respect to the dimensions of this session, it was on-site, as it was held in college, and it was synchronous, as students attended the session while the pitch presentations were made. Additionally, the frameset of this session was frontal class, as each group of students had to deliver their presentation to the rest of classmates and also to the teacher, who had a passive role during this session. Furthermore, the direction of the session was from students to students and teacher, as it is the case for a carousel of pitch presentations.

On the other hand, the evaluation of the seminar was done by the teacher according to three items. The first one was a linear algebra exercise, which was proposed at the end of the first session. It had to be done on an individual basis and its weight in the total score of the seminar was 10%. The second one was a Quirk-based exercise, which was proposed at the end of the second session. It had to be done on a group basis and its weight in the overall score of the seminar was 20%. Finally, the third one was the small Python-based project, which involves the pitch presentation about it and the project itself. It was considered on a group basis, where the weight of the former in the total score of the seminar was 30%, whereas the weight of the latter was 40%. Hence, technical skills were slightly more valued, even though soft skills were also contemplated in the final assessment of the seminar. In summary, the overall score is a composite of the

four weighted scores taken out of the three items described.

Additionally, the different exercises established may be associated with different active learning methodologies, as it is stated in Table 1.

Table 1. Activities proposed -vs- active learning

Session ID	Activity proposed	Active learning methodology
1	Linear algebra activity	Challenge-based learning
2	Quirk-based activity	Team-based learning
3	Interactive class	Interactive lecturing
4	Project delivered	Project-based learning
5	Pitch presentation	Role playing

3. Results

As we are based in Spain, the Spanish grading system applies when evaluating the seminar [13]. Therefore, the scores vary in the range between 0 and 10, whilst 5 is considered as the passing score.

In this context, Table 2 shows the average scores for each of the four component scores considered, along with its weighted value.

Table 2. Average scores for each component

Component score	Session related	Weighted value	Average score
Linear algebra activity	Session 1	10%	7.91
Quirk-based activity	Session 2	20%	7.67
Pitch presentation	Session 5	30%	8.08
Project delivered	Sessions 3-4	40%	8.42

Table 3 displays the overall average score obtained in the seminar, which accounts for the weighted average of the four components shown above, along with the overall variance and the overall standard deviation.

Table 3. Overall average score, overall variance and overall standard deviation

Descriptive statistic	Value
Overall average score	8.12
Overall Variance	1.04
Overall Standard deviation	1.02

Table 4 exhibits the skewness and kurtosis for each of the components of the overall score.

Table 4. Skewness and kurtosis

Component score	Skewness	Kurtosis
Linear algebra exercise	-0.16	-0.61
Quirk-based exercise	0.06	-0.17
Pitch presentation	-0.16	-1.26
Project delivered	-0,74	-0.18

On the other hand, Figure 1 depicts a bar chart with the frequency of the overall scores per student, along with a tendency line.

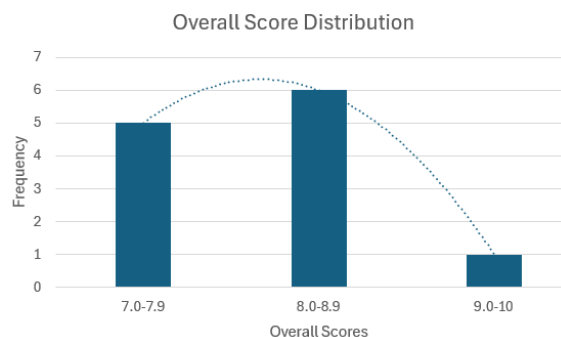


Figure 1. Bar chart for overall score distribution

Furthermore, Figure 2 exposes a histogram of those overall scores per students with 4 bins.

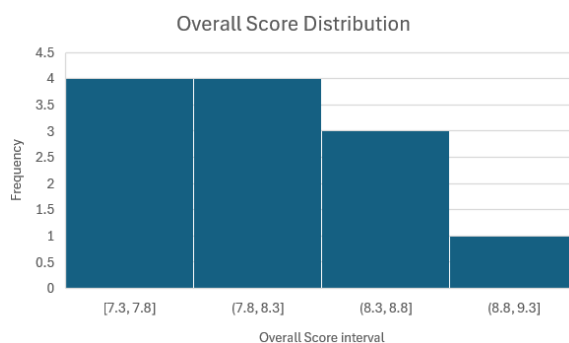


Figure 2. Histogram for overall score distribution

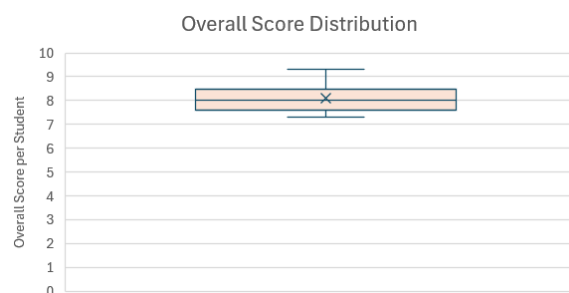


Figure 3. Boxplot for overall score distribution

Additionally, Figure 3 shows a boxplot related to the overall score distribution.

4. Discussion

To start with, Table 2 represents the average score for each component of the overall score. The data show that the average scores obtained in all components are quite close, as the difference between the highest and the lowest values are less than a point. Nonetheless, the component with the lowest score is Quirk-based activity, and this may well have been because none of the students knew this online application before, so they had some difficulties getting used to it.

Afterwards, the component with the second lowest score is the linear algebra exercise, and the reason may have been due to the difficulty in dealing with complex domains for many of the students. On the other hand, the projects presented were quite good considering the lack of initial background of the students in this seminar related to quantum computing, and that is why this component got the highest score. Similarly, the pitch presentations had a good level, probably due to the good level of the projects presented.

Nonetheless, comparing the average scores of the projects and the pitch presentations, the former are slightly higher than the latter, and this could have probably been because STEM students usually have better technical skills than soft skills on a regular basis [14]. In spite of that, both average scores are quite close, so such a difference was insignificant in this case.

Table 3 shows the overall score obtained in the seminar, which accounts for the weighted average of the four components described above. The value was 8.12, which is considered a remarkable score in the Spanish grading system, as it is inside the top interval. On the other hand, the variance obtained is roughly 1 unit, and so is the standard deviation. This fact means that there was some degree of variability in the results obtained in the different components [15], where some students scored top marks, whereas some others did not do it that good.

Table 4 displays the values of skewness and kurtosis for each of the components considered. It is to be reminded that skewness is a measure of the lack of symmetry of a data distribution, which could be seen as looking differently when facing to the left and to the right from the center point of such a distribution. In other words, skewness measures how a data distribution graph differs from the normal distribution. On the contrary, kurtosis is a measure related to whether a data distribution is heavy-tailed or light-tailed with respect to the normal distribution.

By looking at the values shown for both skewness and kurtosis, it is clear that all of them are inside the range going from -2 to +2, which indicates the

assumption of normality of the data [16]. In other words, it can be said that the data distribution from the each score related to the different components follows a normal distribution.

Figure 1 shows a bar chart with the overall score distribution, where 5 students got an average score in the range of [7,8), other 6 students got an average score in the range of [8,9), and 1 student got an average score in the range of [9,10]. Hence, the bar chart is composed of 3 columns, each of them showing the frequency of overall scores with the same integer part. Furthermore, a tendency line is also exposed, which shows a shape somewhat similar to a Gauss bell, even though it could be more related to a degree 2 polynomial.

Figure 2 depicts a histogram with the overall score distribution, where 4 students got an average score in the range of [7.3,7.8), other 4 students got an average score in the range of (7.8,8.3], other 3 students got an average score in the range of (8.3,8.8], and 1 student got an average score in the range of (8.8,9.3]. Therefore, the histogram is composed of 4 columns, each of them showing the frequency of overall scores within the range of half a point, moving from the lowest score to the highest.

Figure 3 exhibits a boxplot with the overall score distribution, where the lower horizontal segment of the box represents the first quartile, the middle horizontal segment within the box stands for the second quartile, namely the median, and the upper horizontal segment of the box accounts for the third quartile [17]. Moreover, the cross within the box shows the overall average, namely the mean, which actually is really close to the median. Furthermore, the horizontal lines shown as the whiskers show the lowest and the highest score as both of them are pretty close to the box. This fact implies that there are no outliers in the graph as the lowest and the highest scores are closer than 1.5 time the interquartile distance from the first quartile and from the third quartile, respectively.

5. Conclusions

In this paper, we describe the setup and the results obtained in a seminar to learn the basics of quantum computing by applying active learning techniques. This seminar was organized for engineering students in the first year, so none of them had previous experience with quantum computing.

The seminar was organized in five sessions, where each of them had a different setup. The first one was an on-site masterclass on linear algebra, specifically designed for quantum computing, where a final exercise was distributed to the students under the challenge-based learning paradigm. The second one was a self-managed workshop, where students worked in pairs in order to design a specific quantum

circuit with an online application, hence the team-based learning paradigm was put in place.

The third one was an online masterclass on a specific library for quantum computing in Python, which was organized as an interactive class. The fourth one was a home activity, where students had to work in pairs in order to complete a small project with the knowledge acquired during the previous session, thus the project-based learning paradigm was used. Finally, the fifth one was a carousel of pitch presentation, where each group of two students had to present its project to the audience, composed by the rest of students and the teacher.

The evaluation of the seminar was carried out with the weighted score of four activities, such that a linear algebra exercise to be done individually was worth 10% of the overall score, an online exercise to be done in pairs was worth 20%, the pitch presentation was worth 30%, and the project delivered was worth 40%. Hence, those four activities were the components of the overall score per student, which was calculated as a weighted average, according to those percentages.

Regarding the results achieved in the seminar, the average scores obtained in all components are pretty similar, as shown in Table 2, where the different average scores per activity differed in less than one point in all cases. Moreover, the overall average obtained for the whole seminar was 8.12 out of 10, whereas the standard deviation was 1.2, as seen in Table 3. Besides, the values of skewness and kurtosis for each component indicates the assumption of normality of such data distribution, as a consequence of the data obtained in Table 4.

With respect to the figures, 11 out of 12 students got overall scores between 7 and 9 according to the bar chart shown in Figure 1, with a tendency line having the shape of a degree 2 polynomial, even though has some degree of similarity to the normal distribution. Moreover, the histogram shown in Figure 2 narrows the interval where the overall scores of 11 out of 12 students are located, thus establishing it between 7.3 and 8.8. Besides, the boxplot shown in Figure 3 shows that the average scores located between the first and the third quartile are quite close to each other, which is indicated by the small height of the box. Also, the mean is really close to the median, whereas the whiskers are also pretty close to the box, which means that there are no outliers because all values are at an acceptable distance from the median. However, the lower whisker is closer to the box than the upper one, which means that the lowest score is closer to the median than the highest score.

In summary, according to the results obtained, we can conclude that the seminar proposed to learn the basics of quantum computing by applying active learning techniques has been a success.

6. References

- [1] Doolittle, P. Wojdak, K., Walters, A. (2023). Defining Active Learning: A Restricted Systematic Review. *Teaching & Learning Inquiry*, 11, 025.
- [2] Dogani, B. (2023). Active learning and effective teaching strategies. *IJASER*, 7(4), 136–142.
- [3] Dubinsky, J.M., Hamid, A.A. (2024). The neuroscience of active learning and direct instruction. *Neuroscience & Biobehavioral Reviews*, 163, 105737.
- [4] Hartikainen, S., Rintala, H., Pylväs, L., Nokelainen, P. (2019). The Concept of Active Learning and the Measurement of Learning Outcomes: A Review of Research in Engineering Higher Education. *Education Sciences*, 9(4), 0276.
- [5] Borte, K., Nesje, K., Lillejord, S. (2020). Barriers to student active learning in higher education. *Teaching in Higher Education*, 28(3), 597-615.
- [6] Ribeiro-Silva, E., Amorim, C., Aparicio-Herguedas, J.L., Batista, P. (2022). Trends of Active Learning in Higher Education and Students' Well-Being: A Literature Review. *Frontiers in Psychology*, 13, 844236.
- [7] Lombardi, D., Shipley, T.F. (2021). The Curious Construct of Active Learning. *Psychological Science in the Public Interest*, 22(1), 8-43.
- [8] Lipton, R.J., Regan, K.W. (2014). Quantum Algorithms via Linear Algebra: A Primer. The MIT Press, 1st edition, pp. 1-208, ISBN: 978-0-262-02839-4.
- [9] Quirk (online quantum circuit simulator), available at <https://algassert.com/quirk> (Access Date: 20 April 2025).
- [10] Introduction to Qiskit (IBM quantum documentation) <https://docs.quantum.ibm.com/guides> (Access Date: 20 April 2025).
- [11] Qiskit (open-source SDK for working with quantum computing) <https://github.com/Qiskit/qiskit> (Access Date: 20 April 2025).
- [12] Clark, C. (2008). The Impact of Entrepreneurs' Oral 'Pitch' Presentation Skills on Business Angels' Initial Screening Investment Decisions. *Venture Capital*, 10(3), 257-279.
- [13] Polytechnic University of Valencia. Spanish grading system, <https://www.upv.es/entidades/OPII/infoweb/pi/info/1147768normali.html> (Access Date: 21 April 2025).
- [14] Roig, P.J., Alcaraz, S., Gilly, K., Bernad, C., Juiz, C. (2024). Design and Assessment of an Active Learning-Based Seminar. *Education Sciences*, 14(4), 371.
- [15] Roig, P.J., Alcaraz, S., Gilly, K., Bernad, C., Juiz, C. (2024). A Generic Multilevel Structure for Educational Escape Rooms. *Education Sciences*, 14(5), 488.

[16] Moghadamnia, M.M., Soleimani-Farsani, B.H. (2023). Relationship between Marital Commitment and Marital Satisfaction with Self-Conscious Emotions and Self-Coherence. *Psychology of Woman Journal*, 4(1), 52-61.

[17] Streit, M., Gehlenborg, N. (2014). Bar charts and box plots. *Nature Methods*, 11(2), 117.

Minimalist Approach: Analysis of “Enigmatic Ruins” Video Game Developed by Game Design Students

Oluwarotimi Abayomi Randle, David Ajayi
GAIC-Lab, University of the Witwatersrand, South Africa

Abstract

The minimalist approach in video games has often been explored from the player’s perspective, with less focus on game developers’ standpoint. This article addresses this gap by examining how game designers use minimalist sound. The study investigates how these sounds are created, the reasons behind their selection, and their alignment with the game’s objectives. Using Affect Theory, the study explores how developers’ creative decisions influence the player’s experience. A video game developer was selected for the study. Facts were obtained through focus group discussions and key informant interviews. The results demonstrate that game developers deliberately choose sounds to enhance the gaming experience and create immersion, although sound can occasionally become distracting in shooting games. The ambience in the selected game creates a sense of place by using mid-frequency sounds. This study provides valuable insights into the minimalist design process, offering aspiring game developers a better understanding of the creative decisions involved in minimalist sound design.

Keywords: sound, video games, minimalist, students, African

1. Introduction

The gaming industry is rapidly expanding the global economy, growing faster than many would expect. It is projected to generate approximately \$363.20 billion by 2027, with billions of gamers worldwide. This has prompted African universities to make significant investments in education programs for video game development [1]. The gamer’s experience is fundamental to designers as they seek to expand the economic scope of game development. Without compromise, the video game industry continues to grow rapidly, attracting more players and generating increasing revenue each year.

In 2023, more than half of the German population played video games regularly, with gamers from different cultures exhibiting a balanced gender ratio. While the video gaming experience is not limited to a

particular age group, it is worth noting that younger age groups (between the ages of 16 and 29) have a high participation rate of 91%. Additionally, at least 19% of gamers around 65 years old enjoy playing video games, resulting in a higher number of gamers within the middle-aged demographic [2].

Sound is an important aspect of gaming, creating an immersive experience for the gamer. It is well known that hearing is the second-most important human sense. It allows us to perceive sounds, which the brain processes to interpret decoded information that can influence our feelings and actions. Factors such as intensity, frequency, and decibel level also play a role in this process. The minimalist approach takes an anti-techno-fetishist stance; instead, it involves reconnoitering and experimenting with new types of game exhibitions through intentional content and design [3]. It is worth noting that indie game developers often experiment with minimalist sounds, stimulating an immersive response in players.

In exploring the role of sound in the gaming experience, music and sound effects in video games are integral to the gamer’s experience, altering their perception of time and encouraging actions and engagement [4]. Additionally, sound can evoke emotions, reactions, and memories in the gamer. This means that an arrangement of low or high frequency sounds can influence specific reactions related to one’s state of mind, such as sadness, cheerfulness, or tension. In game design, minimalist sound refers to a repetitive, digitally created pattern used to stimulate a particular effect. Research has shown that minimal music has a significant effect on the brain [5]. [5] Winnifred Phillips, a game music composer, explains the role of music in video games in her book, noting that minimalist music helps gamers achieve full concentration while playing.

Despite the literature and academic awareness of sound design and minimalist approaches in video games, the minimalist sound approach in game design, particularly from the perspective of developers in South Africa, remains understudied. This article aims to bridge this gap by exploring how game designers structurally used minimalist sound in

the development of the minimalist video games. This paper seeks to gain an in-depth understanding of the motives behind the choice of specific minimalist sound frequencies, considering their intensity and the imaginative effects the developers intended to evoke. It explores how these sounds influence players' emotions to create a new experience through minimalist sound. For this study, a game developed by South African students was selected, with a focus on a structural analytical approach to the game. This study was conducted using Affect Theory as a framework.

2. Literature review on minimalist video game designs

The evolution of sound in video games began with simple origins, when hardware limitations restricted its complexity. The original developer of *Ping Pong* video games diversified the pitch and duration of a single beep sound effect to simulate the player's perception of the ball bouncing off the paddle and hitting the wall [6]. These advancements and innovations in sound creation for video games laid the foundation for using sound and music to stimulate or evoke emotions and influence the audience's actions in powerful ways.

A study of sound in video games identifies three dimensions in which sound, adapted by game designers and developers, interacts with the player and affects the gaming experience [7]. Evidently, this suggests a possible effect of sound on the cognitive interaction of the gamer within the game space. Immersion describes the inclusivity of the gamer while interacting with the game and the possible consequences that sound can have on the player's attitude and actions throughout the game. Similarly, diegetic sound refers to how sounds connect meaningfully to the story taking place within the game world.

To explore minimalism in video games, a minimalist game design methodology was designed [3]. Minimalist games are governed by a set of rules, featuring limited decision spaces, audiovisual illustrations, and a greater focus on the depth of play or the prospect of space [3]. Games with minimalist designs are not limited to a specific game type; instead, they are created by observing basic elements such as interactions, visuals, audio, and game processes – or sometimes by critically examining all of these aspects together. Limited control over the design space leads to more restricted inventions [3], [8].

For minimalist design purposes, the focus is more on the gamer's interaction with objects in the game rather than simply allowing objects to fill space. This

implies that every aspect of the game, whether sound or animation, further enhances and immerses the gamer's experience, thus reflecting the developer's intention to evoke a reaction. Immersion in relation to video games is discussed in [2], which categorizes it into three distinct types. Another study focuses more on the use of minimalist design and does not specifically explore the designer's motive in terms of the choice of sound or music, the projected effect of such sound, frequency ranges, or the targeted audience for a particular selected sound [3]. A third study supports the need to examine the game designer's motive by citing games specifically designed for a visually impaired audience [9]. Such game designers may primarily use audio for communication while creating an inclusive experience that accommodates visually impaired gamers. Sometimes, video features may be omitted, creating a less-than-ideal experience for people without visual impairments and excluding those with hearing impairments. However, the objective of designing games for inclusion based on disability or personal characteristics is not to create separate diverse games but to develop content that enables gamers with different capabilities to enjoy the same experience as their peers.

Investigating the game designer's motive within the framework of minimalist video games is a well-studied area, considering the existing body of literature. Despite extensive research on the use of sound in video games, indie games, minimalism in video games, and various aspects of player involvement, little or no attention has been given to reflecting on the game designer's motives regarding the choice of sound, frequency range, target audience, age group, and general musical knowledge in designed video games. This article attempts to fill the established knowledge gap concerning these objectives.

3. The theoretical framework

Affect Theory seeks to shape a person's emotions, or more broadly, their experiences, into expected, typified physiological and internalized manifestations. It focuses on shaping character through organized means. Affect Theory is attributed to the renowned psychologist Silvan Tomkins. It is based on anticipated or predicted emotions – feelings of pleasure or sometimes displeasure – that originate directly from the consequences of a choice under consideration [10]. A gamer's level of involvement in a game depends on the individual, their abilities, and the general features of the game [11].

In this paper, game characteristics serve as the driving force used by the game designer to stimulate

the gamer's actions or responses to stimuli. The aspects discussed in this article are derived from the principles of minimalist game design from the perspective of the game developer. Affect Theory is used to understand the mindset behind the game's structural architectural design. The article further explores what the game developer aimed to achieve by selecting a particular minimalistic sound for *Enigmatic Ruins* and the intended effect that the game designer wanted to convey to their audience.

3.1. Audiovisual components of the game

The audiovisual components of a game are its major features, as they directly convey information [12], [13]. Two components are involved: the structural design of the game, presented to the gamer in the form of control modes, and the interactive section of the game in which the gamers are engaged.

3.2. Audiovisual components of the game

Games fall into the category of interactive storytelling, utilizing computer-based narrative construction. Interactive storytelling is a method of creating and conveying stories in which the user or player actively participates and can influence the trajectory of the narrative by making choices that affect the outcome. This approach makes the experience more personal and enjoyable for gamers [14]. This shows that the game developer's experience, accumulated over many years, has influenced the narration and the experience they want players to be part of. Over the past 30 years, narrative games and animation have expanded not only in the volume of titles released but also in their exploration of the cultural dimensions of various characters, especially through the nuanced portrayal of ethnically diverse characters and their lived experiences [15]. Video games featuring African-themed riddles and puzzles have been used to explain mathematical and cultural aspects [16]. The benefit is that the narrative influences how the story or mechanics are utilized.

3.3. Immersive level

The term immersive level refers to a gamer's deep engagement and connection, characterized by a sense of being part of the gaming experience. Immersion has been described as the simultaneous perception of sitting in front of a device while being transported to another world within the game [8], [17]. The presence of the gamer in achieving a maximally immersive experience also involves mechanisms such as devotion and intelligent control over the experience [8]. Examples of horror game experiences,

characterized as a state of vigilance that subconsciously puts the gamer in suspense mode, have been described [8].

3.4. Satisfaction and enjoyment

Satisfaction and enjoyment refer to the pleasure and responsive mood experienced during gaming; that is, how engaging the game is. The responsiveness or level of satisfaction experienced may depend on the player's skill level or state of experience [18]. Gamers with a high level of experience may derive more pleasure than novices, for whom increased feedback might include feelings of anxiety, stress, or tension [18].

4. Methods and case study analysis

This paper aims to explore a specific game developer's intention and motive for choosing minimalist sound in the video game *Enigmatic Ruins*. It seeks to reveal the developer's structural ideas through structured questions and focus group discussions, uncovering the motives behind their choices. Primary data were obtained directly from the custodian of the knowledge; that is, the game developer of *Enigmatic Ruins*. The selection process focused on the utilization of minimalism and sound in video games. Out of ten groups, only one used minimalism and sound to guide the development of their game.

4.1. Case study: *Enigmatic Ruins*

4.1.1. Narrative. Set in ancient temple ruins, *Enigmatic Ruins* transitions from a vertical climbing challenge to a horizontal exploration adventure, where players, as treasure hunters, search the temple for hidden keys to escape. This shift introduces new gameplay elements, such as non-player (NPC) interactions, clue-finding, and item collection, designed to evoke curiosity, satisfaction, and tension. The inclusion of NPCs and lore elements encourages players to engage more deeply with the game's story. This narrative integration creates a more cohesive and immersive experience, fostering a stronger emotional connection to the game world.

4.1.2. Visual and audio design. The game's visual style was designed to be both engaging and thematic. In *Enigmatic Ruins*, the abstract environments and physics-based animations from *Mountaineer's Challenge* were transformed into richly detailed, atmospheric settings. This shift required a significant change in art direction, focusing on creating an immersive world that complemented the game's

narrative.

4.1.3. Sound design and implementation. The sound design for *Enigmatic Ruins* aimed to enhance the atmospheric and immersive qualities of the game. Ambient sounds, such as the echoes in the temple, enhance the sense of place.

How were the minimal sounds created? The developer of *Enigmatic Ruins* indicated that the sounds were sourced digitally from online and incorporated to match the visual aesthetics.

What impact do you intend to create with the sound? The minimal sound approach was intended to avoid overwhelming players, allowing them to focus on exploration and puzzle-solving. The sound design of the game successfully contributes to creating a mysterious and immersive atmosphere.

Who is your target audience? *Enigmatic Ruins* is designed for players who enjoy adventure, exploration, and puzzle-solving. The game's mechanics and narrative are designed to appeal to both casual gamers and dedicated players who appreciate a well-crafted story and challenging gameplay.

Is there any integrated cultural sound? During development, indigenous sounds and music were incorporated to enhance the authenticity and cultural depth of the game's setting. These sounds, carefully chosen to match the game's thematic elements, help create a more immersive and respectful portrayal of the fictional culture.

Frequency and age considerations. Sound frequencies and their effects on different age groups were also considered. The music and sound effects were designed to be pleasant and unobtrusive, ensuring a positive experience for players of all ages. This approach makes the game more accessible and enjoyable for a diverse audience.

Player feedback and adaptation of Enigmatic Ruins. Player feedback is fundamental to refining the target audience's considerations. By analyzing player responses and preferences, the developer was able to make informed decisions about including cultural elements and adjusting sound frequencies. This iterative process ensures that the game resonates with its target audience and provides an engaging experience for all players.

5. Conclusion

The analysis of the minimalist approaches in the selected game, *Enigmatic Ruins*, reveals how deliberate minimalist design shapes the player's experience through new mechanics that take factors such as frequency, age, cultural integration, the sound-producing model, and the overall impact of

these sounds into account. This approach creates an immersive atmosphere without detracting from gameplay. The audio cues and their inspired proficiencies, determined by the game developer, further enhance the player's immersion in surprising ways.

This study contributes to the academic literature on video games by exploring designers' intentions and motives, as well as identifying the most suitable minimalist sound for the developed video game. The findings can help game designers consider important factors when designing a video game to achieve a perfect blend of sound and narrative context. It is applicable not only to video games but also to a variety of contexts such as art set-ups and design displays.

This study focuses on the minimalist approach to video game design used by developers. The findings can easily guide game designers in other fields to carefully select appropriate sounds for their games by taking the identified factors into account. Whether crafting video games, art installations, or exhibition experiences, the insights gained from this study offer valuable data for game designers striving to captivate and maintain their audience's immersion by delivering a compelling narrative enhanced with purposeful and suitable minimalist sound.

6. Limitations

This article focuses on video games developed by students in a game design course at a public university. Therefore, this article is limited to students' perspectives regarding video game development and their experiences working on the development of *Enigmatic Ruins*. The level of engagement varies based on the perspectives of the video game developers.

7. References

- [1] Randle, O. (2024). An indigenized framework for game design curriculum for African universities. *Jurnal Bidang Pendidikan Dasar*, 8(1), 25–33. DOI: 10.31004/basicedu.v8i1.2822.
- [2] Haehn, L., Schlittmeier, S. J., and Böffel, C. (2024). Exploring the impact of ambient and character sounds on player experience in video games. *Applied Sciences*, 14(2), Article 583. DOI: 10.3390/app14020583.
- [3] Nealen, A., Saltsman, A., and Boxerman, E. (2011). Towards minimalist game design. In *Foundations of Digital Games*: Cavazza, M., Isbister, K., and Rich, C. Eds., pp. 38–45. DOI: 10.1145/2159365.2159371.
- [4] Cassidy G. and MacDonald, R. (2010). The effects of music on time perception and performance of a driving

game. *Scandinavian Journal of Psychology*, 51(6), 455–464. DOI: 10.1111/j.1467-9450.2010.00830.x.

[5] Szafraniec, K. (2019). Part 1. An article on game music: Where did game music come from: Minimalism. GameMusic.net. <https://gamemusic.net/game/> (Access date: 11 April 2025).

[6] Scarratt, D. (Oct. 24, 2018). The evolution of audio in videogames. ACMI. <https://www.acmi.net.au/stories-and-ideas/evolution-audio-videogames/> (Access date: 11 April 2025).

[7] Munday, R. (2007). Music in video games. In *Music, Sound, and Multimedia: From the Live to the Virtual*: Sexton, J. Ed., Edinburgh: Edinburgh University Press, pp. 51–67.

[8] Fouché, M. (2024). Player experience of minimalist video game design: A case study of indie horror Iron Lung. *Games and Culture*, 21(1), 3–26. DOI: 10.1177/15554120241235437.

[9] Hussain, M. and Javed, M. (2021). Designing accessible games for visually impaired players: A review of game design strategies and challenges. *Journal of Accessibility and Design for All*, 11(2), 45–59. DOI: 10.16910/jadda.2021.11.2.45.

[10] Mellers, B. A., Schwartz, A., Ho, K., and Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science*, 8(6), 423–429. DOI: 10.1111/j.1467-9280.1997.tb00455.x.

[11] Eshuis, S., Pozzebon, K., Allen, A., and Kannis-Dymand, L. (2023). Player experience and enjoyment: A preliminary examination of differences in video game genre. *Simulation and Gaming*, 54(1), 3–22. DOI: 10.1177/10468781231158818.

[12] Caroux, L., Isbister, K., Le Bigot, L., and Vibert, N. (2015). Player–video game interaction: A systematic review of current concepts. *Computers in Human Behavior*, 48, 366–381. DOI: 10.1016/j.chb.2015.01.066.

[13] Caroux, L. (2023). Presence in video games: A systematic review and meta-analysis of the effects of game design choices. *Applied Ergonomics*, 107, Article 103936. DOI: 10.1016/j.apergo.2022.103936.

[14] Heyes, S. (2014, Mar. 18). The four I's of digital storytelling. 8 Million Stories. <https://8ms.com/blog/four-principles-digital-storytelling/> (Access date: 11 april 2025).

[15] Mapondera, E. (2023). Exploring contemporary southern Africa through animation and video games: A case study on KURHWA and The Tale of Stavo. In *Narratives Crossing Boundaries*: Friedmann, J. Ed. Independent Academic Publishing, pp. 281–292. DOI: 10.14361/9783839464861-012.

[16] Randle, O. (2024). The roles of riddles, puzzles, and narratives as pedagogy in video games: A case of African-

themed video games. *Journal of Black Studies*, 55(8), 661–683. DOI: 10.1177/00219347241278383.

[17] Coppi, A. E., Lessiter, J., and Freeman, J. (2014). Get in the game: The influence of embodied play on presence and flow in video games. In *ITAG*, Nottingham, pp. 1–5.

[18] Erlingsson, Á. (2021). Player experience of minimalist video game design: Case study of indie horror Iron Lung. *Simulation & Gaming*, 52(6), 741–758.

[19] Holm, S. (2023). *Preferences, emotions, and visual attention in the first-person shooter game experience* (Doctoral dissertation, University of Turku, Finland). <https://www.utupub.fi/handle/10024/174543> (Access date: 11 April 2025).

[20] Juul, J. (2021). The game of video game objects: A minimal theory of when we see pixels as objects rather than pictures. In *CHI PLAY 2021*, pp. 376–381. DOI: 10.1145/3450337.3483449.

[21] King, G. and Krzywinska, T. (2003). Gamescapes: Exploration and virtual presence in game-worlds. In *Level Up*: Raessens, M. and Raessens, J. Eds. Utrecht University Press, pp. 108–119.

[22] Koenitz, H. (2024). Narrative in video games. In *Encyclopedia of Computer Graphics and Games*: Lee, N. Ed. Springer, pp. 1230–1238. DOI: 10.1007/978-3-319-08234-9_154-1.

Enhancing Project-Based Learning in Engineering Education with Expert-Guided Retrieval-Augmented Generation (RAG)-Based Chatbot

Y.W. Kiang, A.T.L. Lee, V.W.L. Tam, K.H. Lam, V.C.S. Lee, S.M. Yiu

Department of Electrical and Electronic Engineering

The University of Hong Kong

Pokfulam, Hong Kong SAR, China

Abstract

Project-based courses in tertiary education, particularly those involving hardware, present unique pedagogical challenges including misalignment between project and lecture materials, the limited resources for technical support, deviation in students' awareness on the essential technical skills and fundamental knowledge. These factors significantly affect students' learning effectiveness and overall learning outcomes. Accordingly, our team in the Faculty of Engineering at the University of Hong Kong designed and implemented an AI chatbot, utilizing the vector-based Retrieval-Augmented Generation (RAG) with an expert-guided knowledge base, aimed at mitigating these significant challenges and enhancing student learning experiences.

The AI chatbot was implemented as a very supportive and user-friendly learning platform over two consecutive semesters in an undergraduate project-based engineering course involving 134 students. It provided continuous (24/7) and readily accessible guidance with course-relevant responses, powered by the powerful GPT-4o – a state-of-the-art large language model (LLM) – and a carefully structured knowledge base. The knowledge base comprised lecture materials and supplementary question-and-answer (QA) pairs systematically created by experienced course instructors and technical teaching staff, effectively capturing their condensed teaching insights and experiences. The content in the knowledge base was validated by the teaching team to ensure accuracy, relevance, and practical applicability. To facilitate knowledge injection, the expert-guided knowledge base was converted into vector embeddings stored in a vector database and accessed with RAG. RAG is an emerging technique that supplements LLMs with external knowledge without additional model training. When a student query is received, the query will be converted into vector embeddings, knowledge segments with high semantic similarity will then be retrieved and injected into the response generation process. Leveraging both the expert-guided knowledge base and the GPT-4o's extensive pretrained data, the chatbot is capable to deliver factual responses aligned with course materials, to address student project queries without requiring detailed scenario description, and to provide supplementary knowledge essential for project development that might not be directly covered by lecture materials.

To evaluate the effectiveness of the chatbot for engineering education, a combination of both quantitative and qualitative analyses on students' usage patterns and feedback were conducted. A set of questionnaires was carefully designed for the students, which gathered some quantitative ratings regarding the usage frequency, the level of satisfaction and the effectiveness of the chatbot in solving students' course-related queries; and some qualitative comments from students related to the chatbot. To obtain more detailed and valuable feedback from student users, individual interviews were conducted with randomly selected students. In addition to students' feedback, reviews on chatbot interaction logs were conducted by the teaching team, providing supplementary evaluations from educators' perspective. The results reveal the effectiveness of the chatbot in mitigating the pedagogical challenges in project-based engineering courses in tertiary education and enhancing student learning experiences.

This research study represents one of the initial empirical explorations into the application of a vector-based RAG chatbot powered by expert-guided knowledge base in the tertiary engineering education. This work highlights the potential of RAG AI chatbots as practical supporting tools for tertiary or other sectors of education, mitigating real pedagogical challenges and positively influencing student learning outcomes.

Session 2: Learning and Teaching Methodologies

Title: Saudi Secondary Education Tracks: Enhancing Educational Quality and Alignment with International Practices

Author: Khalid Almalhy

Title: Entagogy: Developing a Systems-Theoretical Framework for Autopoietic Co-Construction Between Learners and AI in Posthumanist Education

Authors: Alexander Harris, Stephanie Holt

Title: Using Psychological factors to influence STEM education

Authors: Richard Kyere Asomah, Francis Kwesi Nsakwa Gabriel-Wettey

Title: Teaching to Learning – A Paradigm Shift

Author: Amlan Datta

Title: Development of Mobile Language Learning Applications Using Interactive Multimedia and Gamification for Early Childhood Education

Authors: Yu-Kyung Kim, Yeseo Kim, Hae-Deok Song, Ah-Jeong Hong

Saudi Secondary Education Tracks: Enhancing Educational Quality and Alignment with International Practices

Khalid Almalhy
King Saud University
Saudi Arabia

Abstract

In response to Saudi Vision 2030 and the United Nations Sustainable Development Goal 4 (SDG 4), the Ministry of Education in Saudi Arabia launched the Secondary Education Tracks System in 2021. This reform aims to enhance educational quality, align learning outcomes with labor market demands, and equip students with 21st-century skills. This study investigates the alignment of the Tracks System with international practices and evaluates its impact on secondary education outcomes in Saudi Arabia.

The research adopted a comprehensive mixed-methods approach, surveying 8,232 stakeholders—including educators, administrators, students, parents, and experts from ministries and civil society. Data were collected through structured questionnaires, policy document reviews, and statistical analysis across key dimensions: curriculum content, implementation mechanisms, infrastructure, instructional quality, and student outcomes.

Findings demonstrate strong alignment with global models (e.g., Finland, Singapore, Ontario), particularly in offering flexible learning tracks, integrating digital technologies, and promoting personalized learning. The Tracks System also showed a positive impact on student performance, national test scores, and engagement in project-based learning. However, challenges persist in school infrastructure readiness, equitable access to specialized tracks, and the availability of trained educators. Additionally, there are gaps in academic advising, awareness among students and parents, and integration with postsecondary options.

The study concludes that the tracks system is a transformative initiative with high potential for sustained improvement. It recommends further investment in teacher training, digital resources, and policy clarity to maximize long-term educational impact and maintain alignment with global standards.

Measuring Entanglement: The Entagogy Framework for Bloom-Level Adaptivity in AI Learning System

Alexander Harris¹, Stephanie Holt²
Belgravia Education¹, Brunel University²
UK

Abstract

This paper advances Entagogy as a necessary systems-theoretical framework and build specification for designing human-AI learning systems that learn the learner. Grounded in Luhmannian structural coupling and an entangled reading of Vygotsky's ZPD (e-ZPD), Entagogy formalises the interaction between the Human Cognitive System (HCS) and an AI Semantic Subsystem (AISS) as a closed control loop.

Treating Bloom's "two-sigma" effect as an engineering brief, we outline measurement plans and acceptance criteria that render the framework falsifiable and testable across contexts. The article establishes Entagogy's scope and boundaries, distinguishes it from adjacent lenses (e.g., entangled pedagogy, generic learning analytics), and offers concrete artefacts—state-machine diagrams, threshold tables, and governance checks—to support replication, evaluation, and responsible scaling.

Crucially, we argue that existing educational frameworks: pedagogy, andragogy, heutagogy, cannot be simply transposed onto AI-led learning systems without distortion. Each addresses particular human/AI contexts of teaching and learning, yet none accounts for the autopoietic co-construction that arises when human cognition is structurally coupled with algorithmic subsystems. Research that fails to differentiate which framework is operative risks producing false outcomes, misattributing effects, and obscuring the very mechanisms that drive adaptivity. By advancing Entagogy, we seek to provide a coherent theoretical foundation for future inquiry, one capable of supporting both rigorous scholarship and the responsible development of Intelligent Learning Machines.

Keywords: Artificial Intelligence (AI), AI Semantic Subsystem (AISS), Human Cognitive System (HCS)

1. Introduction

Bloom's seminal claim that one-to-one tutoring yields learning gains two standard deviations above conventional instruction has become a benchmark

for educational innovation. Recent advances in Intelligent Tutoring Systems (ITS), particularly those powered by large language models and adaptive tutoring platforms, have reignited efforts to close the '2 Sigma Gap.' Yet research in this field is hindered by two persistent problems.

First, many studies continue to misclassify learning frameworks, conflating pedagogy, andragogy, and heutagogy with the chronological age of the learner. This simplification obscures the role of autonomy, metacognitive readiness, and self-direction in shaping outcomes, thereby producing misleading attributions of system efficacy.

Second, claims of adaptivity in AI systems are often overstated. Many platforms function responsively, adjusting content in pre-determined ways, without demonstrating the structural coupling or bidirectional feedback characteristic of genuinely adaptive tutoring. Such conflation risks inflating effect sizes and obscures the design conditions necessary to approach Bloom-level gains.

Together, these issues constrain both empirical research and practical development. Without a diagnostic framework to distinguish responsive from adaptive ITS, and without clarity on how frameworks of learning are applied, the field risks generating results that cannot be meaningfully compared or replicated.

This paper introduces Entagogy, a systems-theoretical framework that reconceives AI-mediated learning as a process of autopoietic co-construction. Drawing on Luhmann's approach to structural coupling, Entagogy defines learning as an entangled relationship in which the AI learns the learner while the learner learns the subject. In this formulation, an entangled Zone of Proximal Development (e-ZPD) emerges as the diagnostic threshold for genuine adaptivity.

By operationalizing criteria such as real-time learner-model updates, sustained interactional rhythm, governance permeability, and the presence of exogenous perturbations to prevent autopoietic drift, this paper offers a structured framework for distinguishing between merely responsive and genuinely adaptive AI learning models. This diagnostic clarity ensures that research outcomes can be validly interpreted and compared, as well as providing developers with concrete benchmarks for

design.

This paper therefore positions Entagogy not only as a conceptual advance but also as a practical tool for benchmarking ITS' against Bloom's 2 Sigma challenge. In doing so, it addresses the methodological weaknesses that currently constrain empirical research and supports the development of adaptive platforms capable of sustaining the conditions for transformative learning gains.

2. Empirical Landscape

Bloom's seminal claim that one-to-one tutoring can yield learning gains of two standard deviations above conventional instruction established a benchmark that continues to shape educational innovation [1]. Decades of subsequent research, however, show that while both human and intelligent tutoring systems are highly effective, they rarely exceed gains of around 0.8σ .

A recent systematic review consolidates this evidence, drawing together meta-analyses, controlled trials, design-based experiments, and evaluation frameworks. It shows that AI-powered tutors, including large language model systems, often outperform traditional or even active-learning classrooms, delivering meaningful improvements in engagement, efficiency, and measured outcomes. Yet the gains remain uneven: promising in medical and professional education, positive but modest in K–12 classrooms, and variable across higher education. The review also stresses that what is frequently described as “adaptive” tutoring is often better characterised as responsive branching, with few systems demonstrating real-time bidirectional adjustment or sustained dialogic rhythm [2].

The synthesis identifies recurring methodological limitations. Many interventions are extremely short in duration, creating the risk of novelty effects, and studies often conflate pedagogy, andragogy, and heutagogy with learner age rather than examining autonomy or metacognitive readiness. Evidence also suggests differentiated impacts, with middle school students benefitting more consistently than older learners, and with results concentrated in STEM domains and in US/Asian contexts. Across these cases, the review isolates personalisation, adaptivity, and immediate feedback as the three most reliable drivers of positive outcomes, especially when combined with teacher collaboration and opportunities for self-regulation and mastery learning.

Perhaps most strikingly, none of the reviewed studies addressed ethical concerns such as fairness,

transparency, or accountability. This omission underscores the field's current blind spots. However, taken together, the evidence suggests that AI tutors can now rival the effectiveness of skilled human tutors under specific conditions. Nevertheless, consistent breakthroughs beyond the well-documented plateau remain elusive. Progress will depend on clearer definitions of adaptivity, attention to ethics, and diagnostic standards capable of distinguishing genuine entanglement from surface-level responsiveness [2].

3. Discussion: Introducing Entagogy

Contemporary literature demonstrates the promise of AI-powered personalised learning, yet it also reveals persistent conceptual and methodological weaknesses. Without a shared diagnostic, studies risk conflating responsiveness with adaptivity, overstating system effects and under-specifying the conditions in which they occur. Equally, the widespread reduction of pedagogy, andragogy, and heutagogy to age categories obscures the role of autonomy, metacognitive readiness, and learner agency. These weaknesses generate two consequences: research findings are often difficult to compare across contexts, and intelligent tutoring systems (ITS) are built without clarity on the structural features that underpin lasting learning gains.

A new framework is therefore required to bridge these gaps. For research, such a framework must offer a diagnostic lens that allows outcomes to be interpreted in terms of genuine adaptivity rather than surface-level responsiveness. This enables effect sizes to be contextualised against system properties rather than demographic proxies, reducing the risk of false positives.

For development, the framework must specify measurable criteria that can be engineered into AI learning models, creating a pathway for benchmarking prototypes and iteratively refining them against Bloom's challenge.

3.1. Entangled Learning and the Case for a Luhmannian View

The contemporary shift toward AI-mediated learning requires more than descriptive accounts of interaction between humans and machines. Existing theoretical frames have provided important openings. Latour's Actor–Network Theory collapses the human–machine divide by treating artefacts and people as equally agential nodes [3]

and Postdigital scholars emphasise that these networks are already political, shaped by code, infrastructures, and business models that contour participation [4]. Posthumanist approaches extend this critique by arguing that knowledge itself emerges through material-discursive entanglements rather than circulating through neutral pipelines [5], [6].

Yet description alone is insufficient. In Luhmann's systems theory humans and technologies are operationally closed but structurally coupled, exchanging 'irritations' that allow adaptation while preserving their distinct logics [7]. Education itself can be understood as a self-producing communication system, one that remains viable by coupling with its environment. Here, learners and AI subsystems continuously exchange signals with perturbations from the outside world creating an ecology for learning that moves beyond simple coupling. Adaptive tutoring illustrates this reciprocity: models recalibrate in response to learner performance and affect, while institutions adjust practices and policies in response to AI-mediated shifts.

3.2. Epistemological Framework of Entagogy

Entagogy is grounded in an onto-epistemological position that treats knowledge as relational, co-constructed, and emergent through the continual intra-actions of human and non-human agents. Drawing on Barad's notion of agential cuts [5], it recognises that distinctions such as learner and AI, knowledge and context, are not pre-given but enacted and reconfigured through interaction. Within this view, epistemology is not static but continually renegotiated inside learning ecologies.

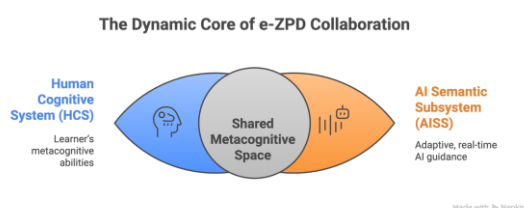


Figure 1. Zone of Proximal Development (ZPD)

At the centre of this framework lies the principle of distributed agency. Rather than privileging either the learner or the AI as the sole locus of autonomy, Entagogy situates learning within the adaptive interplay of structurally coupled human cognitive systems (HCS) and AI semantic subsystems (AISS). This coupling occurs within the entangled Zone of

Proximal Development (e-ZPD), a dynamic space in which human capacities for cognition, affect, and strategy are co-constituted with the adaptive functionalities of the AI system.

Knowledge production is therefore redefined. It is no longer understood as linear transfer or acquisition but as an iterative, recursive process emerging from the reciprocal perturbations and irritations between human and machine [2], [8]. Each exchange produces new conditions for the next, so meanings and understandings evolve through continual co-adaptation. To evaluate such processes, Entagogy calls for robust criteria of knowledge validation: internal coherence within the system, recursive reflexivity across learner and AI interactions, and ethical responsiveness to the broader ecology. Methods must therefore be sensitive to systemic adjustments and capable of tracking co-evolutionary dynamics between learner activity and AI-driven analytics.

Responsible innovation also becomes intrinsic to the framework, ensuring that knowledge claims are not only empirically robust but also socially accountable [9]. In this way, Entagogy situates epistemology and ethics as mutually reinforcing: learning is validated not simply by outcomes, but by the integrity, fairness, and inclusivity of the processes that generate them.

Postdigital scholarship reminds us that HCS/AISS couplings are already political, shaped by infrastructures, business models, and code [4]. Posthumanist theory presses further, locating knowledge within material-discursive entanglements [5] [6]. Luhmann's systems theory adds the essential language of autopoiesis, showing how human and technological systems remain distinct yet adapt through structural coupling [7].

Yet none of these approaches alone resolves the empirical and methodological challenges of AI-mediated learning. ANT and posthumanism describe symmetry and entanglement but lack criteria for persistence. Luhmann explains coupling but remains abstract, without guidance for diagnosis or design. The consequence is that both research and development risk misclassifying responsiveness as adaptivity, overstating outcomes while underspecifying the conditions required for Bloom-level gains.

3.3. Coupling Human Psychology and ITS Led Learning

Vygotsky's Zone of Proximal Development (ZPD) adds to this discourse through the lens of educational psychology: a learner advances by

working with a more knowledgeable other across a span of development that cannot be reached alone [10]. Yet, as Table 1 makes clear, the ZPD in its classical form is insufficient for the realities of AI-rich learning.

The standard ZPD assumes a fixed source of expertise, a relatively stable span of development, and one-directional scaffolding mediated primarily through language. These assumptions collapse in an environment where human and artificial systems are structurally coupled. In such settings, expertise is situational, guidance is polycentric, and mediation extends beyond speech into multimodal and algorithmic cues [8]. The source of expertise is no longer stable but dynamically shared; the zone is not bounded once but redrawn continuously; and scaffolding becomes mutual, with learner actions shaping the AI model while the AI's adjustments reshape learner strategy.

Table 1. ZPD / e-ZPD

Dimension	Classic ZPD (Vygotsky)	e-ZPD (Entagogy)
Mediation	Human "more knowledgeable other" (teacher, peer) provides scaffolding and feedback	Mediation occurs through both human and AI; feedback and support are multimodal, adaptive, and algorithmically mediated
Guidance	Largely unidirectional (teacher/peer guides learner)	Guidance is bidirectional and dynamic; both learner and AI can act as "more knowledgeable other" at different moments
Stability	Zone shifts as learner develops, but structure is relatively stable within an interaction	Zone is fluid, shifting moment to moment in response to both learner and AI adaptation; the boundary is co-constructed
Agency	Primarily human, with teacher/peer as agent of change	Distributed and non-hierarchical; agency is shared, negotiated, and emergent from HCS-AISS coupling
Boundary	Defined by human social, cultural, and cognitive context	Boundary includes technical, algorithmic, infrastructural, and exogenous (social, policy, technical) influences
Reciprocal Learning	Implicit—teacher may learn about the learner, but focus is on learner's development	Explicit—AI and human both adapt and "learn"; each system updates its model in real time
Mode of Feedback	Verbal, social, contextual, often slow	Multimodal: text, speech, visuals, haptic; can be near-instantaneous and deeply personalised

It is for this reason that Entagogy advances the concept of the entangled ZPD (e-ZPD). Unlike a general Luhmannian approach, which offers a broad theory of systems and communication, the e-ZPD provides a usable map for both researchers and designers. It reframes personalisation as structural coupling rather than content adjustment, showing where expertise resides at any moment, how mediation unfolds across multimodal channels, and

under what technical conditions (adaptivity, low latency, permeable governance etc.) the coupling sustains itself.

The diagnostic value is twofold. For researchers, the e-ZPD clarifies what counts as genuine adaptivity: rapid role transfers, structural reciprocity, and continuously shifting boundaries. Without this lens, studies risk misclassifying responsiveness as adaptivity, producing false positives in effect size claims. For developers and psychologists, the e-ZPD defines measurable prerequisites for Bloom-level gains.

In sum, Entagogy's use of the e-ZPD is not a theoretical embellishment but a practical necessity. Where classical ZPD cannot account for the dynamics of AI-augmented learning, and where Luhmann alone remains too abstract, the e-ZPD provides the conceptual and diagnostic bridge. It couples human psychology with machine learning in a way that makes Bloom's challenge not only intelligible but testable.

4. Limitations of Entagogy

Entagogy provides a powerful and versatile theoretical lens to examine how entanglement between an AISS and HCS reciprocally shape learning, it is crucial to acknowledge that not every interaction between learner and AISS constitutes entagogy learning.

Table 2. Required conditions of Entagogy

Dimension	Why it matters for Entagogy	Typical failure signal
Context & Infrastructure	Reliable bandwidth, up-to-date devices and teacher training are prerequisites for low-latency adaptivity.	Technical "hiccups" force the AI to drop into fallback modes, re-humanising the session.
Data Ethics & Privacy	Continuous modelling relies on granular trace data; privacy regimes (COPPA, GDPR) can throttle that flow.	Learners encounter generic feedback because fine-grained logs are masked or siloed.
Bias & Cultural Fit	An AISS tuned on majority-culture discourse may mis-frame language, gesture or prior knowledge.	Minority learners receive irrelevant or pathologising prompts, undermining psychological safety.
Interpretability & Trust	Black-box adaptivity can provoke teacher or learner scepticism, stalling the recursive loop.	Users override or ignore AI advice, reverting to pre-AI routines.
Domain Boundaries	Some competencies (e.g. tacit craft skills, embodied movement) exceed current NLP-centric AISS capacities.	Progress plateaus despite high adaptivity scores, signalling that the target skill needs physical co-presence.
Scaling Effects	What couples well in a pilot may overload compute budgets or moderation capacity at district scale.	Latency climbs and governance locks tighten, collapsing the coupling.

The framework explicitly demands certain structural conditions to be fulfilled before the desired recursive, co-constructive coupling emerges. In other words, simply deploying an advanced AI tool or interacting frequently with an adaptive learning environment does not automatically yield entagologic outcomes.

Central to the Entagogy framework is the notion that meaningful learning arises from the dynamic interplay, or structural coupling, between the Human Cognitive System (HCS) and the AISS [11]. This interplay, however, can only flourish when specific enabling thresholds are consistently maintained. Without satisfying these critical thresholds - adaptivity, latency responsiveness, and governance permeability - the interaction between learner and AISS reverts from a truly co-evolving partnership into a more traditional, unilateral dynamic. Recognising these limits is essential not only for accurately diagnosing when and why entagologic approaches may falter, but also for refining the design of AI-mediated environments to better meet these foundational conditions

4.1. Conditional Coupling

Autopoietic coupling between the Human Cognitive System (HCS) and the AISS constitutes the foundational mechanism of the e-ZPD. However, such coupling is neither automatic nor guaranteed. It depends critically on three enabling thresholds, each of which must be met to sustain a genuinely recursive, co-constructive learning ecology. These thresholds: adaptivity, latency, and governance permeability, define the structural conditions under which learner and AISS can operate as distinct yet mutually responsive systems. Where any one threshold fails, the system reverts to a default mode: either a traditionally human-led scaffold or a heutagogy building AI tutor.

Adaptivity: For autopoietic coupling to take hold, the AISS must operate not as a static feedback engine but as a live epistemic partner [8]. This requires that the system continuously update its internal learner models within the same session, recalibrating predictions, scaffolds, and interface strategies based on ongoing learner inputs; be they cognitive (e.g., answer correctness), behavioural (e.g., pacing), or affective (e.g., sentiment shifts).

If the system's adaptivity is too coarse-grained (e.g., relying on pre-scripted branching paths) or updates only at the end of a session, the illusion of co-construction collapses. The learner ceases to perceive the AISS as a responsive partner and instead experiences it as a static tutor; one that

repeats, rather than evolves. This breaks the recursive loop and inhibits the formation of structural coupling. As a result, the ZPD becomes frozen, linear, and teacher-dependent once again.

Latency: The temporal rhythm of feedback is equally critical. For coupling to feel and be dialogic, the system must respond fast enough to maintain the attentional envelope. Research suggests that this window, particularly for visual and interactive tasks, falls within a 300–600 millisecond range.

The AISS responding too slowly, breaks the interaction's epistemic rhythm. The learner begins to reassert control and autopoiesis is broken. In effect, the AI is demoted from partner to passive tool. Conversely when feedback is instantaneous and semantically relevant, the AISS becomes experienced as an active participant, capable of "thinking with" the learner in real time. This speed-based alignment is essential for establishing the recursive irritations that define autopoietic systems.

Governance permeability: Governance must ensure ethical safety without sterilising the system's responsiveness through programmable transparency, not hard-coded stasis [9]. The capacity of an AISS depends on institutional policy architectures that either permit or prohibit algorithmic reparameterisation. Where audit rules, compliance frameworks, or locked-down governance APIs prevent the system from adapting (for example, in the name of fairness audits or GDPR compliance) the AISS becomes inflexible. It can still recommend, but it cannot learn.

This results in what we might call pseudo-coupling: the learner continues interacting, but the system is no longer responding in kind. The feedback becomes performative, not generative. Worse still, over time the system may default to protecting itself from liability rather than optimising learning outcomes. A rigidly governed AISS risks functioning more like a grading assistant than a cognitive partner.

4.2. Regressive Outcomes in the Absence of Coupling

The integrity of the entangled Zone of Proximal Development (e-ZPD) is contingent upon the simultaneous fulfilment of three structural conditions: adaptivity, latency responsiveness, and governance permeability. Where any of these conditions is absent or compromised, the autopoietic coupling between the Human Cognitive System (HCS) and the AI Semantic Subsystem (AISS) is disrupted. In such instances, the recursive perturbations and mutual irritations that sustain co-

evolution between subsystems are severed, resulting in a reversion to more traditional, linear instructional dynamics. Specifically, the breakdown of each threshold yields distinct regressions:

In the absence of adaptivity, the AISS reverts to pre-scripted instructional patterns. Rather than responding in real time to learner inputs, the system defaults to static, pre-determined sequences. This collapse into a didactic feedback loop inhibits the co-construction of meaning and re-establishes the educator or system as a unilateral source of knowledge.

Latency exceeding the learner's attentional threshold, breaks the illusion of dialogue. Delays in system feedback diminish the perceived responsiveness of the AI, prompting the learner to disengage from co-constructive exchange. In such scenarios, learners are likely to revert to self-scaffolding strategies, effectively reclaiming full agency and undermining the dialogic reciprocity that characterises an entangled ZPD.

If governance frameworks restrict algorithmic re-parameterisation, the AISS is rendered operationally inert [9]. Policy constraints that prevent the system from adapting its internal models in response to emergent learner behaviour result in a rigid and prescriptive interaction. The learner, in turn, is forced to adapt their learning strategies around the fixed constraints of the tool, effectively transforming the AISS from a cognitive partner into a static recommender system.

In each of these cases, the ZPD regresses to a conventional scaffolding paradigm: either a human-led model in which the educator assumes sole responsibility for guidance, or a system-led routine in which the AI dictates fixed pathways without co-adaptive responsiveness. In either instance, the defining characteristic of Entagogy; the recursive, co-autopoietic interaction between learner and intelligent system, is suspended.

4.3. Autopoietic Drift and the Necessity of Exogenous Irritation

Autopoietic systems reproduce their own operations; without perturbation they drift into reductive self-confirmation [7].

To avoid siloed echo-loops, Entagogy hard-wires mandatory irritations from neighbouring function systems. The loops stabilise the learning ecology without breaching operational closure: each subsystem retains its code (e.g., cognitive validation, policy compliance) while remaining sensitive to perturbations.

Table 3. Exogenous Irritations in Entagogy

External system	Irritation channel	Typical cadence	Example signal	Resulting adjustment
Institutional policy	Governance API call-outs	Daily cron or event-trigger	"Equity disparity > 10 %"	AI re-balances content difficulty, teacher receives alert (Steiner-Khamis, 2021)
Peer community	Social annotation layer	Real-time	High divergence in peer answers	AI injects peer exemplar, learner invited to critique
Cultural knowledge base	Live corpus sync	Weekly	New taxonomy term enters discipline	Suggestion engine surfaces updated concept map
Ethical oversight	Bias-audit webhook	Monthly or ad hoc	Fairness test p-value < .05	Model weights recalibrated, bias notice displayed (Rosiek et al., 2024)

5. Approaches to Methodology in Entagogy Research

Both researchers and developers must be able to determine when entagogy is occurring as distinct from pedagogy, andragogy, or heutagogy. Without diagnostic clarity, studies risk both misclassifying learning effects and misleading the development of intelligent tutoring systems.

Entagogy therefore advances not only a theoretical framework but also a research orientation. Rooted in systems theory, it supplies the criteria for recognising the formation of an entangled Zone of Proximal Development (e-ZPD), in which a human cognitive system (HCS) and an AI semantic subsystem (AISS) become structurally coupled [7]. The methodological focus thus shifts from aggregate learning outcomes to the identification of real-time conditions under which entanglement occurs, stabilises, or fails. This requires an inquiry architecture that can track emergent knowledge across cognitive, technical, and socio-material layers.

5.1. The Four Lenses of Entagogy

Rather than prescribing a single protocol, Entagogy proposes a transdisciplinary research stance that is sensitive to three commitments.

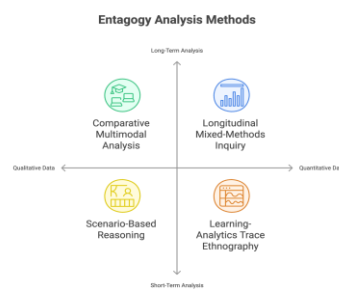


Figure 2. Research Methodology for Entagogy

First, to structural coupling, ensuring that analysis remains centred on reciprocal adaptation rather than

unilateral instruction. Second, to reciprocal metacognition, attending to the ways learner strategies and AI feedback co-constitute one another over time. Third, to ethical reflexivity, embedding concerns of bias, equity, and inclusion within the very criteria used to recognise entanglement [11]. These commitments establish the foundation for a plural yet coherent methodological toolkit (see Figure 2).

Because learning is conceived as a co-autopoietic process in which human cognition and AI semantic subsystems continuously perturb and recreate one another, research designs must operate across scales. Micro-level loops of interaction must be mapped alongside macro-level stabilisations such as curricular adjustments or institutional policy shifts. To achieve this, Entagogy advances a recursive research architecture organised around four complementary lenses.

Scenario-based reasoning functions as the imaginative front-end of entagogic research. Using design fictions, participatory studios, or agent-based simulations, researchers can model hypothetical learning ecologies, surfacing latent variables, ethical inflection points, and threshold conditions such as adaptivity, latency, and governance permeability [11]. Each scenario serves as a stress-test, perturbing the not-yet-built system so that potential couplings and breakdowns are revealed before deployment.

Learning-analytics trace ethnography then examines entagogy at the level of practice. Clickstreams, gaze paths, conversational turns, and affective signals are read simultaneously as quantitative traces and as cultural artefacts [8]. Algorithmic pattern-recognition highlights candidate moments of bidirectional metacognition, while ethnographic annotation situates the events within learners' lived experience. Through this lens, quantitative measures and qualitative narratives inter-animate each other to expose how micro-irritations cascade into meso-level trajectories.

Longitudinal mixed-methods inquiry provides the temporal depth required to establish whether Entagogy has occurred. Repeated measures of cognitive-affective alignment, policy-pedagogy coupling coefficients, and learner-authored reflections reveal how early perturbations consolidate into stable routines or drift into over- or under-coupling. Interviews and classroom observations ground these quantitative patterns in context, ensuring that analysis captures both structural stability and human experience.

Finally, comparative multimodal analysis supplies the lateral perspective needed for generalisation. By aligning discourse analysis,

interface semiotics, and biometric or affective data across multiple sites, researchers can separate universal signatures of entagogic coupling from local adaptations. Cross-case synthesis prevents over-reliance on any single implementation and anchors theory-building in diverse empirical bases.

5.1.2. Recursive Research Architecture. These four lenses are not sequential stages but interlocking components of a recursive research loop. Insights from speculative scenarios recalibrate trace-analytic parameters; emerging ethnographic patterns shape longitudinal designs; cross-case syntheses feed back into the next wave of scenario-building. In this way, the methodological apparatus mirrors Entagogy itself: a dynamic system that learns by continuously perturbing and refining its own operations.

This recursive architecture allows scholars to establish not only whether Entagogy has taken place but also how it stabilises, transforms, or fails. By embedding ethical and equity commitments - psychological safety, inclusion, transparency - at the core of each analytic move, the approach ensures that methodological rigour is inseparable from responsibility.

6. Creating A Diagnostic Protocol for Entagogy

Without a protocol that turns Entagogy's thresholds (adaptivity, guidance rhythm/latency, governance permeability, reciprocity, perturbation, equity) into testable invariants, development teams risk shipping responsive systems that only mimic adaptivity or over-fitting to short-run gains that do not persist. The Diagnostic Protocol formalises those invariants and aligns them with the four research lenses, so results are auditable and comparable across deployments enabling both researchers and developers to accurately gauge when genuine entanglement is occurring, rather than when a system merely looks personalised.

6.1. Integrating a synergy metric to detect entanglement

To operationalise 'when' entagogy occurs this paper integrates findings from Riedl and Weidmann [12]. Introducing a Bayesian Item Response Theory framework that decomposes performance into individual ability (θ) and collaborative ability (κ), validated on interactive human-AI data; critically, κ is distinct from θ and varies with user factors (e.g., perspective-taking), Riedl and Weidmann provide

uncertainty estimates and model comparisons. Their analysis shows substantial human–AI gains and clear differences between AI models (e.g., higher synergy with GPT-4o than Llama-3.1-8B), while demonstrating that collaborative ability can be estimated separately from solo ability. This offers a principled, model-agnostic signal for when synergy/entanglement is present.

We therefore treat $\kappa > \theta$ (beyond a pre-registered margin and with credible intervals) as a necessary *statistical* indicator that must co-occur with structural indicators from Entagogy (within-session adaptivity and role-transfer reciprocity, sub-second guidance rhythm, governance permeability, and stability under exogenous perturbation). This binds outcome-level synergy to the e-ZPD's system properties, preventing false positives from answer-revealing or latency-masked boosts.

6.2. Study-level Operationalisation

At the study level, entanglement can therefore be identified by alternating solo and assisted trials on calibrated or near-isomorphic items so that performance can be decomposed into individual ability (θ) and collaborative ability (κ). The trial schedule should be randomised and sparse enough to avoid practice artefacts while providing sufficient information for Bayesian estimation. Prior to data collection, the entanglement decision rule must be pre-registered - for example, $\kappa - \theta > \delta$ with a 95% credible interval (CrI) and concurrent satisfaction of structural thresholds (within-session adaptivity, sub-second guidance rhythm, evidence of role-transfer reciprocity, and compliant governance). This aligns the outcome signal (κ exceeding θ) with the system properties that Entagogy treats as necessary for an e-ZPD to exist.

Validation ensures that observed κ gains reflect co-construction rather than content leakage or answer-revealing. To this end, learning-analytics can be paired with trace ethnography: dialogic turns tagged for reciprocity (two-way metacognitive prompts, plan–monitor–evaluate moves) and micro handovers (AI↔learner shifts in guidance). These qualitative tags can then be time-locked to the quantitative traces so that κ elevations can be attributed to reciprocal adaptation rather than superficial assistance.

Durability is assessed longitudinally with researchers tracking whether κ is sustained or increasing while hint reliance declines, indicating entanglement without dependency. Robustness can be probed through exogenous irritations (policy or content updates, teacher interventions, or interface

changes) and by examining whether κ and the structural thresholds remain stable under these perturbations.

Equity can further be gauged when κ estimates are disaggregated by subgroup (e.g., language background, gender, connectivity context), with pre-specified margins defining unacceptable gaps and documented corrective actions when thresholds are exceeded. The diagnostic protocol therefore has the potential to embed ethical responsiveness within the same evidentiary framework used to claim entanglement.

Finally, Scenario-based reasoning is able to provide stress-tests as to whether design conditions are sufficient for κ to emerge; trace ethnography instruments latency distributions and role-transfer events that substantiate reciprocity; longitudinal panels test the persistence of κ and structural thresholds under perturbations; and comparative multimodal analyses examine the generalisability of entanglement across domains, infrastructures, and cohorts. Together, these lenses turn the protocol into a recursive apparatus capable of determining when an e-ZPD has formed and whether it is maintained.

6.3. Reporting standardisation

For replication and meta-analysis, studies should report:

- (1) $\kappa - \theta$ estimates with credible intervals and the pre-registered entanglement rule;
- (2) latency statistics (median, P95) and latency-breakdown correlations;
- (3) counts of within-session learner-model updates and scaffold revisions;
- (4) reciprocity markers (AI↔learner role-transfers; two-way metacognitive prompts/responses);
- (5) enacted governance changes/self-tuning bounds;
- (6) equity-disaggregated κ and remediation;
- (7) the log of exogenous irritations and system adjustments.

6.4. Implementation (recursive apparatus)

The Diagnostic Protocol is recursive, not a one-off checklist. Scenario findings inform trial structure and thresholds; trace ethnography captures interaction rhythm, scaffold revisions, and role handovers; longitudinal analysis tests whether κ and structural thresholds hold over time; comparative analysis prevents false generalisation. By triangulating across lenses—and explicitly

incorporating the Bayesian synergy decomposition so that $\kappa > \theta$ serves as a quantitative trigger—researchers can determine with rigour whether an entangled ZPD has emerged and is maintained under varying conditions.

Brief note on mechanism: Riedl and Weidmann also show that Theory of Mind predicts collaborative ability (κ) and fluctuates within users, underscoring the need to model reciprocity dynamically rather than assume static profiles, further aligning with Entagogy's emphasis on bidirectional metacognition and within-session adaptivity.

6.4. Summary of Diagnostic Protocols

The Entagogy Diagnostic Protocol converts Entagogy's thresholds (adaptivity, guidance rhythm/latency, governance permeability, reciprocity, perturbation, and equity) into testable invariants, aligned with four research lenses (scenario, trace ethnography, longitudinal, comparative). Its purpose is to determine when genuine entanglement occurs, not when a system merely looks personalised. To operationalise this, we are indebted to and integrate Riedl and Weidmann [12]: a Bayesian IRT decomposition into individual ability (θ) and collaborative ability (κ). We treat $\kappa > \theta$ (beyond a pre-registered margin, with credible intervals) as a necessary statistical signal that must co-occur with structural indicators (within-session adaptivity and role-transfer reciprocity, sub-second rhythm, permeable governance, stability under perturbation).

We further advance study-level operationalisation through alternate solo and assisted trials on calibrated/near-isomorphic items (randomised, sparse) to estimate θ and κ ; pre-register the entanglement rule (e.g., $\kappa - \theta > \delta$, 95% CrI + thresholds met). Validate via trace ethnography by time-locking reciprocity tags (two-way meta moves; micro handovers) to κ elevations. Test durability longitudinally (sustained/increasing κ with declining hint reliance) and under exogenous irritations (policy/content updates, teacher interventions, interface changes). Ensure equity by disaggregating κ by subgroup with pre-set margins and documenting corrective actions when gaps exceed thresholds.

With $\kappa > \theta$ as the quantitative trigger and Entagogy's thresholds as structural guards, the above protocol rigorously determines whether, and under what conditions, an e-ZPD has formed and is maintained.

7. Conclusion

By shifting attention from age-based pedagogical labels to the structural conditions under which a human cognitive system (HCS) and an AI semantic subsystem (AISS) genuinely couple, the framework of Entagogy resolves two chronic sources of noise in literature: the misclassification of responsiveness as adaptivity, and the false certainty that follows when pedagogy, andragogy, and heutagogy are mapped to demographics rather than autonomy and metacognitive readiness. In doing so, this paper offers researchers and builders a common language for when learning systems actually adapt, rather than merely appear to.

The contribution is both conceptual and operational. Conceptually, the entangled Zone of Proximal Development (e-ZPD) locates learning in the reciprocal perturbations between learner and system, where expertise is shared, roles transfer, and mediation is multimodal. Operationally, the Entagogy Diagnostic Protocol turns that concept into practice: adaptivity, guidance rhythm (latency), governance permeability, reciprocity, perturbation, and equity become measurable thresholds; $\kappa > \theta$, derived from a Bayesian synergy decomposition, becomes the necessary quantitative trigger that must co-occur with those structural signs to claim entanglement. This binds outcomes to mechanisms and protects the field from flattering false positives.

The implications are immediate. For research, pre-registered $\kappa - \theta$ decision rules, trace-ethnographic tagging of reciprocity, longitudinal durability checks, and perturbation tests make studies comparable and auditable across sites and models. For developers, the thresholds function as design requirements: engineer for within-session model updates, sub-second, semantically relevant feedback, and permeable, programmable governance, or accept plateauing at responsive branching.

Entagogy does not eliminate constraints. Its operation remains contingent on infrastructural capacity, regulatory environments, domain specificity, and cultural appropriateness. It also recognises the inevitability of autopoietic drift, embedding exogenous perturbations as a safeguard against closure and stagnation. This paper posits that these are not reasons for retreat, but conditions that demand careful instrumentation, iterative validation, and sustained monitoring.

Progress towards Bloom-level gains is unlikely to be achieved through sudden leaps in model performance. Rather, it will require the disciplined alignment of behavioural design, temporal rhythm,

and governance architectures around demonstrable reciprocity; conditions that can be observed, measured, and reproduced across contexts.

With Entagogy, the central inquiry shifts from the reductive “Did AI increase learning?” to the more rigorous “At what points did entanglement manifest, by what indicators was it verified, and under which conditions did it endure?”

This reframing establishes a standard that is both replicable and auditable. Teachers speak in the language of pedagogy; engineers in the language of algorithms. The bridge between them is absent. Entagogy provides that bridge and establishes a rigorous blueprint for crossing it with equity, transparency, and scale.

8. References

- [1] Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 13(6), 4–16. <https://web.mit.edu/5.95/www/readings/bloom-two-sigma.pdf> (Access Date: 1 May 2025).
- [2] Létourneau, A., Deslandes Martineau, M., Charland, P., et al. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. *npj Science of Learning*, 10, 29. DOI:10.1038/s41539-025-00320-7.
- [3] Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.
- [4] Bozkurt, A. (2024). Postdigital Educational Technology. In P. Jandrić (Ed.), *Encyclopedia of Postdigital Science and Education*. Cham: Springer. [Online]. DOI:10.1007/978-3-031-35469-4_57-1.
- [5] Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC: Duke University Press.
- [6] Morales, L., and Zarabadi, S. (2024). *Posthuman Educational Mappings: Agential Realism and Educational Practices*. New York: Routledge.
- [7] Luhmann, N. (1997). *Die Gesellschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- [8] Burriss, S. K., and Leander, K. M. (2024). Critical posthumanist literacy: Building theory for reading, writing, and living ethically with everyday AI. *Reading Research Quarterly*, 59(1), 123–144.
- [9] Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

[10] Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

[11] Steiner-Khamsi, G. (2021). Externalisation and structural coupling: Applications in comparative policy studies in education. *European Educational Research Journal*, 20(6), 806–820.

[12] Riedl, C., and Weidmann, B. (2025, September 22). Quantifying Human-AI Synergy. *PsyArXiv*. DOI:10.31234/osf.io/vbkmt_v1.

Acknowledgements

The authors would like to thank the World Academy of Tirana (WAT), and Deutsche Schule Bombay (DSB) for their support during the writing and presentation of this paper.

Using Psychological Factors to Influence STEM Education

Richard Kyere Asomah, Francis Kwesi Nsakwa Gabriel-Wetley
University of Ghana
Ghana

Abstract

Although STEM education implementation has been running for decades, only a few studies evaluate the prospects of optimism, confidence, and motivation is shaping students' interest in STEM related fields of employment. The study examines the relationships among Motivations, Optimism and Confidence and how these affect student's interest in STEM related field of employment in Ghana. Semi-structured interviews and questionnaires were employed to gather data from 351 students, sampled using stratified sampling technique. CFA and EFA attested to the validity and reliability of the questionnaires. Quantitative data was assessed using Descriptive and structural equation modeling (SEM), and qualitative data was analyzed using word descriptions. The findings of the study established that, Motivation is positively correlated with students' interest in STEM related fields of employment ($\beta = .61, p < .01$). Similarly, Confidence and Optimism were also found to be positively correlated ($\beta = .67, p < .01$) and ($\beta = .58, p < .01$), respectively. Finally, it was discovered that, Confidence in STEM has the strongest relationship with Extent of Interest in STEM ($\beta = 0.67, p < 0.001$). This finding supports the importance of fostering motivation, optimism, and confidence in STEM education to promote student interest and engagement. Implications of the study for policy, theory, practice and further research are discussed.

Teaching to Learning – A Paradigm Shift

Amlan Datta
Lone Star College
Texas, USA

Abstract

Education systems worldwide have long relied on traditional, teacher-centered methods. Here, the teacher is the primary source of knowledge, the Sage-on-Stage, and students are passive recipients. However, research has shown that active participation in the learning process leads to better understanding and retention. As Benjamin Franklin lucidly stated: “Tell me and I forget, teach me and I may remember, involve me and I learn”. Simply put, this involves a paradigm shift from our traditional understanding of teaching, to one of student learning. As the late ‘Tabla’ maestro, Ustad Zakir Hussain, once explained: “The teacher never teaches, the student learns ...” In this interactive session, I’d like to share my experiences and experiments with teaching versus learning – both in the face-to-face classrooms, and in the online format. This proposal aims to highlight the importance of shifting from a teaching-centric to a learning-centric approach to improve educational outcomes and foster lifelong learning. This will be an interactive session and sharing of ideas, for a shift from teaching to learning based on some of the experiences of the presenter.

Traditional methods often rely on lectures, memorization, and standardized testing. These methods can limit creativity, critical thinking, and the ability to apply knowledge in real-world situations. Many students struggle to stay engaged in a passive learning environment. Lack of motivation can lead to poor academic performance and disengagement from the learning process.

Studies have shown that learner-centered approaches lead to better academic performance, higher retention rates, and improved understanding of complex concepts. By consciously making a paradigm shift from teaching to learning, we can help students develop essential skills such as critical thinking, problem-solving, communication, and collaboration.

The need for a paradigm shift from teaching to learning emphasizes the benefits of learner-centered approaches in enhancing educational outcomes and preparing students for the future. The call to action is to encourage educators (as well as policymakers, and stakeholders) to support and invest in this transition. It highlights the importance of innovation and collaboration for continuous improvement in the education system.

Development of Mobile Language Learning Applications Using Interactive Multimedia and Gamification for Early Childhood Education

Yu-Kyung Kim, Yeseo Kim, Hae-Deok Song, Ah-Jeong Hong
Chung-Ang University
Seoul
South Korea

Abstract

While excessive exposure to media-based information during childhood has been strongly linked to poor literacy outcomes, the ability to access, understand and apply meaningful information is increasingly recognized as a critical skill in the current digital age. In light of this, relying solely on restricting media access to improve children's literacy skills has significant limitations. Instead, attention is shifting to harnessing the educational potential of technology-enhanced learning as a transformative tool for enhancing literacy in the digital age.

The aim of this study is to develop a mobile Korean language learning application designed to increase the learning engagement of children aged 5 to 6. The application was designed to support learners' voluntary participation and engagement using interactive multimedia. Gamification principles were incorporated into the language learning application to effectively capture children's interest and encourage meaningful interaction.

The development of our digital language learning application followed to the ADDIE model, a widely recognized instructional design framework that includes phases of analysis, design, development, implementation, and evaluation. This structured approach ensured that our learning platform was tailored to meet the specific needs and engagement levels of 5-year-old children. The results showed that children showed increased engagement, such as proactively exploring word meanings and showing interest in the Korean alphabet. Teachers also gave positive feedback on the pedagogical potential of the content and interface.

This research highlights the potential of digitally integrated, playful approaches to early childhood language learning. In conclusion, the integration of technology-enhanced learning into early childhood education holds great promise for improving literacy. By harnessing the interactive nature of digital tools and aligning them with proven instructional design models such as ADDIE, educators can create engaging and effective learning experiences for young learners..

Session 3: Pedagogy and AI

Title: The Compulsion to Create in Artists' Narratives: A Psycho-Pedagogical Perspective
Author: Agnieszka Janiszewska

Title: Engaging Young Children with Mathematics – Is Agency a Requirement?
Authors: Audrey Cooke, Maria Johansson, Eva Noren

Title: AI's Impact on Higher Education: Transforming Research, Teaching, and Learning
Author: Alyse Jordan

Title: AI Digital Textbooks in Korean Classrooms: The Codle Platform in Action
Author: Justin Eom

The Compulsion to Create in Artists' Narratives: A Psycho-Pedagogical Perspective

Agnieszka Janiszewska

*Department of Educational Studies, Faculty of Educational Sciences, University of Lodz
Poland*

1. Scope

During the presentation, I will share the findings of a study that explored the characteristics of the compulsion to create phenomenon and its significance in the process of artists' self-actualization. The theoretical framework is grounded in concepts from positive psychology and positive education, with particular reference to Abraham Maslow's hierarchy of needs, Mihaly Csikszentmihalyi's theory of flow, and Scott Barry Kaufman's revised hierarchy of needs. The study employed an exploratory, qualitative approach, using free-form interviews as the research method. The sample comprised 24 artists, representing a variety of artistic disciplines. Based on the data collected, I conceptualized the compulsion to create as a deeply felt, recurring need for artistic expression that motivates individuals to regularly engage in creative activity. This drive frequently generates highly positive emotional states, particularly the experience of flow. The analysis led to the identification of key characteristics of this phenomenon and culminated in the development of a model describing the relationship between the need to create, lower-level human needs, and self-actualization. Regarding the second objective of the study, a major conclusion is that the fulfilment of the compulsion to create is, to a significant extent, synonymous with the fulfilment of the need for self-actualization.

The presentation will conclude with proposals for future research directions and recommendations for educational practice. These include, among others:

- workshops for artists who experience a strong need to create that is insufficiently integrated with other fundamental life needs (e.g., relational or material security);
- workshops for artists who feel a strong creative drive but are blocked by creativity inhibitors, particularly those of an intrapsychic nature (e.g., negative beliefs or low self-esteem in the creative context);

- workshops or individual consultations for artists and their families experiencing relational conflict stemming from the high prioritization and deep involvement in creative work by the artist.

2. Objective and Motivation

The key objective of my presentation is to share the findings of my original exploratory research on the compulsion to create and its role in the self-actualization of artists with an international academic audience. As both a researcher in pedagogy and creatology, and a practicing artist – a novelist – I believe that the phenomenon of the compulsion to create remains insufficiently recognised, despite its significant impact on the well-being of creative individuals. I argue that this phenomenon deserves greater attention, as a deeper understanding of the specific developmental needs of those who experience a strong creative drive may encourage more effective educational and psychological support. The practical goal of my research was to derive insights from empirical data that could serve as a foundation for developing educational programs and coaching interventions tailored to artists and their families.

Having recently completed my PhD in the field of pedagogy, I am highly motivated to engage with scholars and practitioners in education, to critically exchange perspectives, and to further develop the theoretical and practical dimensions of my work through interdisciplinary and cross-cultural dialogue. Additionally, with a background in English philology, I am particularly interested in establishing research collaborations within the Anglo-Saxon academic community.

Engaging Young Children with Mathematics – Is Agency a Requirement?

Audrey Cooke¹, Maria Johansson², Eva Noren³

¹Curtin University, Australia

²Luleå Tekniska Universitet², Sweden

³Stockholms Universitet³, Sweden

Abstract

The focus of this research paper is on early childhood education, and it presents some initial findings that consider whether children's agency is required for young children to engage in mathematics. Mathematics is considered in terms of Bishop's mathematical activities as reframed for use with very young and preverbal children in a preschool setting. A discussion of agency follows, which draws the reader's attention to how young children's agency may be connected to play, learning, intentionality, materialities, and entangled intra-actions. Initial findings from research using 360-degree video collected in a preschool is then analysed to see whether more mathematics (as per Bishop's mathematical activities) is evident when a young child has greater agency.

1. Introduction

This paper presents research being undertaken in a preschool setting in Sweden. This is a setting typically attended by children aged 1 to 5 years of age and is commensurate with an early learning centre in the Australian context. In these countries the setting is a mix of ages, which differs to some other countries, such as Singapore for example, the setting is structured around separate ages. The aim of the research is to investigate how young children's mathematical engagement might be evident in activities they complete in the preschool setting. Bishop's [1, 2] mathematical activities and the reframed version proposed by Cooke and Jay [3] were used to assist in identifying the mathematics that may be evident. This paper investigates whether young children's agency is a necessary condition for the emergence of mathematical activities in preschool settings. Specifically, we ask: Does greater agency lead to a broader range of mathematical activities being observed?

2. Early childhood and mathematics

Young, and very young, children can engage in

mathematical thinking [3]. Björklund [4] proposes that toddlers engage in mathematising, where these very young children use mathematics to make sense of their world. The mathematics that these very young children use can be identified through Bishop's [1, 2] mathematical activities as reframed by Cooke and Jay [3].

2.1. Mathematical activities

Bishop [1] identifies six mathematical activities – counting, locating, measuring, designing, playing, and explaining – that he described as fundamental and universal, and as being used in a deliberate and sustained manner either individually or combined together (the latter of which he describes as more significant as they would interact). He proposes that the first four “have been, and still are, the bases of science, engineering, manufacturing industry, trade, agriculture, war, etc.” [1, p. 57], while playing and explaining are considered as just as important – playing incorporating modelling and analysis of life that promotes understanding of the social world and explaining being “a thoroughly refined and sophisticated activity” [1, p. 58]. In considering the child and mathematics, Bishop [1] stated that mathematics education should consider children “as active learners, engaged in developing their cultural knowledge through social interaction with other people within the cultural group, who are the carriers and bearers of ideas, norms and values of that culture” [1, p. 88].

Cooke and Jay [3] reframed the six mathematical activities to focus on the actions of young and preverbal children. For example, Cooke and Jay [3] propose that counting incorporates the child quantifying objects, such as identifying how many could be held or changes in how many there are and using patterns as may be evident in their routines or as moving to music; locating incorporates the child's body movement or positioning in space and the movement of objects; measuring incorporates the child identifying attributes of objects and other elements in their environment, as well as the amount

of the attributes the object or elements contain; designing incorporates the child recognising attributes related to functionality, which can involve a particular purpose; playing incorporates the child's capacity to recognize and use rules or procedures, and to predict what may occur in their environment; and explaining incorporates the child showing their understanding of their environment or world and the objects and elements within it.

2.2. Intra-action and materiality

In this study, children's agency is understood through a posthumanist lens where humans and materialities are mutually entangled. Barad [5] introduces the concept of intra-action to describe how individuals and materials interact in ways that dissolve the boundaries between subject and object. When children engage with materials (such as the pretend fishpond and fishing rod), their actions are not solely expressions of individual intent but are shaped through dynamic encounters with the material world.

2.3. Agency in early childhood education – and mathematics education

The national preschool curriculum of Sweden [6] states that teachers should not only communicate mathematical concepts to the children but create situations or engage themselves in play leading to specific learning objects. Teachers are also required to introduce mathematical concepts in specific contexts and to scaffold children to build mathematical understanding by investigation and discovery. The curriculum also addresses values and content, providing direction for teachers and principals regarding what they ought to do in order for the children to reach the goals in the curriculum.

The framework for early learning in Australia, *Belonging, Being and Becoming: The Early Years Learning Framework for Australia* [7], emphasises children's agency as contributing to both the child's sense of self and of belonging, and their learning. The Australian Government Department of Education [7] advocates children having agency when engaging in play and learning as well as in their everyday lives and worlds. Children's agency is often considered together with intentionality and their right to be active participants, and is considered essential in helping children recognise their capacity to impact the world in which we live. This reflects elements of Biesta and Tedder [8], who describe agency as something individuals achieve rather than have in unique or specific contextual situations – that is, they view it as achievement-of-agency.

Several reviews of agency in the literature for younger children in preschool have been published [9]. In these reviews, agency has been defined in

relation to a vast variety of theories. Therefore, agency's theoretical richness is complicated because there is accordingly a huge diversity of meanings attached to the concept of agency. Varpanen [9] suggests that children's agency and theory must be understood in terms of its use in practice. "An encyclopaedia definition for agency is the capacity to act intentionally" (Schlosser, 2015, as cited in [7, p. 2]) but, according to Varpanen, "The fundamental tension between the individual and social aspects is a necessary background for understanding the different conceptualisations of agency." [7, p. 2]. Barad [5] takes this a step forward, saying that agency is relational; children, materialities and context are entangled in intra-actions.

Delacour [10] explores how teachers in preschool impact children's understanding of mathematical concepts. Materiality and language effect the mathematical activities, both if they are planned by a teacher or spontaneously emerge from the children. Simply said, Delacour posits that in a planned activity, the teacher utilises control, and children follow instructions; whereas in a spontaneous activity, like free play, the teacher follows and the children 'lead'. When the children were allowed "to move and investigate in various directions, to follow their imagination and not remain stuck in what the teacher had in mind, /.../, a devolution process occurred." [9, p. 9].

In the same vein, McCormick Smith and Chao [11], emphasise that play-based undertakings position learners' agency at the forefront and reduces the position for the teacher as central instructor. Typically play-based preschool rooms are designed into sections, such as doll house, block or pillow area, water table, light table, reading/book space, spaces for play with for example Lego, puzzles, dinosaurs or other common toys, games and materials. Here, children can freely investigate materials and spaces "while building autonomy, self-regulation, and social and emotional skills." [10, p. 6]. McCormick Smith and Chao [11] highlight that the teaching of mathematics in early childhood usually has focused on counting and cardinality. Today, research shows that preschool children can grasp more complex mathematics [12]. As an example, research by Clements and Sarama [13] demonstrate that young children can comprehend complex patterns and geometrical manipulations.

de Freitas and Sinclair [14] consider embodied agency, where "Diagramming and gesturing are embodied acts that constitute new relationships between the person doing the mathematics and the material world" [13, p. 134]. In doing so, they propose that although the learners' capacity to use diagramming to communicate their ideas is a skill that should be developed, it is more where "a highly material process of becoming entwined and enfolded with the material surfaces engaged in the encounter.

It is precisely these encounters that we believe substantiate an embodied mathematical agency.” [13, p. 151]. That is, the learner has agency in how they represent their mathematical understanding, and they achieve this by physically replicating their mathematical understandings on the material surface. In a similar way, Norén [15], considers agency as being “highly dependent on the material world for its meaning-making” [14, p. 150], with an emphasis on “the ‘doing’ in the mathematical activities and the materiality of social interaction” [14, p. 150]. Within the mathematical activity, the agency exists in the entanglement between the learner and the materiality.

2.4. The tension against schoolification

According to a more than one hundred-year-old Swedish preschool practise, activities in the preschool have been understood in terms of a holistic view of play, care, development and learning [16], [17]. Helenius [18] adds to this, stating that preschools in Sweden have been grounded on a solid play-base, and pedagogy in preschools has been derived from children’s interest. Dale Tunnicliffe and Gkouskou [19, p. 57] consider play as the work that children do, and emphasise the importance of investigating the understandings children may demonstrate during play.

Unfortunately, the term teaching (in Swedish - undervisning), introduced into the preschool curriculum 2018, might lead to more formalised teaching, that is, a more restricted/instrumental teaching approach. Mathematics teaching for young children is often associated with school mathematics. Eidevald et al. [20] emphasize that the perceptions of both education and teaching in preschool need to be charged, with distinctive meanings based on preschool’s long and well-established traditions and more recent preschool didactic research.

2.5. The aims of this research

The question of how children’s agency relates to the mathematics that may be evident in their actions and behaviours is explored through analysis of video. This preliminary analysis considers a segment of video that was collected as part of the larger project. The video was selected as it contains a play experience that starts with the children choosing the to engage with specific materialities (in this instance, a pretend fishpond set up on a table in the room) and then progresses to teacher guidance. That is, during the 5 minutes and 26 seconds, there are changes in how the materialities were entangled with the children, the amount of structure and guidance provided by the teacher, and the children’s choice and agency in the materialities they intra-acted [5] with. Specifically, the children engage initially by

their own initiative with a fishpond setup and later with teacher guidance. The study explores whether the children’s agency influences the presence of mathematical activities in children’s actions through video analysis.

3. Method

Video has been used to identify Bishop’s [1, 2] mathematical activities. Johansson et al. [21] and Johansson et al. [22] use Bishop’s [1] six mathematical activities to identify mathematics that young children in a Swedish preschool may engage with. Specific mathematical activities from Bishop [1] have also been identified in videos. Helenius et al. [23] focus on measurement via temperature in the video they analyse. Meaney [24] analyses two videos of two much younger children in a preschool setting, although the focus was on Bishop’s [1] mathematical activity of locating.

Björklund [25] and Lynch and Stanley [26] outline the benefits and use of video when examining the actions and occupation of very young children. Björklund [23, p. 86] states that using video enables a more thorough examination of the actions, often in a more natural or everyday environment. Likewise, Lynch and Stanley [24, p. 58] describe how video can be used to scrutinise the occupation of very young children in an everyday home setting. Both Björklund [23, p. 86] and Lynch and Stanley [24, p. 64] espouse the capacity of video to be viewed and reviewed when being analysed. However, Lynch and Stanley [24, pp. 58-59] warned of several issues – namely, the power of the one who is videoing in terms of what is focused upon, whether the person holding the video camera is a participant or not, and the impact of the act of being recorded on the participants.

This project uses a 360-degree camera that was suspended from the ceiling of the room that the participants occupied. The camera was set to record either manually or remotely as the participants were entering or in the room. The camera was visible to participants but was not reachable. The camera recorded all of the session when the participants were in the room. All of the room (and participants) were visible and audible on the video. When the video was rendered, it produced a 360-degree video. The platforms used to store the video were secure and enabled the video to be manipulated by the viewer to focus in on a specific section of the room or on specific participants.

The collection of video using the 360-degree camera enabled most of the issues raised by Lynch and Stanley [24, pp. 58-59] to be addressed. As the camera was attached to the ceiling, although in the line of sight of the participants, it reduced the impact of the positioning of the video on the participants. As no researchers were in the room during the session,

the role was one of observer without participation, more a non-participant observer [27, p. 5]. The camera captured the entire room, which reduced an element of power of the researchers, but power still exists in terms of what the researchers choose to focus upon when viewing the video.

3.1. Participants and situation

Several videos were collected from two locations in Sweden. The videos were collected from the preschool after appropriate consent was given by the centre managers, teachers, and parents and guardians of the children in the video. The video analysed for this paper features one teacher and two children aged 2 years and 3 years, with the children both described as having sufficient language to communicate verbally. The room has a large doll house, a light table, and a pretend fishpond created on a table, as well as assorted toys and shelves. The video is 24 minutes and 26 seconds long but the focus of the video is on a section that starts at 2 minutes and 12 seconds and ends at 7 minutes and 38 seconds. The next section discusses some initial findings investigating whether agency is required for young children to engage in mathematics. Although there are two children, Child 1 is the focus.

4. Results

The event starts with a preschool teacher and two children, Child 1 (age 3) and Child 2 (age 2) sitting on a mat in one of the corners of a rectangular room. Child 1 looks at the teacher and Child 2, rises up and puts hands on hips, starts walking, looks in a mirror on the wall, and then walks directly to a fishpond located in the corner diagonally from the first corner. Child 1 goes down on knees, getting closer to and looking at photographs placed lower on the wall. Child 1 looks at two of photographs on the wall, the photos represent underwater sea life. Then Child 1 points at the first and then the second; saying “look here, those are the last here”. In this event at least four of Bishop’s [1, 2] activities are present, *locating*, *measuring*, *counting*, and *explaining*. Child 1 moves in the room, sits, rises, walk, sits on knees; *locating* in the spatial environment. Child 1 also explores their relationship with the items of interest when moving closer to the photographs on the wall, enacting *measuring* through recognising they need to decrease the distance between themselves and the photographs on the wall. Child 1’s stating of “look here, those are the last here” points to *counting*. Lastly, Child 1’s act of verbalizing their interpretation of the photograph is *explaining*. Considering the event as intra-action with Child 1 and ‘things’/materialities that are constantly influencing and working inseparably, the child’s agency, specifically in the start, is a “dynamism of

forces” [5, p. 141].

In a second event a few seconds later, Child 1 turns to the fishpond table, bend down on one knee, then straightens to stand with their hands on the of the table. Child 1 starts to pick up the a paper fish in the fishpond with their right hand and touches it with left hand, then picks up the fishing rod with left hand. They put the fish onto the magnet at the end of the string attached to the fishing rod, bouncing the rod twice and the fish falls off. Five of Bishop’s [1, 2] activities are present in this event. Similar to the first event, *locating* and *counting* are present, as well as *measuring*, using objects by putting the rod and fish together; *playing*, bouncing the rod, touching the fish; *designing*, putting the fish on to the rod, choosing which fish to take. Also in this event, Child 1 and materialities are entangled in intra-actions – as an example, there would be no bouncing without the rod and the child; likewise, choosing a fish among other fish is an intra-action with Child 1 and the different fish. This demonstrates several relations in intra-actions with non-human materialities and Child 1’s enacted agency.

Further on Child 1 picks up a blue fish with their hands and moves it towards the magnet at the end of the rod string but drops it. In that moment the teacher has moved to the fishpond and takes the fish. The fish has a drawn picture of a hand with five fingers on it, and the teacher says: *It says here, how many fingers? Pointing to the fish 1, 2, 3, 4, 5, you have to show the full hand.* Child 1’s relational agency shows no intra-action with the teacher who seems to want Child 1 to count out loud and to show fingers while counting out loud. These events provide rich illustrations of how the child’s agency and engagement with materialities can support multiple mathematical activities. In the following discussion, we interpret these findings through the lenses of Bishop’s [1, 2] mathematical activities and Barad’s [5] notion of intra-action. Child 1 displays a variety of Bishop’s mathematical activities while freely interacting with the fishpond materials. However, when the teacher intervened and directed Child 1 toward counting fingers on a fish image, the child’s mathematical engagement became narrowly focused on counting.

5. Discussion

Agency and play are in focus for early childhood in Sweden and Australia [6], [7]. Agency has been considered as part of mathematics education in terms of embodied acts [14] and the ‘doing’ to make meaning [15]. In this research, mathematics was evident in the intra-actions of Child 1. As had been previously found [21], [22], Bishop’s [1] six mathematical activities are useful in helping identify the mathematics that may be evident in young children’s play in a preschool setting. This research

found that the child, Child 1, engaged with all six of Bishop's [1] mathematical activities when they had agency in their intra-actions with the materialities of the fishpond. Specifically, Bishop's [1] mathematical activities of *counting*, *locating*, *measuring*, *designing*, *playing*, and *measuring* were all identified when the child had agency in their intra-actions. When the teacher engaged with the child, the mathematics that was evident constricted, likely due to the focus on counting that the teacher brought to the intra-action.

To search for mathematical activities [1, 2] and agency in children's actions in micro minutes, with the lens of Barad [5], led us to discover the variety of mathematical activities in which a single child, Child 1, and materialities are entangled. The young child's enacted agency shows how much mathematics (as per Bishop's [1, 2] mathematical activities) is evident in the intra-actions. This is in great contrast to the teacher's more instructional way of trying to make Child 1 count and show a full hand of five. Although Walkerdine's [28] differentiation of pedagogical and instrumental experiences was not a focus of this paper, it seems that the teacher was engaging more in a way that reflects Walkerdine's [28] pedagogical task, where the teacher was focused on a specific area of mathematics – on how many fingers were present in the image on a fish. This resulted in the intra-actions incorporating a focus on Bishop's [1, 2] mathematical activity of *counting* but seemingly not Bishop's [1, 2] five other mathematical activities.

The difference in which of Bishop's [1, 2] mathematical activities were present is stark. In line with what de Freitas and Sinclair [14] have noted, the child had greater agency in how they demonstrated their mathematical understanding when they were engaging with the fishpond and could physically represent their mathematical understandings through how they intra-acted with the materialities of the fishpond. In having agency, the child had choice in their intra-actions with the materialities of the fishpond. This resulted in intra-actions where the child had agency in using mathematics to achieve their desired outcomes. Consequently, we propose that more mathematics is evident when a young child enacts greater agency. The findings show that agency and free exploration enabled a richer, more diverse set of mathematical activities. When adult intervention shifted the focus to counting, other mathematical activities were less evident. These results align with de Freitas and Sinclair's [14] insights about embodied mathematical agency and support educational approaches that prioritize children's autonomy in learning.

6. Conclusion and limitations

As has been stated previously [19], situating play

as the work that children do means that the understandings that children may have can and should be identified whilst they are playing. This helps situate the findings of this research, specifically, that all of Bishop's [1, 2] mathematical activities were evident when Child 1 had greater agency in their engagement with the fishpond, and not when the teacher endeavoured to engage the child with mathematics using the resources provided with the fishpond. It may be that the focus of the teacher on engaging the child with providing how many fingers are pictured on the fish reflects Walkerdine's [28] framing of a pedagogical approach to mathematical experiences. This lead to the experience incorporating Bishop's [1, 2] mathematical activity of counting, and the narrowing of the mathematics that the child could demonstrate. This study highlights the importance of supporting children's agency in early childhood mathematics as, by enabling children to freely interact with materials, teachers can foster a broader and richer range of mathematical understandings than when learning is more heavily teacher-directed.

7. Limitations

Two broad limitations are evident in this research. The first is that the researchers determine which sections of video are analysed and, within that video, what to focus on. The use of the 360-degree video expanded what was captured but it was still the researchers who determine what is focused on in the 360-degree video. The second limitation is that, although the use of the 360-degree video emphasised the benefits previously outlined [25], [26], the use of 360-degree video and the resultant positioning of the researchers as non-participant observer [27] removes the opportunity for the teacher to provide their rationale for their engagement with the child, and a consequence is that the reasoning behind their engagement is not available. Planned future research intends to investigate this by giving the teacher the opportunity to share their thoughts on the footage captured on the 360-degree video and the decisions they undertook both before and during the experience that was captured on the 360-degree video.

8. References

- [1] Bishop, A. (1988). Mathematics education in its cultural context. *Educ. Studies in Math.*, vol. 19, no. 2, pp. 179–191. DOI: 10.1007/BF00751231.
- [2] Bishop, A. (1991). *Mathematical enculturation: A cultural perspective on mathematics education*. The Netherlands: Kluwer Academic Publishers. DOI: 10.1007/978-94-009-2657-8.

- [3] Cooke, A., Jay, J. (2021). Bishop's (1988, 1991) mathematical activities reframed for pre-verbal young children's actions. *The 14th Int. Congress on Math. Educ. ICME14*, Shanghai, China, July 11-18.
- [4] Björklund, C. (2018). Powerful frameworks for conceptual understanding. *Forging Connections in Early Math. Teaching and Learning*, V. Kinnear, M. Y. Lai, and T. Muir, Eds., pp. 37–53. doi: 10.1007/978-981-10-7153-9.
- [5] Barad, K. (2007). *Meeting the universe half way: Quantum physics of the entanglement of matter and meaning*, Duke Universal.
- [6] Skolverket. (2018). *Curriculum for the Preschool Lpfö 18*. <https://www.skolverket.se/getFile?file=4049> (Access Date: 12 February 2025).
- [7] Australian Government Department of Education, "Belonging, Being & Becoming: The Early Years Learning Framework for Australia," Version 2, 2022. [Online]. <https://www.acecqa.gov.au/nqf/national-law-regulations/approved-learning-frameworks>. (Access Date: 2 Mary 2025).
- [8] Biesta, G., Tedder, M. (2007) "Agency and learning in the lifecourse: Towards an ecological perspective," *Studies in the Educ. of Adults*, vol. 39, no. 2, pp. 132-149. DOI: 10.1080/02660830.2007.11661545
- [9] Varpanen, J. (2019). What is children's agency? A review of conceptualizations used in early childhood education research. *Educ. Res. Rev.*, vol. 28, 100288. DOI: 10.1016/j.edurev.2019.100288.
- [10] Delacour, L. (2011). Mathematics in Swedish multilingual preschool: human and nonhuman actors and the devolution process. *Phil. of Math. Educ. J.*, vol. 39.
- [11] McCormick Smith, M., Chao, T. (2018). Critical Science and Mathematics Early Childhood Education: Theorizing Reggio, Play, and Critical Pedagogy into an Actionable Cycle. *Educ. Sci.*, vol. 8, no. 4, 162. DOI: 10.3390/educsci8040162.
- [12] Park, J., Ramirez, G., Park, D. (2024). "Effect of preschool teacher's math anxiety on teaching efficacy and classroom engagement in math," *Psych. in the Schools*, vol. 61, no. 6, pp. 2600-2611. DOI: 10.1002/pits.23182.
- [13] Clements, D.H., Sarama, J. (2014). *Learning and Teaching Early Math: The Learning Trajectories Approach*. Routledge.
- [14] de Freitas, E., Sinclair, N. (2012). Diagram, gesture, agency: Theorizing embodiment in the mathematics classroom. *Educ. Studies in Math.*, vol. 80, pp. 133-152. DOI:10.1007/s10649-011-9364-8.
- [15] Norén, E. (2018). Agency, Materiality, and Mathematics Learning in a Preschool Classroom. *Inside the Math. Class*, U. Gellert, C. Knipping, and H. Strahler-Pohl, Eds., pp. 145-164, Springer. DOI: 10.1007/978-3-319-79045-9_7 E.
- [16] Helenius, O., Johansson, M.L., Lange, T., Meaney, T., Riesbeck, E., Wernberg, A. (2016). When is young children's play mathematical? *Math. Educ. in the early years: Results from the POEM2 conf. 2014*, T. Meaney, O. Helenius, M.L. Johansson, T. Lange, and A. Wernberg, Eds., Springer International Publishing, pp. 139-156. DOI: 10.1007/978-3-319-23935-4_8.
- [17] Tallberg Broman, I. (19919). När arbetet var lönen. En kvinnohistorisk studie av barnträdgårdsledarinnan som folkuppfostrare. (Dissertation). Almqvist & Wiksell International.
- [18] Helenius, O. (2018). Explicating professional modes of action for teaching preschool mathematics," *Res. in Math. Educ.*, vol. 20, no. 2, pp. 183-199. DOI: 10.1080/14794802.2018.1473161.
- [19] Dale Tunnicliffe, S., Gkouskou, E. (2020). Science in action in spontaneous preschool play—an essential foundation for future understanding. *Early Child Dev. and Care*, vol. 190, no. 1, pp. 54-63. DOI: 10.1080/03004430.2019.1653552.
- [20] Eidevald, C., Engdahl, I., Frankenberg, S., Lenz Taguchi, H., Palmer, A. (2018). Omsorgsfull och lekfull utbildning och undervisning i förskolan. *Ingår i: Undervisning i förskolan: en kunskapsöversikt*, 81, S. Sheridan and P. Williams, Eds.
- [21] Johansson, M.L., Lange, T., Meaney, T., Riesbeck, E., Wernberg, A. (2014). Young children's multimodal mathematical explanations. *ZDM*, vol. 46, no. 6, pp. 895-909. DOI: 10.1007/s11858-014-0614-y.
- [22] Johansson, M.L., Lange, T., Meaney, T., Riesbeck, E., Wernberg, A. (2016). What maths do children engage with in Swedish preschools? *Math. Teach.*, vol. 250, pp. 28–33. <https://www.proquest.com/trade-journals/what-maths-do-children-engage-with-swedish/docview/1807741647/se-2>. (Access Date: 31 March 2025).
- [23] Helenius, O., Johansson, M.L., Lange, T., Meaney, T., Wernberg, A. (2016). Measuring temperature within the didactic space of preschool. *Nordic Studies in Math. Educ.*, vol. 2, no. 4, pp. 155–176. DOI: 10.7146/nomad.v21i4.148751.
- [24] Meaney, T. (2016). Locating learning of toddlers in the individual/society and mind/body divides. *Nordic Studies in Math. Ed.*, vol. 21, no. 4, pp. 5-28. DOI: 10.7146/nomad.v21i4.148740.
- [25] Björklund, C. (2008). Toddlers' opportunities to learn mathematics. *Int. J. of Early Childhood*, vol. 40, no. 1, pp. 81–95. DOI: 10.1007/BF03168365.
- [26] Lynch, H., Stanley, M. (2018). Beyond words: Using qualitative video methods for researching occupation with young children," *OTJR: occupation, participation and health*, vol. 38, no. 1, pp. 56–66. DOI: 10.1177/1539449217718504.
- [27] Crotty, M. (1988). *The foundations of social research: Meaning and perspective in the research process*. Allen & Unwin.

[28] Walkerdine, V. (1988). *The mastery of reason: Cognitive development and the production of rationality*. London: Routledge.

AI's Impact on Higher Education: Transforming Research, Teaching, and Learning

Alyse Jordan
Encore Knowledge Group
USA

Abstract

The swift advancement of artificial intelligence (AI) presents significant opportunities and challenges for higher education. In a recently published open educational resource by Frontiers in Higher Education, both editors and authors examined the diverse impacts of AI on academic institutions, emphasizing ethical, social, and pedagogical dimensions. This initiative aims to explore how AI improves academic performance, enriches learning experiences, and drives innovation in educational methodologies. By showcasing groundbreaking research and innovative teaching strategies, we will illustrate AI's global influence, its enhancement of instructional practices, and the ethical issues that arise.

This research fosters an investigation into AI's profound effects on teaching, learning, and research, providing insights into curriculum development, assessment practices, and faculty workflows. By addressing critical ethical, social, and pedagogical considerations, we aspire to inspire further research and influence policy.

The research covers a wide range of topics, including AI-driven pedagogical transformations, personalized learning experiences, and interdisciplinary collaboration. We welcome a variety of contributions, from original research articles to systematic reviews and case studies, enhancing our understanding of AI's potential to promote educational equity and excellence. Join us in exploring the transformative possibilities of AI in academia and navigating its complexities for a brighter educational future.

AI Digital Textbooks in Korean Classrooms: The Codle Platform in Action

Justin Eom
Team Monolith
South Korea

Abstract

The The Korean computer science AI Digital Textbook platform revolutionizes high school computer science education in Korea with generative quiz creation, in-class AI tutors for Python coding and written support, and post-class learning analytics dashboards—key features required by Korea's AI digital textbook certification. Instructors can upload materials to auto-generate practice problems, receive step-by-step debugging hints for student code, or get contextual prompts for open-ended written questions.

In practice, the Korean computer science AI Digital Textbook implements a teaching–learning cycle that integrates these core functions into a seamless Pre–During–Post model. Before class, students complete AI-generated quizzes to activate prior knowledge. During class, AI tutors offer just-in-time debugging assistance and scaffolded prompts for written tasks, guiding learners from error diagnosis to solution exploration without revealing answers. After class, the analytics dashboard visualizes individual and class-wide performance, enabling educators to assign targeted remediation or enrichment activities.

Students report tangible benefits from AI tutor interactions: “AI hints helped me learn to debug on my own,” and “the follow-up questions deepened my explanations.” Teachers observe that learners who struggled with syntax errors now apply systematic debugging strategies, while more advanced students explore extension tasks suggested by the AI. These cases demonstrate how real-time, context-aware feedback transforms both confidence and competence.

Underpinning these experiences is a prompt engineering framework: every student prompt bundles the original problem statement and the learner's draft response, which is sent to the server for template-based processing. The template layers in problem comprehension, learner-level assessment, motivational language, and additional clues—mimicking the nuanced guidance of a human tutor. This design also embeds social–emotional support messages to sustain engagement and self-efficacy.

Team Monolith's national digital transition teacher training program has upskilled over 13,000 educators in workshops covering AI textbook integration strategies, prompt-based feedback design, and data-driven lesson planning with analytics dashboards. Participants actively share classroom success stories and refine best practices collaboratively.

This presentation will detail the teaching–learning model leveraging the Korean computer science AI Digital Textbook's certified core functions, illustrate student AI tutor usage cases, showcasing how this platform accelerates Korea's digital education innovation and serves as a global exemplar of AI-integrated pedagogy.

Session 4: Inclusive Education

Title: Vocational Challenges in Teaching Students with Learning Difficulties: A Comprehensive Literature Review

Author: Manal Aldaihani

Title: Key Functions for the Development of a Health Promotion Programme in the Primary School

Author: Veronique Regnier

Title: Korean STEAM Education: The Unique Role of Emotion

Author: Nicole E. Anderson

Title: Assessment of Road Accessibility Barriers of People with Disabilities in Nigeria

Authors: Aondoseer A. Atoo, Folake O. Akintayo

Vocational Challenges in Teaching Students with Learning Difficulties: A Comprehensive Literature Review

Manal Aldaihani

PAAET

Kuwait

Abstract

Teaching students with learning difficulties is a highly specialized and demanding vocation, requiring educators to navigate complex pedagogical, emotional, and institutional challenges. This study presents a comprehensive literature review examining the vocational difficulties faced by teachers in this field. Drawing on a diverse body of peer-reviewed research, the study explores critical themes, including insufficient access to specialized training, the implementation of individualized education plans (IEPs), high levels of occupational stress, and the emotional and psychological impact of managing diverse student needs.

The review also highlights systemic barriers, such as limited institutional support and a lack of resources, which exacerbate these challenges. In addition, it examines the role of collaborative practices, professional development, and policy interventions in mitigating these difficulties. By synthesizing existing research, this study provides a nuanced understanding of the professional struggles faced by teachers of students with learning difficulties and underscores the need for comprehensive support systems.

The findings call for actionable strategies to enhance teacher preparation and retention, improve job satisfaction, and ultimately create more effective and inclusive learning environments. This research contributes to the broader discourse on educational equity and the professional sustainability of specialized educators.

Key Functions for the Development of a Health Promotion Programme in the Primary School

Veronique Regnier
Jean Monnet University
France

Abstract

The Alliance for Health Research project aims to implement a comprehensive approach to health promotion in schools and in the community and to evaluate its impact. This project involves 130 French primary schools in the Auvergne-Rhône-Alpes region. It is an intersectoral project involving regional education authorities, the regional public health agency, local policy makers, researchers and local health promotion operators. The proposed intervention is co-constructed with the educational partners on the basis of a personalized diagnosis for each school.

From September 2019 to June 2024, primary school teachers received training, support and resources to develop a health-promoting environment at school, in line with the health-promoting school framework. At the same time, school representatives and local authority staff involved in extracurricular activities received training, support and resources in the same way. Quantitative and qualitative data on child health indicators and processes were collected to evaluate the intervention in 2021, 2023 and 2024. This poster focuses on the results of the implementation mechanism based on 61 semi-structured interviews. The sample includes district education advisors, school principals, school assistants, extracurricular officers, extracurricular program and activity managers, and health promotion project managers.

Collective work dynamics, practice exchanges and the professional identity of health promotion are essential elements of implementation. Lack of time is perceived as a structural barrier. Communities emphasized that perceived legitimacy and competencies were the cornerstones of their engagement. These results enable us to define “key functions” that are useful for the implementation of the program in new academies.

Korean STEAM Education: The Unique Role of Emotion

Nicole E. Anderson
Brigham Young University, USA

Abstract

STEAM education is STEM Education (Science, Technology, Engineering, and Mathematics) infused with the Arts. Western science has largely prioritized STEM education, with STEAM education beginning later than STEM but increasingly making inroads [5]. In contrast, in 2011 Korea made STEAM, and not STEM, education a national priority, as there is a long history of arts and aesthetics in Korean culture [1], [8].

1. Introduction

This inclusion of the Arts into STEM to produce STEAM education introduces an emotional component into the research equation. Creative design and emotional experience are key pillars of the Korean STEAM education framework, based on guidance from STEAM education's founder, Georgette Yakman [16] (see Figure 1). Therefore, it is not surprising that Korean STEAM education curriculum designers and educational researchers have made emotion a priority.

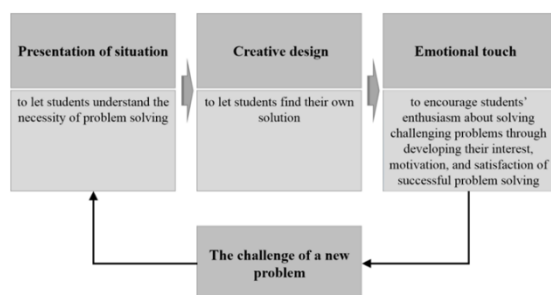


Figure 1. Pillars of Korean STEAM Education Framework

Note: Reprinted from Yakman and Lee (2012)

The importance of Social-Emotional Learning (SEL) principles to STEM and STEAM education are noted in both the West [7] and the East [9]. SEL is a better predictor of school performance than even IQ [4]. According to the USA Committee for Children, SEL involves:

- a) maintaining cooperative relationships,
- b) making responsible decisions,
- c) managing strong emotions and exerting self-

- control,
- d) communicating clearly,
- e) solving problems effectively,
- f) recognizing emotions in oneself and others, and
- g) having empathy.

However, recent Western systematic reviews and syntheses of literature [6], [13] have noted the lack of assessment of SEL measures in Western STEM and STEAM as part of a lack of assessment of these measures in Western education generally. According to Garner and Gabitova [6], “very little is known about whether learning in STEM is enhanced by the infusion of SEL content or whether SEL is a consequence of well-designed STEM education” (p. 1). And in a rare meta-data synthesis of STEAM education research, Rikoon et al. [13] noted that “as a proportion of the number of times SEL spaces were raised [in STEAM education research], they were assessed relatively infrequently”.

In addition to being a world leader in STEAM education [1], Korea is one of the most technologically advanced countries in the world [12] and has one of its best education systems [15]. Korean also has a strong evaluative arm, with researchers, teachers, and even students contributing to STEAM education research. Yet analysis of rigorous, sophisticated research designed studies of Korean STEAM education by Western researchers is rare, as titles and abstracts of Korean scientific studies are written in English, but the rest of these articles are written in Korean [1].

2. Method

Reading titles and abstracts written in English of studies found in Korea's national science database, KoreaScience, using the search term “STEAM education” and searching for the word “emotion” in abstracts, I found $n = 26$ STEAM education studies on emotion from 2012 to 2024. I then translated these entire articles from Korean to English using Google Translate. Information was gathered in terms of a) positive and negative research findings of STEAM studies involving emotions in quantitative and qualitative studies, and b) emotion in STEAM studies measured as either IVs or DVs.

3. Results

Benefits of emotionally engineered lessons included increased or high levels of satisfaction, interest, and enjoyment of science and STEAM, similar to what has been found in the West. In contrast, negative findings of emotionally enhanced STEAM education lessons included a) a risk of conflicts in team activities [18], b) students with emotional issues having a hard time on a STEAM fieldtrip that emphasized emotional experiences and quiet time [19], and c) a student on the same STEAM field trip feeling that “there was not enough time to think deeply” where the goal was to create an emotional experience to facilitate creative thinking [19]. A qualitative study of expert STEAM teachers reported similar concerns with rigorously studying emotion. As one teacher asked, “If emotional experience is an attitude that arises during the learning process, when will immersion occur? How can we understand emotional experiences that occur other than at the learning stage?”. Another teacher stated, “If one of the elements of emotional experience is the challenge of a new task, it can only be confirmed with consecutive assignments. Since this is a one-time class, confirmation is not possible” [14].

In terms of the second research aim, $n = 3$ studies manipulated emotion as an independent variable, or IV (“emotionally engineered” STEAM lesson versus traditional STEAM or science lesson). All three studies demonstrated increases in relevant outcomes from pre to post, including science interest [2], emotional intelligence [11], and science-related attitudes [10]. One program involved reading poetry to students for six sessions. Several cognitive outcomes were assessed but the only demonstrated improvements were in logical thinking [3]. Another STEAM study targeting SEL improvement [17] found increases in relationship skills but not SEL outcomes.

4. Discussion

Koreans are very systematic and evaluative in their everyday educational efforts and are more rigorous in their quantitative research designs in STEAM education studies than Western researchers [1]. This culture has resulted in evaluations from teachers and even students assessing the effectiveness of STEAM education lessons that are not generally found in the West. In Western research, researchers may already be adding “emotional touches” informally to their STEAM education lessons but not operationally defining lessons as “emotionally engineered.” Additionally, the lack of improvement of SEL outcomes in this study corresponded to a lack of quality of SEL program elements found by Western researchers [7] and point

to the dire need for more research. Western STEAM education researchers may therefore benefit from studying emotion more rigorously, such as by asking more in-depth qualitative questions on emotion or by conducting experimental research with IVs that involve emotional engineering of lessons.

5. References

- [1] Anderson, N. (2021). A systematic review of STEAM education research: Comparing American and Korean studies. *Academia Letters*, Article 1039. DOI: 10.20935/AL1039.
- [2] Bae, Y., Park, P., Moon, G.S., Yoo, I., Kim, J.W., Lee, H. and Shin, S. (2017). An instructional design of STEAM programs using 3D Printer and analysis of its effectiveness and satisfaction. *Journal of the Korean Association of Information Education*, 21(4), 475-486.
- [3] Bae, S. H. and Kim, H. S. (2017). The effect of STEAM program on the logical thinking ability of middle school students. *Journal of the Korean Society of Earth Science Education*, 10(1), 17-25.
- [4] Duckworth, A. L. and Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16, 939-944. <https://doi.org/10.1111%2Fj.1467-9280.2005.01641x>
- [5] Galligan, A. (2014). U.S. Workforce policy builds up ‘STEAM’. *CultureWork!* 18(2), 1-6.
- [6] Garner, P. W. and Gabitova, N. (2022). *Social and Emotional Learning and STEM-Related Education*. Routledge.
- [7] Garner, P. W., Gabitova, N., Gupta, A., and Wood, T. (2018). Innovations in science education: Infusing social emotional principles into early STEM learning. *Cultural Studies of Science Education*, 13(4), 899-903.
- [8] Kim, H. and Chae, D.H. (2016). The development and application of a STEAM program based on traditional Korean culture. *Eurasia Journal of Mathematics, Science and Technology Education*, 12(7), 1925-1936.
- [9] Lee, S.K. and Bong, M. (2017). Social and emotional learning as a solution for adolescent problems in Korea. *Social and emotional learning in Australia and the Asia-Pacific Perspectives, programs and approaches*, 233-251.
- [10] Lee, S. Y. and Lee, H. C. (2013). The effects of science lesson applying STEAM education on the

creativity and science related attitudes of elementary school students. *Journal of Korean Elementary Science Education*, 32(1), 60-70.

[11] Oh, D. J., Bae, J. H. and Park, S. H. (2016). The effects of science based enrichment STEAM gifted program on creative thinking activities and emotional intelligence of elementary science gifted students. *Journal of Korean Elementary Science Education*, 35(1), 13-25.

[12] Pranav, S. K. and Katiyar, A. (2018). Transformation of South Korea from a third world agrarian economy to a technologically advanced first world country. *International Journal for Advanced Research and Development*, 3(1), 166-170.

[13] Rikoon, S., Finn, B., Jackson, T., and Inglese, P. (2018). Crosscutting literature on STEAM ecosystems, expectancy value theory, and social emotional learning: A metadata synthesis. *ETS Research Report Series*, 2018(1), 1-15.

[14] Son, M. and Jeong, D. (2019). Limits of STEAM education and its improvement alternative: Based on the viewpoints of STEAM expert teachers. *Journal of the Korean Association for Science Education*, 39(5), 573-584.

[15] Sorensen, C. W. (1994). Success and education in South Korea. *Comparative education review*, 38(1), 10-35.

[16] Yakman, G. and Lee, H. (2012). Exploring the exemplary STEAM education in the US as a practical educational framework for Korea. *Journal of the Korean Association for Science Education*, 32(6), 1072-1086.

[17] Lee, H., Lee, S.-Y., Jung, J., Lee, S., Choi, E., Kwak, E.-R., ... Chang, H. (2022). Effects of Out-of-school STEAM Programs Based on Social-Emotional Learning. *Journal of Korean Elementary Science Education*, 41(4), 740–753. DOI: 10.15267/KESES.2022.41.4.740.

[18] Choi, Y., Hong, S. (2015). Effects of STEAM Lessons Using Scratch Programming Regarding Small Organisms in Elementary Science-Gifted Education. *Journal of Korean Elementary Science Education*. 34. DOI: 10.15267/keses.2015.34.2.194.

[19] Yoon, S., Lee, Y., Han, C. et al. Usefulness of the Patient Health Questionnaire-9 for Korean Medical Students. *Acad Psychiatry* 38, 661–667 (2014). DOI: 10.1007/s40596-014-0140-9.

Assessment of Road Accessibility Barriers of People with Disabilities in Nigeria

Aondoseer A. Atoo¹, Folake O. Akintayo²
Joseph Sarwuan Tarka University¹
University of Ibadan²
Nigeria

Abstract

This study assessed the road accessibility barriers of people with disabilities within a Nigerian city (Makurdi metropolis) using interviews and road network/features observation. A total of 180 persons with disabilities were interviewed with 6 road networks studied for two weeks where accessibility features were considered. The most common disability around were the physically challenge which was 53.3% with the least being those with hearing impairment. The people agree that road accessibility is a barrier that hinders their free movement. Physical barriers, transport service barriers, barriers related to Policy and technological barriers were the most dominant. Physical barriers could be resolved through inclusive design and construction of road facilities that will consider the disabled persons, with government and policymakers enacting laws and policies to improve road accessibility for people with disability.

Keywords: Accessibility, people with disabilities, accessibility barriers, inclusive transport, Makurdi metropolis

1. Introduction

Accessibility can be seen as the possibility of any individual to reach any destination of choice without hinderance or barriers [1]. Therefore, road accessibility can be said to be the ability of a person or individuals to reach and make use of places without structural, physical, or any other barrier on urban roads, irrespective of the person's physical ability.

Globally, transportation accessibility for people with disabilities (PWDs) is an issue as it affects mobility [2]. Though much have been done through the United Nations Convention on the Rights of Persons with Disabilities (CRPD) and the Sustainable Development Goals (SDGs) emphasizing the importance of inclusive transportation systems [3], this is not the case in low-income country like Nigeria. Public transport system lacks physical features to accommodate people with disabilities and mobility impairments, such features are ramps, elevators, tactile guidance for the visually impaired and bus stop facilities [4], [5]. This study assesses the road

accessibility barriers experienced by PWDs within the study area in other to improve on them to meet world standard.

2. Methodology

This research adopted a mixed-method approach that combines interviews with people with disabilities and observational survey where the accessibility challenges or barriers of PWDs on road infrastructure was observed. This observation and interviews were conducted for two weeks in Makurdi metropolis of Nigeria.

3. Results and Discussion

The survey, through interviews, interacted with 180 persons within the urban city of Makurdi in Nigeria. The disability types includes hearing impairment, physically challenged, visual impairment and albino were interviewed. The most dominant interacted with were the physically challenge (those in wheelchairs, tricycle chairs, and those on crutches), which made up 53.3% with hearing impairment making the least at 13.9%. They confirm that apart from safety on the road, accessibility to good roads and transport system is a major barrier on the road (See figure 1).

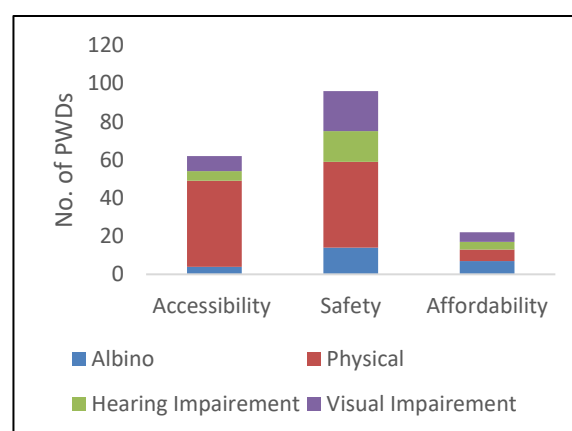


Figure 1. PWDs Barriers on the road

Six (6) major roads within the study area were studied to observe the accessibility challenge available, taking note of global practices. Bus stops were studied for PWDs reservations, ramp available at bus stops and road side walk pavement, tactile pavement, traffic signals, public transport service and government policies.

Details of physical accessibility barriers include:

- 6 bus stops with shields have no ramps at bus stops for easy access of people in wheelchairs. No reserved space for wheelchair persons which confirms with Akintayo and Adibeli [4] and Zhang et al., [5].
- Lack of paved pedestrian paths on some roads with paved ones not having ramp for easy access on and off the path.
- Some bus stops have no shields which aligns with Akintayo et al., [6].
- Road side market activities on pedestrian walkways.
- Lack of traffic signal on roads with the few traffic signal available not considering pedestrian crossing.
- Lack of crossing points on the road.
- Road median not having a well-defined, designed, and constructed crossing point.

The physical accessibility barriers can be improved through improve inclusive design and construction of road with PWDs factors in mind (design for all).

The transport system does not have specialized services for PWDs and the buses has no lowering doors, no space for wheelchairs in the bus. Use of specialized vehicles will correct this. Technological barriers include the lack of assistive device in mobility, lack of traffic signals with sounds ability. This can be made available to help tackle this issue.

Lack of policy implementation backing the use of roads for PWDs and weak implementation is also a barrier. This agrees with Mogaji and Nguyen [7].

4. Conclusion

This study shows that people with disability face several barriers in their day-to-day activities which range from physical, technological, transport service, policies, including design and construction that need to be looked into to improve the mobility of PWDs within the urban centres in Nigeria. Inclusive transportation system and network design to include PWDs will improve their mobility as well as improve technology, good policies will go a long way in helping PWDs.

5. References

[1] Gaglione, F., Gargiulo, C., Zucaro, F. (2019). Elders' quality of life. A method to optimize pedestrian accessibility to urban services. *TEMA Journal of Land Use, Mobility and Environment* 12 (3), 295–312. DOI:10.6092/1970-9870/6272.

[2] Ghosh, S., Chakraborty, L., Basu, K. (2022). Intersectional discrimination against women and girls with disabilities in educational opportunities in India. *World Development Perspectives* 26, 100412. DOI:10.1016/J.WD P.2022.100412.

[3] Imam, I., and Abdulraheem-Mustapha, M. A. (2016). Rights of People with Disability in Nigeria: Attitude and Commitment. *African Journal of International and Comparative Law*, 24(3), 439–459. DOI:10.3366/AJICL. 2016.0163.

[4] Akintayo, F. O., and Adibeli, S. A. (2022). Safety performance of selected bus stops in Ibadan Metropolis, Nigeria. *Journal of Public Transportation*, 24, 100003. DOI:10.1016/J.JPUBTR.2022.100003.

[5] Zhang, Y., Farber, S., Young, M., Tiznado-Aitken, I., and Ross, T. (2024). Travel behaviour differences among people with disabilities: A cluster analysis of accessible taxi users before and during the COVID-19 pandemic. *Journal of Transport and Health*, 35. DOI:10.1016/J.JTH.2023.101 753.

[6] Akintayo, F. O., Adibeli, S. A., Oyewale, P. K., and Olawuyi, O. A. (2019). Bus stops characteristics in selected local government areas in Ibadan Metropolis. Presentation at the 1st International Conference on Engineering and Environmental Sciences (ICEES), November 5–7. <https://ocs.uniosun.edu.ng>. (Access Date: 30 June 2025).

[7] Mogaji, E., and Nguyen, N. P. (2021). Transportation satisfaction of disabled passengers: Evidence from a developing country. *Transportation Research Part D: Transport and Environment*, 98. DOI:10.1016/J.TRD.2021. 102982.

Session 5: Cyber Security

Title: Secure Software Brownfield Engineering – Physical & Logical Partition Identification
Authors: Jadish Singh, Aspen Olmsted

Title: Cyber Immune Model for Combating DDOS Attacks
Authors: Abereowo Oladimeji, Akinwomi Akintoba, Kuboye Bamidele, Alese Boniface K.

Title: Cybercrime Detection and Analysis Model for Evidence Collection
Authors: O. Y. Ogunlola, I. D. Bolarinwa, B. K. Alese

Title: Enhancing Individual Cyber Resilience: A Holistic Approach to Online Safety and Security
Authors: Arwa Alsaiani, Steven Furnell, Helena Webb

Secure Software Brownfield Engineering – Physical & Logical Partition Identification

Jadish Singh, Aspen Olmsted
School of Computer Science and Data Science
Wentworth Institute of Technology
 USA

Abstract

Understanding the physical and logical architecture of existing "brownfield" software systems is a significant challenge, particularly when documentation is scarce or outdated. This paper presents a novel approach for reverse-engineering the physical partitioning of legacy PHP applications by analyzing their runtime execution traces. We leverage the comprehensive trace functionality of XDebug, a widely used PHP extension, to capture detailed records of function calls, file inclusions, and module interactions during various operational scenarios. Our methodology involves parsing these XDebug trace logs to identify recurring patterns of inter-file and inter-directory dependencies. By observing which PHP files and directories are frequently accessed together or sequentially during specific use cases, our approach infers logical groupings that correspond to physical partitions, such as modules, sub-systems, or architectural layers. This automated identification of cohesive code units provides crucial insights into the system's underlying structure, even without prior knowledge of its design. The derived physical partitions serve as foundational elements for subsequent architectural modeling, enabling developers and architects to construct more accurate and up-to-date representations of the brownfield system. This understanding is vital for effective maintenance, refactoring, and strategic planning for modernization. This research demonstrates a pragmatic, tool-assisted method for extracting architectural knowledge from opaque legacy PHP codebases, facilitating more informed software evolution.

Keywords: cyber-security; software engineering; secure software development

1. Introduction

The vast majority of software systems in active use today fall under the category of "brownfield" applications. These are mature, often critical, systems that have evolved over extended periods, typically lacking the pristine architectural clarity of newly

developed "greenfield" projects. A pervasive challenge with brownfield software is the significant gap in current and accurate documentation regarding its physical architecture. Over time, original design documents become outdated, key personnel move on, and the organic growth of the codebase obscures the initial structural intent. This lack of understanding about how the software is physically partitioned into modules, components, or layers severely impedes maintenance, refactoring efforts, and any strategic planning for modernization or migration. Without a clear map of the system's internal organization, even seemingly minor changes can introduce unforeseen complexities or vulnerabilities, making the evolution of these vital assets a high-risk endeavor.

Traditional methods for understanding brownfield architectures often involve laborious manual code reviews, which are time-consuming, expensive, and prone to human error, especially in large codebases. Static analysis tools can provide insights into code structure, but they may struggle to infer the dynamic relationships and actual runtime dependencies that define effective physical partitions. There is a pressing need for automated, pragmatic approaches that can effectively reverse-engineer the physical architecture of legacy systems, providing developers and architects with actionable insights into their structure.

This paper addresses this critical need by proposing a novel methodology for identifying physical partitions within brownfield PHP applications. Our approach leverages the powerful, yet often underutilized, trace functionality of XDebug, a standard debugging and profiling tool for PHP. XDebug's ability to record every function call, file inclusion, and module interaction during a script's execution provides an unprecedented level of detail about the system's runtime behavior. By dynamically observing which parts of the codebase are activated and interact during specific operational scenarios, we can infer the logical groupings that constitute physical partitions.

The core of our methodology lies in the intelligent parsing and analysis of these XDebug trace logs. We

focus on identifying recurring patterns of inter-file and inter-directory dependencies. For instance, if a set of files within a particular directory consistently co-execute or interact in sequence across various user actions, this strongly suggests that they belong to a cohesive physical partition or module. This automated inference of architectural units provides a data-driven basis for understanding the system's structure, even in the absence of explicit design documentation. The derived physical partitions serve as foundational elements for subsequent architectural modeling, enabling the creation of accurate and up-to-date representations that are indispensable for informed decision-making in software evolution. This research offers a practical, tool-assisted solution to a pervasive problem in legacy software management, facilitating more effective maintenance, refactoring, and modernization initiatives.

The paper is organized as follows. Section II describes the related work and the limitations of current methods. Section III describes a motivating example for our work. Section IV discusses the implementation of our parser. We conclude and discuss future work in Section V.

2. Related Work

The field of secure software engineering has seen significant advancements, particularly in the realm of greenfield development, where security can be integrated from the ground up. However, the unique challenges posed by brownfield systems necessitate a distinct focus. This section reviews existing literature pertinent to secure software development, with a particular emphasis on contributions that address the complexities of integrating security into established, often undocumented, codebases, as well as the role of dynamic analysis and visualization in this context.

A foundational aspect of secure software engineering is the proactive integration of security considerations throughout the software development lifecycle. Aspen Olmsted's seminal work, "Security Driven Software Development" [1], provides a comprehensive framework for embedding security into every phase of development, from requirements gathering to deployment and maintenance. This book emphasizes the importance of a security-first mindset and offers strategies for identifying and mitigating risks early. While primarily focused on new development, the principles outlined by Olmsted, such as threat modeling and secure coding practices, are equally relevant to brownfield remediation efforts. Our proposed methodology, which aims to improve understanding of existing brownfield code, directly supports the application of such security principles by making the implicit explicit.

Further elaborating on the importance of early security integration, Olmsted's paper, "Secure software development through non-functional

requirements modeling" [2], highlights the critical role of non-functional requirements (NFRs) in achieving software security. This work argues that security, often treated as an afterthought, should be explicitly modeled as an NFR from the outset. In this context, the use of formal specification languages like Object Constraint Language (OCL) and UML stereotypes can significantly enhance the precision and verifiability of security requirements [1]. For brownfield systems, where NFRs may not have been formally captured during initial development, our approach of generating UML Sequence Diagrams helps in reverse-engineering the system's behavior. By visualizing execution flows, it becomes possible to infer how security-related NFRs (e.g., access control, input validation) are currently being handled, or where they are conspicuously absent. This retrospective analysis can then inform the re-establishment of security NFRs and guide remediation efforts.

In a broader context, Olmsted's paper titled "Secure Software Development—Models, Tools, Architectures and Algorithms" [3] offers a comprehensive view of the various components necessary for robust and secure software development. This paper discusses multiple models for security, the tools that aid in analysis, architectural considerations for building secure systems, and the algorithms underlying security mechanisms. Our research aligns with this broader vision by introducing a practical tool-based approach (leveraging XDebug) to generate a specific model (UML Sequence Diagrams) that aids in understanding the architecture and behavior of brownfield PHP applications. The generated diagrams serve as a valuable input for applying various security models and algorithms, allowing analysts to trace the flow of data and control to identify vulnerabilities that might otherwise remain hidden within the complex logic of legacy code.

While existing literature extensively covers secure software development, a gap remains in practical, low-overhead methods tailored explicitly for thoroughly understanding the runtime behavior of brownfield applications for security purposes. Static analysis tools (e.g., SAST solutions) are effective at identifying patterns of vulnerabilities but often struggle with context and produce false positives in complex legacy code. Dynamic Application Security Testing (DAST) tools interact with running applications to identify vulnerabilities, but they typically operate at a higher level of abstraction and may not provide the granular, function-level insights that XDebug traces offer. Our work complements these existing tools by providing a detailed, visual representation of execution flow, which is particularly beneficial for manual security auditing, threat modeling, and understanding the impact of discovered vulnerabilities within the intricate tapestry of brownfield PHP applications. The re-purposing of

XDebug for this task offers a unique advantage, as it leverages a widely available and familiar developer tool, minimizing the learning curve and integration overhead for teams working with legacy PHP systems.

3. Motivating Example

Identifying Physical Partitions in SuiteCRM with Program Traces: To demonstrate the practical application of our methodology for identifying physical partitions, we again turn to SuiteCRM, a complex brownfield PHP application. Its extensive codebase and often undocumented structure make it an ideal candidate for reverse-engineering its physical organization. By analyzing XDebug traces of everyday user interactions, we can infer the underlying modularization and component boundaries, which are crucial for architectural understanding and future development efforts.

Consider the following common user scenarios within SuiteCRM, and how program traces can illuminate their underlying physical structure:

Scenario 1: User Creates a Contact

- **User Action:** A sales representative navigates to the "Contacts" module, fills in various contact details (e.g., name, email, phone number, address), and submits the form to save the new contact record.
- **Trace Insight for Partitioning:** When XDebug tracing is enabled for this operation, the trace log captures every PHP file that is included, required, or executes functions during the process. The key insight for physical partitioning comes from observing which files and directories are consistently activated together to complete the "Create Contact" operation. The trace would reveal:
 - **Core Module Files:** Calls to files within modules/Contacts/ (e.g., Save.php, Contact.php model files).
 - **Shared Utility Files:** Frequent interactions with files in common directories like include/ (for common libraries, helper functions, validation logic), data/ (for data access objects or database interaction layers), and potentially custom/ (for customizations).
 - **Framework Components:** Activation of core framework files, if SuiteCRM uses an underlying framework structure, revealing its interaction with the application-specific logic.

By aggregating these file accesses across multiple "Create Contact" operations, our parser can identify a

strong co-occurrence pattern, suggesting that modules/Contacts/, include/, and specific files within data/ collectively form the "Contact Management" physical partition. This helps to delineate the boundaries of this functional area within the sprawling codebase.

Scenario 2: User Schedules a Meeting

- **User Action:** A user accesses the "Meetings" module, enters details such as the meeting subject, time, date, duration, and invites participants (e.g., other users, contacts, leads), then saves the meeting record.
- **Trace Insight for Partitioning:** Tracing this scenario provides a rich set of file and directory interactions that highlight the "Meeting Management" partition and its dependencies. The trace would show:
 - **Meeting Module Files:** Extensive activity within modules/Meetings/ (e.g., Meeting.php, Save.php).
 - **Inter-Module Dependencies:** Crucially, it would show calls to files within modules/Users/ and modules/Contacts/ as participants are associated. This reveals the explicit runtime dependency of the "Meetings" partition on the "Users" and "Contacts" partitions.
 - **Calendar and Notification Logic:** If separate directories or files handle calendar integration or email notifications, the trace would expose their involvement, indicating distinct sub-partitions or shared service components.

The analysis of these traces allows us to group modules/Meetings/, along with specific interactions with modules/Users/ and modules/Contacts/, into a coherent "Meeting Scheduling" partition, demonstrating how different functional areas physically interact.

Scenario 3: User Creates an Invoice

- **User Action:** An accountant generates a new invoice through the "Invoices" module, associating it with a specific client (Account), adding various products or services, specifying quantities and prices, and saving the invoice.
- **Trace Insight for Partitioning:** This scenario involves complex financial logic and relationships, making it excellent for identifying the "Invoice Management" partition and its connections to other business entities. The XDebug trace would capture:
 - i. **Invoice Module Files:** Heavy usage of files within

within modules/Invoices/ (e.g., Invoice.php, Save.php, calculation logic files).

- ii. **Related Entity Modules:** Direct and frequent interactions with files in modules/Accounts/ (for client details) and modules/Products/ (for product information). This clearly delineates the physical components that form the "Invoice Processing" functionality.
- iii. **Reporting/Export Components:** If invoice generation involves creating PDFs or exporting data, the trace would reveal calls to dedicated reporting libraries or export utility files, potentially residing in separate directories (e.g., include/SugarPDF/).

By analyzing these co-occurring file accesses, our method can identify the "Invoice Management" partition, along with its strong physical coupling to the "Account Management" and "Product Catalog" partitions. This level of detail about physical organization is invaluable for understanding the codebase, planning refactoring efforts, or even migrating specific modules in a brownfield modernization project. The automated derivation of these partitions from runtime traces provides an objective and up-to-date view of the system's true physical architecture.

4. Parser Implementation

The successful identification of physical and logical partitions within brownfield PHP applications hinges on the robust and intelligent parsing of XDebug trace logs. This section details the implementation of our Java-based parser, specifically designed to extract and analyze file and directory access patterns from these traces, thereby inferring the system's physical modularization.

i. XDebug Trace File Format Revisited

As previously discussed, XDebug trace files (typically .xt format, computer readable format 1) provide a chronological, line-by-line record of execution events. For the purpose of physical partition identification, the most critical fields are:

- **File Name:** The absolute path to the PHP file being executed or included.
- **Type:** Indicates the event type, particularly 0 (function call) and 2 (include) which directly involve file execution.

Our parser focuses on these fields to build a comprehensive map of file and directory interactions during runtime.

ii. Parser Design for Partitioning Analysis

Our Java parser is structured to efficiently process large trace files and build a data model optimized for co-occurrence analysis. The architecture follows these steps:

iii. Efficient Trace File Reading:

Similar to the sequence diagram parser, a `BufferedReader` is used for line-by-line reading to handle potentially massive XDebug trace files without excessive memory consumption. Each line is processed sequentially.

iv. Event Filtering and File Path Extraction:

For each line, the parser first performs lexical analysis by splitting the line by tabs. It then filters events, primarily focusing on Type 0 (function/method calls) and Type 2 (file inclusions/requires). For these relevant events, the absolute File Name is extracted. This file path is the primary data point for our partitioning analysis.

v. Normalized Path Storage and Directory Mapping:

To effectively identify physical partitions, file paths need to be normalized. This involves:

- Removing the common base path of the application (e.g., `/var/www/html/suitedcrm/`). This allows relative paths to be compared and grouped.
- Extracting the directory component of each file path. For instance, `/var/www/html/suitedcrm/modules/Contacts/Save.php` would yield `modules/Contacts/`.

The parser maintains a data structure (e.g., a `Set` or `Map`) to store all unique normalized file paths and directories encountered during the trace.

vi. Co-occurrence Matrix/Graph Construction:

This is the central component for identifying partitions. As the parser processes the trace file, it builds a representation of which files and directories are accessed together within a short time window or within the scope of a single user action.

Sliding Window Approach: A simple yet effective method is to use a sliding window. For every `N` consecutive file access events in the trace, we consider all unique files and directories within that window to be "co-occurring."

Event-Based Grouping: A more precise approach is to group file accesses by the "top-level" user action or request. For example, all file accesses occurring between the start and end of the "User Creates a Contact" scenario are considered a single co-

occurrence set. This requires external knowledge of when a scenario begins and ends, or inferring it from the trace (e.g., based on the initial entry point PHP script).

The co-occurrence data can be stored in a co-occurrence matrix (where rows and columns represent files/directories, and cell values indicate frequency of co-occurrence) or, more flexibly, as an undirected graph. In the graph, nodes represent files or directories, and an edge exists between two nodes if they co-occurred. The weight of the edge can represent the frequency of co-occurrence.

vii. Clustering and Partition Identification:

Once the co-occurrence graph is constructed, standard graph clustering algorithms (e.g., Louvain method, Girvan-Newman algorithm, or even simpler connected components for a basic approach) can be applied. These algorithms identify dense subgraphs, where nodes (files/directories) within a subgraph are highly connected to each other but less connected to nodes in other subgraphs. Each such dense subgraph represents a strong candidate for a physical partition.

The output of this clustering step is a set of identified partitions, each consisting of a collection of normalized file paths and directories. For example, one cluster might contain `modules/Contacts/`, `include/validation.php`, and `data/ContactDAO.php`, suggesting a "Contact Management" partition.

5. Conclusions and Future Work

Some challenges and considerations for future work include:

- **Defining "Co-occurrence":** The definition of "co-occurrence" (e.g., window size, event grouping) significantly impacts the identified partitions. Experimentation and domain knowledge are crucial to fine-tune this.
- **Noise Reduction:** XDebug traces can be very verbose, including many generic framework files or common libraries that are part of every operation. The parser needs mechanisms to filter out such "noise" (e.g., by excluding files from known common libraries or by weighting co-occurrences based on their specificity to a module).
- **Granularity of Partitions:** The level of granularity (e.g., module-level, sub-system level) depends on the clustering algorithm and its parameters. The parser should allow for configurable thresholds or parameters for the clustering algorithm.
- **Visualization of Partitions:** While the parser outputs the identified partitions programmatically, a visualization component (e.g., using Graphviz or

a custom drawing library) would be beneficial to represent the discovered partitions and their inter-dependencies visually. This could be a separate post-processing step.

By implementing this specialized Java parser, we can systematically analyze XDebug trace data to reveal the implicit physical partitions within complex brownfield PHP applications. This automated approach transforms raw execution logs into actionable architectural insights, providing a foundation for more effective software evolution and maintenance. diagramming library or tool, enabling the automated generation of visual sequence diagrams that are invaluable for understanding and securing brownfield PHP applications. The modular design ensures that the parser can be extended to support additional XDebug features or adapted for different trace formats in the future.

6. References

- [1] Olmsted, A. (2024). *Security-Driven Software Development: Learn to analyze and mitigate risks in your software projects*, Birmingham, UK: Packt Publishing.
- [2] Olmsted, A. (2010). Secure software development through non-functional requirements modeling, in *Proceedings of the 2010 International Conference on Software Engineering Research and Practice (SERP)*.
- [3] Olmsted, A. (2012). Secure Software Development—Models, Tools, Architectures and Algorithms. *Journal of Software Engineering and Applications*, pp. 743-750, 2012.

Cyber Immune Model for Combating DDOS Attacks

Abereowo Oladimeji, Akinwomi Akintoba, Kuboye Bamidele, Alese Boniface K.
Cyber Security Department, Federal University of Technology, Akure, Nigeria

Abstract

The advent of Industry 4.0 has brought forth a transformative wave, revolutionizing manufacturing processes through the integration of digital technologies, automation, and data-driven decision-making. The foundation of Industry 4.0 lies in interconnectedness—machines communicating seamlessly, systems sharing data autonomously, and processes optimizing themselves in real-time. However, the proliferation of interconnected devices, the emergence of the internet of things (IoT), and the extensive reliance on cloud computing have amplified the vulnerability of industrial systems to cyber threats. While these advancements bring unprecedented efficiency, they also introduce a complex network of vulnerabilities. Cyber threats, ranging from ransomware attacks to data breaches and system manipulations, pose substantial risks, potentially causing severe operational disruptions, financial losses, and jeopardizing sensitive information. DDoS attacks have increasingly become a significant threat to online services, disrupting business operations and service availability. In 2020, Amazon Web Services experienced a massive attack, peaking at 2.3 tbps. Also in 2023, hackers targeted the U.S. government with similar attacks to draw attention to political causes. In 2024, Cloudflare mitigated the largest recorded DDoS attack, peaking at 3.8 tbps.[9] This susceptibility demands a paradigm shift in cybersecurity strategies, urging the creation of a robust and adaptive cybersecurity immune system that mimics the human immune system, tailored explicitly for the dynamic landscape of industry 4.0.

Keywords: Hacktivists, Cytokines, Reinforcement Learning, Industry 4.0, Cyber-Physical Systems (CPS)

1. Introduction

The advent of Industry 4.0 has brought forth a transformative wave, revolutionizing manufacturing processes through the integration of digital technologies, automation, and data-driven decision-making.[2] The foundation of Industry 4.0 lies in interconnectedness that is machines communicating seamlessly, systems sharing data autonomously, and processes optimizing themselves in real-time.[2] However, the proliferation of interconnected devices, the emergence of the internet of things (IoT), and the extensive reliance on cloud computing have amplified the vulnerability of industrial systems to cyber

threats[6]. While these advancements bring unprecedented efficiency, they also introduce a complex network of vulnerabilities. Cyber threats, ranging from ransomware attacks to data breaches and system manipulations, pose substantial risks, potentially causing severe operational disruptions, financial losses, and jeopardizing sensitive information. DDoS attacks have increasingly become a significant threat to online services, disrupting business operations and service availability. In 2020, Amazon Web Services experienced a massive attack, peaking at 2.3 tbps. Also in 2023, hackers targeted the U.S. government with similar attacks to draw attention to political causes. In 2024, Cloudflare mitigated the largest recorded DDoS attack, peaking at 3.8 tbps. This susceptibility demands a paradigm shift in cybersecurity strategies, urging the creation of a robust and adaptive cybersecurity immune system that mimics the human immune system, tailored explicitly for the dynamic landscape of industry 4.0.[1].

The human immune system is a complex network of organs, cells, proteins, and tissues that work together to defend the body against harmful pathogens such as bacteria, viruses, fungi, and other foreign invaders[8]. Its primary function is to recognize and eliminate these pathogens while distinguishing them from the body's own healthy cells. The defense mechanism employed by the immune system is in 3 basic levels

The Physical Barrier: The body's first line of defense, this includes physical barriers such as the skin and mucous membranes that prevents pathogens from entering the body [7].

Innate Immune System: This is the body's immediate, non-specific defense mechanism. It includes cells like macrophages, neutrophils, natural killer cells, proteins like complement proteins. They quickly respond to pathogens, neutralizing or engulfing them.

Adaptive Immune System: This system provides a more targeted and specific response to pathogens. It involves specialized cells (T and B lymphocytes) that recognize specific pathogens or antigens. B cells produce antibodies that bind to pathogens, while T cells destroy infected cells directly.

The concept of a cybersecurity immune system draws an analogy from the human immune system, aiming to defend against digital threats in a manner akin to the body's defense against infections. Hence its functionalities include:

i. **Continuous Monitoring:** Similar to the body's constant surveillance for potential threats, cybersecurity systems continuously monitor networks, devices, and data for any anomalies or suspicious activities.

ii. **Threat Detection and Analysis:** Just as the immune system identifies foreign invaders, cybersecurity systems employ various tools like intrusion detection systems (IDS), artificial intelligence (AI), and machine learning algorithms to detect and analyse potential cyber threats.

iii. **Adaptive Response Mechanisms:** Similar to the immune system's ability to adapt to new pathogens, a cybersecurity immune system responds dynamically to cyber threats, utilizing automated responses and adaptive defenses to neutralize or contain attacks.

2. Related works

Some of the related works that were reviewed during the course of the research are listed below:

An Immunology-Inspired Network Security Architecture by Quan *et al.* 2020 The paper aims to enhance defense capability by consistently contesting with invaders in a real environment and actively mutating/evolving attack strategies. Additionally, the authors present a reference architecture designed based on the principles derived from immunology. The authors studied the immune system in humans, and drew parallels between the immune system and network security system. They derive three design principles: maintaining the balance between security and availability, achieving situation-awareness, fast response through community cooperation among heterogeneous nodes, and enhancing defense capability through consistently contesting with invaders in a real environment and actively mutating/evolving attack strategies. The research explores the concept of a cyber immune system as a vision for cyber security, which includes automatic defense mechanisms based on incomplete attribution, continuous monitoring, pattern recognition, and the application of a set of rules designed to isolate or destroy attackers. The paper provided a detailed process of designing a cyber immune system, but the proposed system has not been tested in a real environment scenario to actually see if the system would work as it is on paper.

We also reviewed another paper titled Work-in-Progress: The Cyber-Physical Immune System by Zhou *et al.* 2021. The objective was to enhance the reliability and security of CPSs by deploying the CPIS on top of a conventional CPS. Also to alert the user of any threats or anomalies detected by the CPIS. The CPIS comprises a collection of hardware and software elements deployed on top of a conventional CPS. The CPIS utilizes data-driven techniques to identify

threats and adapts to the changing environment. It collects data from the conventional CPS, alerts the user of any threats or anomalies, and deploys threat-mitigation strategies.[5] The CPIS comprises a collection of hardware and software elements deployed on top of a conventional CPS. The CPIS utilizes data-driven techniques to identify threats and adapts to the changing environment. It collects data from the conventional CPS, alerts the user of any threats or anomalies, and deploys threat-mitigation strategies. The CPIS could not determine what to keep and communicate to the main node. Also, the system is not able to take countermeasures against attack invasion, just like the immune system does.

One of the other reviewed works was Cyber Immunity A Bio-Inspired Cyber Defense System. The paper aims to develop a cyber immune system that is capable of learning attack with unknown signatures, mimic the AIS using machine learning to detect attacks based on what was learnt. The researcher developed a cyber immune system that was trained to learn what normal network looks like, once trained will be able to calculate the probability that a certain defiant is malicious.[7] The cyber immune system detected new, unknown cyber-attacks and also provided a powerful defense mechanism for the cybersecurity systems after being trained to recognize this threat pattern. Their system was excellent in detecting unknown signatures but is not capable of detecting unauthorized use of administrator rights since this is considered normal behavior.

3. Objectives

The specific objectives of the research are to:

- i. design a cyber defense model using human immune approach.
- ii. design a model to identify, pre-empt, and neutralize DDoS Attacks.
- iii. implement the designs in i and ii; and
- iv. evaluate the developed models using some standard metrics

4. System Design

In order to draw up a model, we need to understand how the immune system works. We have summarized the processes of the human immune system into six steps as seen below:

- **Physical Barriers:** The body's first line of defense includes physical barriers such as the skin and mucous membranes that prevent pathogens from entering the body.
- **Innate Immune System:** This is the body's

immediate, non-specific defense mechanism. It includes cells like macrophages, neutrophils, natural killer cells, and proteins like complement proteins. They quickly respond to pathogens, neutralizing or engulfing them.

- **Adaptive Immune System:** This system provides a more targeted and specific response to pathogens. It involves specialized cells (T and B lymphocytes) that recognize specific pathogens or antigens. B cells produce antibodies that bind to pathogens, while T cells destroy infected cells directly.
- **Memory Cells:** Both the innate and adaptive immune systems create memory cells that "remember" specific pathogens. If the same pathogen invades the body again, these memory cells enable a quicker and more robust response, often preventing illness or reducing its severity.
- **Lymphatic System:** This system includes lymph nodes, lymphatic vessels, spleen, and thymus. It plays a crucial role in the circulation of immune cells, filtering lymph (a clear fluid) to remove pathogens and cellular debris.
- **Cytokines and Chemical Mediators:** These signaling molecules are released by immune cells to communicate with each other, regulating immune responses and inflammation.

We will also imbibe the following principles for building a healthy cyber immune system:

- i. Hygiene: a good hygiene precedes a strong and healthy cybersecurity immune system. This will help provide a quality check on the individual components of the ecosystem to ensure they are fully functioning so the whole defense infrastructure doesn't break down when attacked. This calls for regular security audits, investment in improving current defense, processes and protocols, and continuous training to keep security practitioners current at all times.
- ii. Visibility: Establishing visibility into all systems at the network-packet and metadata level via a single pane of glass will allow security defense to contain the threat effectively.
- iii. Threat detection: just like how a virus spreads across organs in the human circulatory system, malware often targets lateral movement between vital computing systems and devices in the corporate network, the malware will attempt to evade detection but with adequate threat identification capabilities, behavioral anomalies will be detected. The suspicious activity is triangulated and compared against known bad behavior to identify the intruder.

- iv. Prediction: next step will be to include the development of predictive capabilities to identify motives. The motive behind attacks can often reveal repeatable behavioral patterns that will allow the security infrastructure to respond appropriately.
- v. Response: once the motive is uncovered, the appropriate incident response can be taken to contain the threat and for further analysis.

From the above stated principles and the summarized functionalities of the human immune system, we are able to design our model in a way that it is able to identify, pre-empt, and neutralize DDoS Attacks.

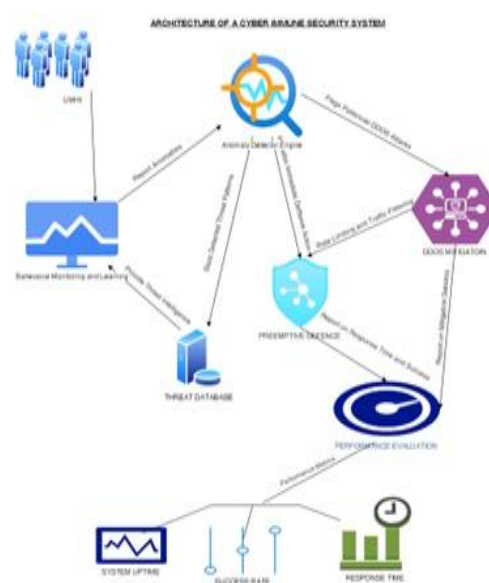


Figure 1. Architecture of a cybersecurity immune system

Our model is able to perform its operation through the following means slated below:

Behavioral Monitoring and Learning monitors patterns of system behavior, adapts based on the latest threats, and refines detection models.

we will setup Machine learning pipeline using TensorFlow framework to learn from past events.

The system will be modeled using reinforcement learning, where it receives feedback (rewards) based on its detection accuracy [3].

Let the state S represent the system state, and the action α represent the defense actions taken. The goal is to learn a policy $\pi(s)$ that maximizes the expected reward:

$$\pi(s) = \arg \alpha^{max} E[R(s, \alpha)]$$

This allows the system to adjust its defense strategies dynamically based on the evolving threat landscape and continuously improves the Anomaly Detection Engine based on new intelligence.

Anomaly Detection and Behavioural Analytics

Monitors network traffic and system behaviors to detect anomalies. We make use of k-means clustering to classify legitimate vs. anomalous activities.

Clustering (Machine Learning):

Assume $X \in R^d$ represents the system features (e.g., network traffic, CPU load).

Apply k-means clustering to classify system behavior:

$$d(x, c_k) = \|x - c_k\|^2$$

where c_k is the centroid of cluster K , and anomalies are points that are far from any cluster centroid:

If $\min_k d(x, c_k) > \text{threshold}$, mark x as an anomaly.

This allows real-time detection of abnormal network traffic.

Preemptive Defense Layer

Responds immediately to detected anomalies, blocks malicious traffic, and implements rate-limiting, firewall rules and traffic shaping.

Rate-Limiting:

The system monitors incoming traffic and calculates the rate of requests per second:

$$\text{Rate}(t) = \frac{\text{Number of Requests}}{\Delta t}$$

If the rate exceeds a threshold (e.g., 1000 requests per second), the defense mechanism is triggered to drop packets:

If $\text{Rate}(t) > \text{Threshold}$, drop incoming packets.

Firewall Rule Application:

Let x_i represent an incoming packet, where features include source IP, destination port, and protocol.

The firewall implements rules f such that:

$$f(x_i) = \begin{cases} 1 & \text{if } x_i \text{ matches known malicious patterns} \\ 0 & \text{otherwise} \end{cases}$$

The firewall blocks the packet if $f(x_i) = 1$.

DDoS Mitigation System

The system identifies and neutralizes DDoS attacks by analyzing traffic patterns. DDoS attacks often result in abnormal traffic surges, therefore tools like Cloudflare DDoS Protection, AWS Shield, or Akamai Kona. Uses traffic rate-limiting, IP filtering, and traffic scrubbing to differentiate between legitimate and malicious requests.

Volumetric DDoS Detection:

Let $T(t)$ represent the traffic volume at time t . We compute the traffic rate:

$$R(t) = \frac{T(t)}{\Delta t}$$

Compare the current traffic rate with historical data (e.g., mean and standard deviation):

$$R_{\text{baseline}} = \frac{1}{n} \sum_{i=1}^n R(t_i)$$

If $R(t) > R_{\text{baseline}} + 3\sigma$, trigger the DDoS mitigation system to block traffic or apply rate-limiting.

Feedback Loop (Memory and Adaptation)

The feedback loop continually improves the system by adjusting based on detected attacks. It feeds data back into the Threat Database and Behavioral Monitoring to continually improve the system's defenses.

Updating Threat Database

After an attack is detected, the system extracts the features x_{new} and adds them to the threat database if classified as a new attack.

Compute the **centroid** of attack vectors S_{attack} and update it:

$$C_{\text{new}} = \frac{1}{S_{\text{attack}}} \sum_{i=1}^{S_{\text{attack}}} x_i$$

Use the new centroid for future comparisons.

Metrics and Evaluation Module

This module evaluates the effectiveness of the system using performance metrics like mean-time-to-detection (MTTD) and false-positive rate (FPR).

Mean Time to Detect (MTTD):

Let t_{detect} represent the detection time for each attack.

MTTD is the average detection time:

$$MTTD = \frac{1}{N} \sum_{i=1}^N t_{\text{detect}}$$

False Positive Rate (FPR):

Let FP represent the number of false positives and N the total number of legitimate events. FPR is computed as:

$$FPR = \frac{FP}{N}$$

5. Implementation

The research is still ongoing, in order to be able to learn the behavior of the attack, and the motives of the attacker, we created a honeypot on the network. This we have been using to train our model to be able to differentiate between a normal behavior and an anomaly on the network. Even though the research is still ongoing, our model is growing in accuracy over time. Once the model has been trained then to differentiate anomalies, we will proceed to the detection and mitigation stage.

6. Expected contribution to knowledge

The research is still ongoing and hopes to achieve a system that is to establish a cyber immune model that mimics the human immune system to defend against DDOS, learn from previous attacks, preempt the attacks and adapt the system to be able to neutralize future attacks.

7. References

- [1] Schrom et al. (2023). Challenges in cybersecurity: Lessons from biological defense systems. *Mathematical Biosciences*, 362, 0–2. DOI: 10.1016/j.mbs.2023.109024.
- [2] Petrenko, S. (2022). Developing a cybersecurity immune system for Industry 4.0. In *Developing a cybersecurity immune system for Industry 4.0*. DOI:10.1201/9781003337874.
- [3] Kalra, P. (2022). Mathematical models for machine learning techniques. 9(11), 57–64.
- [4] National Cyber Security Centre. (2022). *Active cyber defense - The fifth year*.
- [5] Pang, et al. (2021). The cyber-physical immune system: Work-in-progress. *Proceedings - 2021 International Conference on Embedded Software, EMSOFT 2021*, 43–44. DOI:10.1145/3477244.3477621.
- [6] Pourmoafi, S., and Vidalis, S. (2021). Bio-cyber operations inspired by the human immune system. *European Conference on Cyber Warfare and Security*, 534–541. DOI:10.34190/EWS.21.089.
- [7] System, A. B. C. D. (2018). *Cyber immunity - A bio-inspired cyber defense system*. April 2017. DOI:10.1007/978-3-319-56154-7.

[8] Zhang et al. (2018). Key issues in designing cyber security proxy gateways for digital substation non-immune bay layers. *2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems, CYBER 2017*, 1309–1312. DOI: 10.1109/CYBER.2017.8446509.

[9] NCS. (n.d.). Six steps to building a healthy cyber immune system. NCS.<https://www.ncs.co/en-sg/knowledge-centre/articles/six-steps-to-building-a-healthy-cyber-immune-system/> (Access Date: 28 December 2024).

Cybercrime Detection and Analysis Model for Evidence Collection

O. Y. Ogunlola, I. D. Bolarinwa, B. K. Alese

*Department of Cybersecurity, Federal University of Technology, Akure
Nigeria.*

Abstract

The Cybercrime Detection and Analysis Model (CDAM) is a tool developed to improve detection and digital forensic capabilities in the digital era. It uses advanced machine learning techniques to analyze cyber threats like malware and phishing attacks. The model is trained on publicly available datasets and a web-based application, allowing real-time evidence analysis. This advancement in cybersecurity practices is crucial for law enforcement and professionals.

Keywords: Digital Era, Cybercrime, Threats, Analysis Model, Government and Integrity

1. Introduction

In an era where digital transformation is ubiquitous, the specter of cybercrime looms large, posing significant threats to individuals, organizations, and nations alike. The imperative to safeguard digital assets and maintain the integrity of information systems has never been more critical. This research introduces the Cybercrime Detection and Analysis Model (CDAM), a pioneering initiative aimed at enhancing the capabilities of cybercrime investigation units in detecting, analyzing, and collecting digital evidence. Cybercrime encompasses illegal activities carried out through computers or other communication devices, which aim to instil fear, cause distress, or inflict damage on individuals and property. Cybercrimes fall into two main types: computer-assisted crimes, such as child pornography, fraud, money laundering, and cyberstalking; and computer-focused crimes, like hacking, phishing, and website defacement [1].

The first cybercrime incident, in which computer codes were replicated, took place in the 1960s. Many fraud and forgery cases were reported after 1970 when a bank teller at New York's Union Dime Savings Bank embezzled over \$1.5 million from customer accounts. A creeper virus was developed by Bob Thomas in 1971 to infect the systems of the Advanced Research Project Agency Network (ARPANET) was the first network with packet-switching technology and the TCP/IP protocol [9].

In the digital age, cybercrime has emerged as a significant threat to individuals, businesses, and governments worldwide. The proliferation of internet

connectivity and sophisticated digital tools has given rise to a new breed of crime that transcends traditional physical boundaries. Cybercrimes encompass a wide range of illegal activities; the detection and analysis of cybercrimes are paramount for the security and integrity of digital infrastructures. As such, developing effective models for cybercrime detection and evidence collection is crucial for law enforcement and cybersecurity professionals.

2. Related Works

[10] developed a digital forensic framework for IoT environments to improve investigation techniques and tackle unique challenges in IoT devices. They tested the framework with IoT scenarios, contributing to knowledge. [5] did a deep learning-based solutions for detecting cyberbullying with high accuracy and efficiency. Both studies highlight the need for continuous updates to handle evolving cyberbullying tactics and the need for further validation with diverse IoT devices. [11] worked on a framework for accountability and auditability in cloud computing to improve trust and security. [13] provided a comprehensive list of security considerations for cloud-supported IoT environments, aiming to address security challenges in IoT-cloud integration. However, these methods require further validation and may require updates to address evolving privacy concerns. The frameworks and methods contribute to knowledge but may require continuous updates and varying effectiveness with different configurations.

[3] developed SAIRF, a system using fuzzy min-max neural networks to improve cyber-attack detection and understanding. The research was motivated by the increasing complexity and frequency of cyber-attacks and the need for better methods to recognize and understand attack intentions. [4] compared deep learning and support vector machine algorithms for detecting port scan attempts, providing insights into their strengths and weaknesses. [2] focused on detecting and predicting phishing websites using classification mining techniques to enhance the accuracy and efficiency of phishing detection systems. However, limitations include varying effectiveness with different datasets and the need for

continuous updates to address new phishing tactics. [6] developed a digital forensic model for cloud computing to address cybercrime investigations. The model aims to improve the effectiveness of investigations in cloud environments.

3. System Design

This research design focuses on a Cybercrime Detection and Analysis Model for Evidence Collection, utilizing advanced machine learning techniques to enhance detection and facilitate evidence collection. The model addresses the need for sophisticated detection systems to address modern cyber threats, as traditional methods are insufficient. Machine learning offers significant advantages in the realm of cybersecurity due to its ability to analyze vast amounts of data and uncover patterns that are not immediately apparent through manual inspection (see Figure 1).

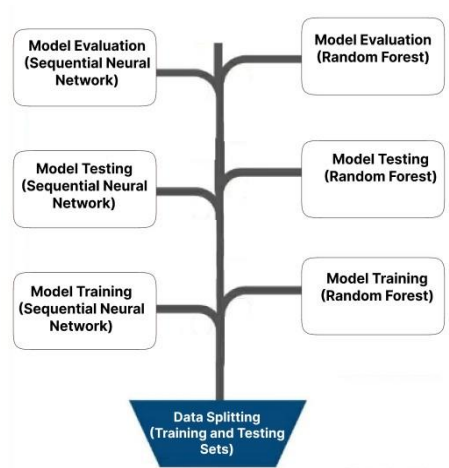


Figure 1: An illustration of the Cybercrime Detection and Analysis Model development process

Data preprocessing is a crucial step in developing a Cybercrime Detection and Analysis Model for Evidence Collection, transforming raw data into a suitable format for analysis. It addresses issues like missing values, noise, and inconsistencies. Data cleaning involves identifying and correcting errors, including handling missing values due to corruption, manual entry errors, or incomplete data collection processes.

3.1. Deployment of the Web Application

A web application is developed using Flask, a lightweight WSGI framework in Python, to streamline the Cybercrime Detection and Analysis Model. The application will allow users to upload evidence files and use trained models for detection and evidence collection. Flask's modular design reduces overhead and simplifies maintenance, making it ideal for deploying machine learning models. The application

will extract features from files using tools like nmap, scapy, and tldextract, and analyze using pre-trained models to identify potential cyber threats (see Figure 2).

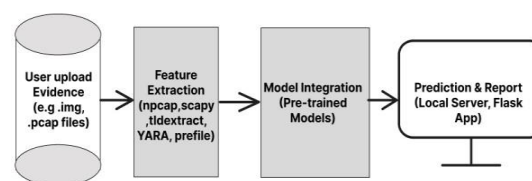


Figure 2. Architecture for Web Application Deployment of Cybercrime Detection and Analysis Model for Evidence Collection.

9. Implementation

The Cybercrime Detection and Analysis Model was developed using a high-performance development machine with an Intel Core i9-10900K processor, 32 GB DDR4 RAM, and 1 TB NVMe SSD. The machine was designed to efficiently process large datasets, perform complex computations, and train sophisticated machine learning models. The software environment included Python 3.8 for programming, machine learning libraries like scikit-learn and TensorFlow/Keras, data processing tools like Pandas and NumPy, visualization libraries like Matplotlib and Seaborn, and a lightweight web application framework called Flask. Cybersecurity tools like Scapy, tldextract, and YARA were used for network packet manipulation, phishing detection, and malware research. These tools ensured efficient processing, testing, and deployment of the system.

Data collection was the foundational step in building the machine learning model. For the Cybercrime Detection and Analysis Model, three high-profile datasets from Kaggle were utilized: the Network Intrusion Dataset, the Malware Dataset, and the Phishing URL Dataset. Each dataset served a distinct purpose and collectively covered a wide spectrum of cyber threats (see Figures 3, 4 and 5).

```
# Load the dataset
file_path = r'C:\Users\Owner\cybercrime_detection\data\malware\Kaggle-data.csv'
df = pd.read_csv(file_path)
```

Figure 3. A screenshot of malware dataset loaded on notebook

```
if __name__ == "__main__":
    # Define the file path
    data_dir = r'C:\Users\Owner\IDS\data\raw\CSV'
    file_name = "KDDTrain+.csv"
    file_path = os.path.join(data_dir, file_name)

    # Load the dataset
    df = load_data(file_path)
```

Figure 4. A screenshot of network intrusion dataset loaded on notebook



Figure 5. Preprocessed Datasets

5. Result

This section provides a detailed evaluation of the Random Forest and Feedforward Neural Network models' effectiveness across three cybercrime detection tasks: malware detection, network intrusion detection, and phishing detection. Key evaluation metrics considered include accuracy, precision, recall, and F1-score. The analysis aims to deliver a comprehensive understanding of each model's performance in detecting various cyber threats, highlighting their strengths and weaknesses. Models were evaluated based on the following metrics, which provided insights into their performance and reliability.

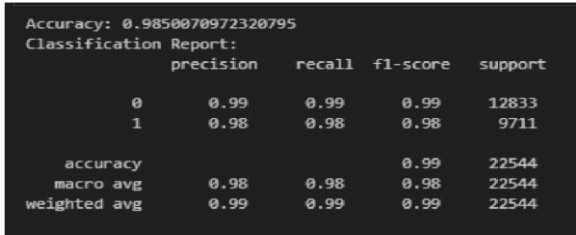


Figure 6. Results of evaluation metrics for Random Forest Network Intrusion Detection Model

Machine learning techniques for detecting cybercrimes, such as network intrusions, malware, and phishing attempts. The integration of Random Forest and Feedforward Neural Networks into a webbased application proved effective in processing and analyzing real-world data. The models achieved high accuracy and were able to detect various types of cyber threats with considerable effectiveness.

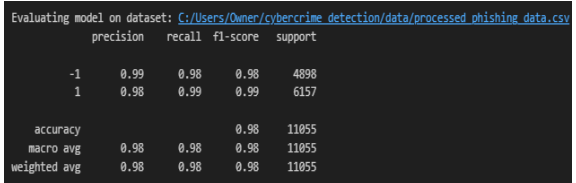


Figure 7. Results of evaluation metrics for Random Forest Phishing Detection Model

The project highlights the potential of machine learning in cybersecurity, providing a foundation for further developments in the detection and prevention of cybercrimes. The use of a user-friendly web

application for deploying the models ensured that the system was practical and accessible for real-world applications. Overall, the research has contributed valuable insights into the application of machine learning for enhancing cybersecurity measures.

6. Reference

[1] Al-Mousa, M. (2019). Analyzing Cyber-Attack Intention for Digital Forensics Using Case-Based Reasoning. ArXiv, abs/2101.01395. DOI: 10.30534/ijatse/2019/9 2862019.

[2] Al-Diabat, M. (2017). Detection and Prediction of Phishing Websites using Classification Mining Techniques. International Journal of Computer Applications, 147, 5-11. DOI: 10.5120/IJCA2016911061.

[3] Ali, A., and Mohammed, M. (2017). SAIRF: A similarity approach for attack intention recognition using fuzzy minmax neural network. J. Comput. Sci., 25,467-473. DOI: 10.1016/j.jocs.2017.09.007.

[4] Aksu, D., and Aydın, M. (2018). Detecting Port Scan Attempts with Comparative Analysis of Deep Learning and Support Vector Machine Algorithms. International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), 77-80. DOI: 10.1109/IBIGDELFT.2018.8625370.

[5] Iwendi, C., Srivastava, G., Khan, S., and Maddikunta, P. (2020). Cyberbullying detection solutions based on deep learning architectures. Multimedia Systems, 29, 1839 - 1852. DOI: 10.1007/s00530-020-00701-5.

[6] Hemdan, E., and Manjaiah, D. (2018). Digital Forensic Approach for Investigation of Cybercrimes in Private Cloud Environment. Advances in Intelligent Systems and Computing. DOI: 10.1007/978-981-108639-7_3.

[7] Yanbo Wu, Dawei Xiang*, JiangMing Gao and Yun Wu. (2018). Research on investigation and evidence collection of cybercrime Cases. IOP Conf. Series: Journal of Physics, IOP Publishing. DOI:10.1088/1742-6596/1176/4/042064

[8] Shisheng, Y. (2014). Countermeasures and Suggestions for the difficulty in investigation and evidence collection of network crime cases. Employee legal world: under, 0(9):38-38.

[9] Karie, N., and Venter, H. (2015). Taxonomy of Challenges for Digital Forensics. Journal of Forensic Sciences, 60. DOI: 10.1111/1556-4029.12809.

[10] Kebande, V., and Ray, I. A. (2016). Generic Digital Forensic Investigation Framework for Internet of Things (IoT). IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud), 356-362. DOI: 10.1109/FiCloud.2016.57.

[11] Ko, R., Lee, B., and Pearson, S. (n.d).Towards Achieving Accountability, Auditability and Trust in Cloud Computing. 432-444. DOI: 10.1007/978-3-642-22726-4_45.

[12] Lindell, Y., and Pinkas, B. (2021). Privacy Preserving

Data Mining. Journal of Cryptology, DOI: 10.1007/s00145-001-0019-2.

[13] Singh, J., Pasquier, T., Bacon, J., Ko, H., and Eyers, D. (2016). Twenty Security Considerations for Cloud Supported Internet of Things. IEEE Internet of Things Journal, 3, 269-284. DOI: 10.1109/JIOT.2015.2460333.

Enhancing Individual Cyber Resilience: A Holistic Approach to Online Safety Security

Arwa Alsaiani, Steven Furnell, Helena Webb
School of Computer Science
University of Nottingham
United Kingdom

Abstract

Nowadays, interest in organisational cyber resilience has risen, but little is known about people's resilience against cyber threats. Individual cyber resilience in different environments is essential because most individuals may encounter threats or risks when using technology. Defending against cyberattacks, recovering from them, learning from past attacks is critical for individuals. However, there is a lack of studies on people's cyber resilience, partly because it is unclear how to define this term that should encompass both cybersecurity online safety. The goal of this paper is to characterise the linkage between cybersecurity online safety by exploring definitions related to security safety. Also, this report explores how variants of safety security are defined across various sources to clarify the overlaps inconsistencies between the two concepts. Using visual tools such as word clouds, keyword analysis helps identify both common ground conceptual gaps. Additionally, it emphasises the importance of channels in promoting public awareness of online safety cybersecurity.

Keywords: Cybersecurity, Safety, Threat, Risk, Individual, Resilience

1. Introduction

In the digital age, the internet has become an integral part of daily life, encompassing education, shopping, work. Additionally, as reliance on the global network increases, so do concerns about cybersecurity online safety. According to a study by [1], 47.9 million adults in the UK used computers, tablets, smartphones to access the internet. Cybersecurity protects systems, devices, people, information/data from cyberattacks, while internet safety protects people from online harm. However, people are more exposed to cyber risks vulnerabilities due to their growing reliance on online platforms. This highlights the need to address individuals' lack of security knowledge understanding of the Internet. Cyberattacks can have significant consequences, including financial loss,

reputational damage, the compromise of sensitive information. Furthermore, interest has developed in

Cyber resilience in recent years, representing an organisation's ability to manage the impacts of cyberattacks recover rapidly from an event, has seen a notable increase in study interest. However, cyber resilience focuses on how systems organisations can be cyber resilient. According to Joinson [2], there is a lack of studies on what constitutes cyber resilience in individual technology users who may confront cybersecurity problems at home or outside of work. Furthermore, individuals need to develop their cyber resilience to better face cyber threats. This study will highlight the conceptual relationship between security safety terminology by analysing their definitions comprehensively, aiming to identify overlaps distinctions.

2. Literature Review

The digital world presents challenges for individual users navigating online environments. This literature review examines existing research on personal cyber resilience, including cybersecurity online safety, with a focus on individual users rather than organisations. The review is guided by specific questions that help frame our understanding of how these concepts are currently treated in academic literature identify a notable gap regarding individual cyber resilience as an integrated concept.

This review systematically analyses keyword searches for terms such as cyber resilience, cybersecurity, online safety, threats, attacks.

The concept of cyber resilience in current literature is focused on organisational issues. A critical finding of this review is that the term "cyber resilience" is applied at the organisational or system level rather than the individual level. According to Linkov et al. [3], cyber resilience is defined as "the ability of a system to prepare for, absorb, recover from, adapt to adverse conditions, stresses, attacks, or compromises on cyber resources." Similarly, Björck et al. [4] characterise it as an organisation's capacity to deliver intended outcomes

continuously despite adverse cyber events.

While cybersecurity online safety are often regarded as distinct domains with separate focuses:

- i. Cybersecurity literature emphasises technical protection, system integrity, defence against cyber-attack [5]. Topics commonly include network security, encryption, authentication mechanisms, malware protection [6].
- ii. Online safety literature focuses on behavioural aspects, psychological impacts, social aspects of digital engagement [7]. This includes cyberbullying, privacy risks, content exposure[8]. Furthermore, few studies provide an integrated framework that addresses the cybersecurity online safety dimensions of individual digital protection. At the same time, some researchers acknowledge connections between these domains [9].

Moreover, individuals encounter various threats, such as cybersecurity online safety concerns, including malware ransomware attacks aimed at personal devices data [10]. Furthermore, phishing social engineering exploit human psychology [11]. Research also highlights non-technical threats, including social behavioural issues, that affect individual well-being, such as online cyberbullying across age groups [12]. The literature indicates that these threats are often addressed in isolation rather than as interconnected challenges that require comprehensive approaches. Technical cybersecurity research tends to downplay behavioural factors, while online safety research may overlook technical vulnerabilities.

The literature review has identified that existing research typically considers cyber resilience as an organisational or system-level concept rather than one focused on individuals. Additionally, cybersecurity online safety are often treated as separate rather than integrated domains. The review has identified the diverse threats individuals face online, alongside common safety recommendations, underscoring the need for a more cohesive approach. A key finding of this literature study is the lack of a clear definition of personal cyber resilience. This term should encompass both the technical aspects of cybersecurity the behavioural elements of online safety at an individual level. Bridging these technical behavioural domains is crucial for individuals' digital well-being in increasingly complex online environments.

3. Methodology

This research investigates how key concepts of “safety” “security” are defined communicated by various organisations websites. The methodology is divided into three stages: data collection, visualisation, keyword analysis, evaluation.

3.1. Data Collection

Definitions of “safety” include online internet, E-safety, digital safety. Additionally, “security” encompasses cybersecurity, information security, computer security, IT security, compiled from various official reputable sources, including governmental agencies, non-governmental organisations, educational institutions, prominent public-facing websites.

3.2. Word Cloud Visualisation

Keywords were extracted from the definitions to identify common themes terminologies. A word frequency analysis was conducted to determine which words appeared most frequently. To enhance the interpretability of the keyword analysis, word clouds were generated for both “safety” “security” definitions. This step helped identify overlapping distinct concepts used in conceptualising safety versus security. Following the keyword analysis evaluation, the research examined whether the identified keywords were clearly explained within the sources. This was done by reviewing each definition evaluating it based on whether the keywords were explicitly defined or left ambiguous. Supporting examples or explanations were provided. The language used was understandable for non-experts.

This assessment aimed to determine how effectively each source communicates safety security concepts to the public.

4. Safety: Definition Scope

The internet can enable different forms of online harm or abuse, such as bullying or grooming. Online safety, internet safety, E-safety, digital safety are often used interchangeably.



Figure. 1. Word clouds for (a) Online Safety, (b) Internet Safety, (c) e-Safety (d) Digital Safety

Table 1 establishes a set of definitions of these safety synonyms from different sources. Moreover, Figure 1

Table 1. Definitions of variants of 'safety' terms

Online Safety	
Source	Definition
SWFGL What is Online Safety? SWGfL	"protects the people using them from harm by the devices networks (therefore third parties) through awareness, education, information technology."
Learning.Learning.co m What is Online Safety: Definition & Basics Learning.com	"is the ability to underset recognize threats that exist on the internet have the skills knowledge to avoid these threats. "
Online Safety Act www.gov.uk/governm ent/publications/onlin e-safety-act- explainer/online- safety-act-explainer	"The Online Safety Act 2023 (the Act) is a new set of laws that protects children adults online."
Safeguarding Network https://safeguarding. network/content/saf eguardingresources/o nline-safety	"Is being aware of the nature of the possible threats that you could encounter whilst engaging in activity through the internet."
Internet Safety	
Source	Definition
SwellCSP https://www.swelcsp. org.uk/key- safeguarding- issues/exploitation- 2/internet-safety/	"Is trying to be safe on the internet."
Cumberl Safeguarding Children Partnership https://cumbriasafegu ardingchildren.co.uk	"Is the act of staying safer online, includes being aware of the risks associated with your online activity employing a few strategies to prevent or avoid these risks".
E-Safety	
Source	Definition
Edinburgh Learns Digital. https://edinburg	"Internet safety is about teaching our learners to be safe, showing them how to be smart online, being kind to others while interacting online"
Volunteer Now https://www.voluntee rnow.co.uk/app/uploa ds/2022/05/E-Safety- Factsheet-2.pdf	"As the safe responsible use of technology."
Gov.au https://www.esafety.g ov.au/about-us/what- we-do	"Helps people of all ages to underset how to use online services platforms as safely as possible get help if they are harmed"
Digital Safety	
Source	Definition
The International Journal Scientific Advances https://doi.org/10.515 42/ijscia.v2i5.21	"Is a branch of cyber security that deals with people the levels of online comfort, convenience, privacy. It may be regarded as the act of staying safer online"
Safe Online https://safeonline.glo bal/what-is-digital- safety-/	"Involves preventing minimising harm in the online environment. This includes moderating illegal or harmful content, promoting responsible platform design governance, empowering users to customise their digital experiences".

Table 2. Definitions of variants of 'security' terms

Cybersecurity	
Source	Definition
National Initiative for Cybersecurity Careers Studies https://niccs.cisa.gov/cy bersecurity-career- resources/vocabulary	"The activity or process, ability or capability, or state whereby information communications systems the information contained therein are protected from /or defended against damage, unauthorized use or modification, or exploitation".
University of Cambridge Glossary of information cyber security terms IT Help Support	"Ability to protect or defend the use of cyberspace from cyber-attacks"
Iso.org ISO/IEC 27032:2023(en), Cybersecurity — Guidelines for Internet security	"Safeguarding of people, society, organizations nations from cyber risks"
NCSC https://www.ncsc.gov.uk/ search?q=cyber%20secu rity	"The protection of devices, services networks - the information on them - from unauthorised access, theft or damage".
Information Security	
Source	Definition
IT Governance. Information Security vs Cyber Security: The Difference IT Governance	"Is a superset of cyber security. It aims to protect <i>all</i> information, including Digital information, Hard-copy information; Unwritten information/knowledge".
Microsoft.com. https://www.microsoft.com/ en- gb/security/business/security -101/what-is-information- security-infosec	"is a set of security procedures tools that broadly protect sensitive enterprise information from misuse, unauthorized access, disruption, or destruction."
Nist.gov. https://csrc.nist.gov/glos sary/term/INFOSEC	"The protection of information information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide confidentiality, integrity, availability".
Gov.uk. https://www.gov.uk/service- manual/technology/securing -your-information	"is the technologies, policies practices you choose to help you keep data secure"
UK Cyber Security Council https://www.ukcybersecurit ycouncil.org.uk/glossary/	"The regime of keeping one's organisation's data safe away from theft prying eyes".
Computer Security	
Source	Definition
Acronis. (2022). What is computer security? - Definition, Protection, Tips, Cyberthreats	"Is the activity of preventing unwanted access, theft, damage, interruption to computer systems, networks, data"
IT security	
Source	Definition
Cisco Threat Response Overview (2018)	"Is a set of cybersecurity strategies that prevents unauthorised access to organisational assets such as computers, networks, data"
Ibm.com https://www.ibm.com/think/ topics/it-security	"is the practice of protecting an organization's IT assets—computer systems, networks, digital devices, data—from unauthorized access, data breaches, cyberattacks other malicious activity."

provides visual insight into the most common keywords associated with online safety, internet safety, E-safety, digital safety using a Word Cloud tool.

5. Security: Definition Scope

After determining the keywords in online, internet, E-safety definitions, the next step is to examine how different sources address Cyber security, Information security, computer security, IT security definitions, as shown in Table 2. Additionally, figure 2 visually represents. The most frequent words related to security, as defined in the Word Clouds.



Figure 2. Word clouds for (a) Information Security, (b) IT Security, (c) Computer Security (d) Cyber Security

6. Safety security keywords Analysis

Table 3 highlights the safety-related keywords investigated in the sources listed. Most sources cover or explain the keywords "risk," "content," "prevent." Some sources may also refer to risk as a threat or harm. Additionally, the table indicates whether they are simply mentioned (M) or more fully defined (D). Also, Table 4 shows security keywords in each source, which support explain to people. Threat prevention are among the clarified keywords in several sources. Also, some keywords, such as unauthorized, defend, access, are not covered in these sources, as shown in the table. Additionally, the table indicates whether they are simply mentioned (M) or more fully defined (D).

7. Discussion

This indicates that some sources found that security safety focus only on instructions that help to protect users from encountering risks or threats. However, they do not sufficiently address the risks that threaten safety security. There are no clear instructions on how individuals should respond to specific security threats or safety risks. Additionally, the security keyword "Defend" is used without any explanation, so these sources do not focus on people's

responses to threats. This makes theoretical knowledge insufficient, as users need actionable steps to help them handle threats risks effectively.

On the other hand, Cyber incidents may significantly influence our daily lives. Thus, developing our abilities to assess create individual cyber resilience is crucial. The relationship between cybersecurity online safety represents a comprehensive approach to build personal cyber resilience. According to NIST Special Publication 800-160 Volume 2, cyber resilience functions as an overarching concept that integrates cybersecurity measures online safety practices, offering a comprehensive framework for organisational protection Ross et al. [13]. These practices cover technical personal defence, creating a robust holistic definition against digital threats. It is important to integrate security safety to enhance cyber resilience build through ongoing awareness learning. [14] supported that one of the main elements of the CR-SAT is learning, which acknowledges that in a cyber-resilient organisation, adverse events may be used as opportunities for learning. Together, these elements create a comprehensive framework that enables individuals to anticipate threats more effectively, prevent them, respond to them.

8. Awareness Guided Channels for End-Users

End users learn safety security guidelines from various channels, including advertisements, notifications, websites, apps, campaigns that contain security advice. Additionally, individuals can be classified based on how they find or receive information, into people actively searching for information or people encountering information as part of their day-to-day activities. These two types can be referred to as active or passive methods. The passive method can reach a broader audience but may suffer from lower levels of understanding or engagement. [15] suggests that passive methods, such as email or newsletters, are often insufficient for educating users. Not everyone may understand or know how to apply the information. Therefore, there is a need for active methods their integration with passive methods to enhance effectiveness. Significant effort is made to encourage people to follow this guidance or advice. However, an important consideration is whether the advice provided to people by these channels is consistent across different sources. Also, little is known about the quality of this advice: Does it help people increase their cybersecurity or online safety awareness. Does it aid in recovery from incidents. To address these questions, this research must conduct an observational study to identify the most effective routes for the public.

9. Conclusion

To summarise, digitisation has made cyber concerns a

Table 3. Prominent keywords presented in the sources of safety

Source	T	R	D	C	P	H	Observation Threat (T), Risk(R), Devices (D), Content (C), Prevent (P), Harm (H).
SWFGL							Does not cover any of the keywords
Learning.Learning.com					M		Presents "prevent" that the skills that users should learn can help prevent threats such as not sharing personal information, not clicking links, or opening attachments from strange people, to protect against Malware.
Online Safety Act				D	D		It provides types of content with explanations, such as Illegal content content that is harmful to children. Also, it explains "prevent", which means that children must be prevented from accessing harmful content should be given age-appropriate access to Priority content.
Safeguarding Network		M	M	M		M	Mentions some keywords, risk, or harm, provides types such as content, conduct, contact, commerce. Gives content examples, for instance, fake news, racism, misogyny, self-harm. Also included are devices such as smart speakers, wearables (e.g., smart watches), toys with voice or image recognition, robots, drones, other mechanical toys, smart TVs, games consoles.
SwellCSP							Does not cover any of the keywords
Cumberl Safeguarding Children Partnership					M		Mentions keywords "prevent" as tips for users from risk, such as monitoring your child's internet use, keeping a track of their browser history, choosing strong passwords to secure information Keep your privacy settings high ensure you are up to date with all the latest security software.
Edinburgh Learns Digital.							Does not cover any of keywords
Volunteer Now		M		M			Provides types of risk content
Gov.au					M		Mentions prevention could be raising public awareness about online safety issues, empowering individuals communities through training education, producing free, evidence-based, easy-to-use information for general public.
The International Journal Scientific Advances					M		Provides "prevent" ways to protect people such as using strong passwords backing up data regularly.
Safe Online							Does not cover any of the keywords

Table 4. Prominent keywords presented in sources of security

Source	T	U	P	D	A	D	Observation Threat (T), Unauthorized(U), Prevent(P), Defend(D), Access(A), Devices (D)
National Initiative for Cybersecurity Careers Studies							The source does not cover any of the keywords, it also does not explain any of them clearly
University of Cambridge	D						The source provides a threat definition does not cover any of the other keywords
Iso.org	D						The Source defines threats threat types, such as social engineering, privacy attacks, malware (hacking), spyware, other potentially unwanted software.
NCSC							The source does not cover any of the keywords.
IT Governance.							The source does not cover any of the keywords.
Microsoft.com.							Mentions types of threats
Nist.gov.							The source does not cover any of the keywords.
Gov.uk.							The source does not cover any of the keywords.
UK Cyber Security Council							Provides threat definition.
Acronis. (2022).			M			M	Provide "prevent" as a practice users can learn to protect themselves, mention some examples of devices such as a mobile phone tablet
Cisco Threat Response Overview (2018)	M					M	Gives some examples of threats devices
Ibm.com							Provides types of threats, prevent devices

concerns a significant part of our everyday lives. Cyber incidents can impact these individuals, making it essential to enhance our ability to assess develop individual cyber resilience. This paper has characterised the linkage between cybersecurity online safety by exploring definitions related to security safety. In addition, it highlighted key terms, differences, commonalities in safety security using word clouds. Additionally, it examined indirect methods for raising cybersecurity online safety awareness among the general population in everyday contexts.

10. References

- [1] Ofcom. (2023). Online Nation. www.ofcom.org.uk/siteassets/resources/documents/research--data/online-research/online-nation/2023/online-nation-2023-report.pdf (Access Date: 6 January 2025).
- [2] Joinson, A.N., Dixon, M., Coventry, L. Briggs, P. (2023). Development of a new 'human cyber-resilience scale'. *Journal of Cybersecurity*, 9(1), p. tyad007.
- [3] Linkov, I., Eisenberg, D.A., Plourde, K., Seager, T.P., Allen, J. Kott, A. (2013). Resilience metrics for cyber systems. *Environment Systems Decisions*, 33, pp.471-476.
- [4] Björck, F., Henkel, M., Stirna, J. Zdravkovic, J. (2015). Cyber resilience—fundamentals for a definition. In *New contributions in information systems technologies* (pp. 311-316). Cham: Springer.
- [5] Von Solms, R. Van Niekerk, J. (2013). From information security to cyber security. *Computers & Security*, 38, pp.97-102.
- [6] Furnell, S. Clarke, N. (2012). Power to the people? The evolving recognition of human aspects of security. *Computers & Security*, 31(8), pp.983-988.
- [7] Livingstone, S. Smith, P.K. (2014). Annual research review: Harms experienced by child users of online mobile technologies: The nature, prevalence management of sexual aggressive risks in the digital age. *Journal of Child Psychology Psychiatry*, 55(6), pp.635-654.
- [8] Jones, L.M. Mitchell, K.J. (2016). Defining measuring youth digital citizenship. *New Media & Society*, 18(9), pp.2063-2079.
- [9] Hartikainen, H., Iivari, N. Kinnula, M. (2019). Children's design recommendations for online safety education. *International Journal of Child-Computer Interaction*, 22, p.100146.
- [10] Zimba, A., Wang, Z. Chen, H. (2020). Reasoning crypto ransomware infection vectors with Bayesian networks. *Journal of Cyber Security Technology*, 4(3), pp.151-170
- [11] Ferrara, E., Varol, O., Davis, C., Menczer, F. Flammini, A. (2019). The rise of social bots. *Communications of the ACM*, 59(7), pp.96-104.
- [12] Kowalski, R.M., Limber, S.P. McCord, A. (2019). A developmental approach to cyberbullying: Prevalence protective factors, 45, pp.20-32.
- [13] Ross, R., McEvilley, M., Carrier, J. (2019). *Developing Cyber Resilient Systems: A Systems Security Engineering*.
- [14] Carías, J. F. (2021). Cyber resilience self-assessment tool (CR-SAT) for SMEs. Retrieved from IEEE Access, 9, pp.80741-80762.
- [15] Alharbi, T. A. (2021). Assessment of Cybersecurity Awareness among Students of Majmaah University. *Big Data Cognitive Computing*, 5(2).

Session 6: ICT in Education

Title: The Agency and Resilience of Teacher Educators in the Age of AI: A Case Study of Teacher Educators in Taiwan

Author: Hung Jui-Hsuan

Title: Designing a CTF-style Activity as a Self-study Exam in a Computing Course

Authors: Pedro J. Roig, Salvador Alcaraz, Katja Gilly, Cristina Bernad, Carlos Juiz

Title: University Orientation and Conversational Technologies: A Study on the User Experience of an Embodied Conversational Agent

Author: Joy Ciliani

The Agency and Resilience of Teacher Educators in the Age of AI: A Case Study of Teacher Educators in Taiwan

Hung Jui-Hsuan
Center for Teacher Education
National Tsing Hua University
Taiwan

Abstract

In the age of artificial intelligence (AI), Taiwan's teacher educators face an increasingly complex professional landscape. Beyond traditional teaching, research, and service responsibilities in higher education, they must also respond to evolving demands from teacher education policies, qualification examinations, and rapidly changing educational technologies. Integrating AI tools into education brings opportunities and challenges, compelling teacher educators to rethink pedagogical approaches, redefine professional identities, and engage in continuous digital upskilling. They are also tasked with bridging theory and practice by fostering meaningful connections with school settings while coping with international academic pressures.

Under such multifaceted demands, teacher educators' agency and resilience become crucial for their professional development and their ability to adapt to policy and technological changes. This study explores how teacher educators in Taiwan develop agency and resilience in the AI era and examines the structural and enabling factors that shape these attributes. Using the theoretical lens of the "structure-agency" dialectical relationship, the study focuses on three key dimensions: work engagement, policy responsiveness, and identity adaptation.

A qualitative research approach was employed, with semi-structured interviews conducted with 23 teacher educators across Taiwan. The findings reveal: (1) institutional type and structural characteristics of teacher education units significantly influence educators' work engagement, policy responsiveness, and identity adaptation—with identity adaptation emerging as most critical; (2) regardless of institutional differences, teacher educators exhibited strong policy responsiveness and a clear sense of professional mission as "teachers of teachers," a role further intensified by the emerging need to incorporate AI literacy into teacher preparation; and (3) while multiple role expectations and work pressures challenge educators' resilience, they also open opportunities for innovation and deeper integration of theory, practice, and technology.

Based on these insights, it is recommended that policymakers refine teacher education policies with greater precision and flexibility to ease structural burdens while supporting AI integration. University-school partnerships should be strengthened through shared platforms and resources facilitating exchanges between practical experience and evolving digital pedagogy. For teacher educators, professional development mechanisms—such as AI-focused workshops, peer learning networks, and psychological support systems—are essential to fostering resilience and sustaining agency. Practice-oriented, interdisciplinary research should also be encouraged to bridge academic inquiry with digital teaching innovation. Finally, higher education institutions must provide supportive environments that reduce structural constraints and enhance global visibility, especially in an era where AI reshapes education worldwide. Future research may expand to include cross-cultural comparisons of agency and resilience in the context of AI and incorporate mixed methods to validate emerging factors influencing teacher educators' adaptation.

References

- [1] Aagaard, T., Bueie, A., & Hjukse, H. (2022). Teacher educator in a digital age: A study of transformative agency. *Nordic Journal of Digital Literacy*, (1), 31-45.
- [2] Archer, M. S. (1995). *Realist social theory: The morphogenetic approach*. Cambridge university press.
- [3] Archer, M. S. (2003). *Structure, agency and the internal conversation*. Cambridge University Press.
- [4] Archer, R. (2000). The place of culture in organization theory: Introducing the morphogenetic approach. *Organization*, 7(1), 95-128.
- [5] Bartell, T., Cho, C., Drake, C., Petchauer, E., & Richmond, G. (2019). Teacher agency and resilience in the age of neoliberalism. *Journal of teacher education*, 70(4), 302-305.
- [6] Cochran-Smith, M., Craig, C. J., Orland-Barak, L., Cole, C., & Hill-Jackson, V. (2022). Agents, agency, and teacher education. *Journal of Teacher Education*, 73(5), 445-448.
- [7] McDonough, S., Papatraianou, L., Strangeways, A., Mansfield, C. F., & Beutel, D. (2021). Navigating changing times: Exploring teacher educator experiences of resilience. *Cultivating teacher resilience*, 279.

Designing a CTF-style Activity as a Self-study Exam in a Computing Course

Pedro J. Roig¹, Salvador Alcaraz¹, Katja Gilly¹, Cristina Bernad¹, Carlos Juiz²

¹*Miguel Hernández University, Av. Universidad s/n (Ed. Alcudia), Elche, SPAIN*

²*University of the Balearic Islands, Ctra. Valldemossa km 7.5, Palma de Mallorca, SPAIN*

Abstract

Capture The Flag (CTF) competitions were first created in the cybersecurity domain in order for participants to be able to face real life security issues regarding computing and networking in a controlled environment. However, this type of format has been extended to the educational area by means of the introduction of CTF-style activities. In this paper, this format has been used in the design of a self-study exam in order for the students of a college course related to the introduction of computer science to get better prepared for the final exam. A weighted scoring system was put in place by combining the points achieved in the questions, where the use of hints in a question penalized a potential correct answer, and the time elapsed to complete the exam. Results showed an increase of around 10% in the scores of the final exam with respect to the previous edition of the course. However, an inferential study in those results concluded that they were not statistically significant, whereas the sample size was insufficient for the effect size achieved. Therefore, further research with more students involved is needed.

1. Introduction

Capture The Flag (CTF) competitions were initially created to improve the technical skills within the cybersecurity community, as CTF challenges allow to test the knowledge and skills of the contestants related to cybersecurity without having to get into unauthorized networks or systems [1]. This way, participants could face real life issues regarding computing and network security in a controlled environment [2]. Some of the most common topics in this field are cryptography, steganography, web exploitation, or reverse engineering [3].

There are different formats of CTF challenges, such as attack-defense, king of the hill, hack quest, or jeopardy, where the latter consists of a bunch of teams competing to answer questions or solve challenges [4]. A typical format of jeopardy CTF consists of a mission, composed of a series of tasks to be carried out, where some hints may or may not be available for

each task [5]. Competitors get a certain amount of points for solving each task, which is inversely proportional to the time spent in order to solve it, such that the lesser the time spent, the more points are obtained [6]. Hence, in this type of format, the CTF winner is usually the first individual competitor or team to complete the whole mission, as it is usually the one getting more points.

However, in case there are hints available, then each time a participant takes a hint in order to get a clue about how to solve a task, then a certain penalization is applied, thus meaning that a certain amount of points are subtracted from its score, which is directly proportional to the number of hints unveiled [7]. Therefore, the CTF winner might not be the contestant spending less time to clear the mission, as the CTF runner-up might end up with more points if there is not much time difference and the number of hints consumed is lower [8]. In fact, it can be stated that the score of each participant depends on two variables, such as the time elapsed and the number of hints consumed, although the amount of points tied to each variable depends on the CTF designer [9].

It is to be noted that CTF-style activities have been adopted in recent years from the cybersecurity field to other education fields because of the versatility of its characteristics and its gamified environment [10]. In fact, CTF-style activities can provide an alternative way to evaluate students, as they can be used as a summative assessment, in a similar way as traditional written exams [11].

In this paper, a CTF-like activity has been designed in order to provide a self-study exam in a computer course at a college level, whose aim is to get learners trained for a traditional written exam to be carried out at a later stage. The main research goals in this paper are to first present a novel layout for self-study exams whose structure is inspired by the CTF competitions, as well as comparing the academic performance in a course dedicated to the introduction to computer science where an instance of this type of self-study exams was introduced with a similar course without it.

The remaining of this paper is organized as follows: section 2 presents the method, section 3 displays the results, section 4 presents the discussion, and finally, section 5 draws the final conclusions.

2. Method

We were running a college course about the introduction to computer science during the present academic year 2024-2025. The evaluation scheme included the delivery of a number practical cases, where coding designs had to be developed, along with a final exam to be completed through a Learning Management System (LMS). Such an exam consisted on a series of multi-choice questions, where the correct answer had to be selected among a group of possible statements, along with short-answer questions, where the correct answer must be written. In the last academic year, students were unsure about the type of questions they would face during the final exam. Therefore, a CTF-like self-study exam was designed in order to help them get ready for it.

The structure of this self-study exam was composed of 10 questions, randomly drawn out of a pool of a thousand questions in order to minimize the chances for any pair of students to get the same questions. Each question was presented on a separate screen, such that an answer was required before passing to the next question. Regardless of the type of question, namely multi-choice or short-answer, three hints were available for each question in case an answer was not known by a student. However, those hints were initially hidden, although a student could unveil them in a sequential manner so as to get extra information about the question proposed. This way, the first clue was more generic, whilst the second one was slightly more specific, whereas the third one was more focused.

The points assigned for each question was 1 point for a correct-answer question without using any hint, or otherwise, 0 points for an incorrect-answer question regardless the number of hints being used. However, each time a clue was unveiled, then 0.25 points were taken out of a possible correct-answer question. This way, if one hint was unveiled, then a correct-answer question was worth 0.75, whilst if two hints were unveiled, then a correct-answer question was worth 0.50, whereas if three hints were unveiled, then a correct-answer question was worth 0.25.

Regarding the evaluation system used, the assessment of multi-choice questions was straightforward, as if the amount of possible right answers was higher than one, then such an amount was quoted within each question. Hence, for multi-choice single answer (MCSA) question types, the way to get a correct-answer question was just to select the right answer. On the other hand, for multi-choice multiple answers (MCMA), the way to get a correct-answer question was to select all the right answers, or

otherwise, it was considered as an incorrect-answer question.

With respect to short-answer questions, the assessment was done through cosine similarity. This way, the string with the right answer established by the teacher and the string with the answer given by a student were confronted for text matching by means of cosine similarity. The outcome ranged from 0 to 1, where the former implied no matching at all between both strings and 1 implied full matching between both strings. The threshold established was 0.70, which is quite commonly used in this kind of measure of similarity for any given pair of non-zero vectors. The reason for the popularity of this threshold is because the cosine of 45° is the square root of 2 divided by 2, which occurs to be approximately 0.707, and 45° is located just in the middle of the range of angles between 0° and 90° . It is to be noted that the cosine of 0° is 1, while the cosine of 90° is 0, which are basically the upper and lower bounds of the range defined in this case. Hence, if a cosine value is 0.7, then it can be said that it corresponds to an angle of nearly 45° .

Therefore, for a short-answer question to be considered as correct-answer question, then the cosine similarity between the answer given by the student and the right answer established by the teacher must be at least 0.7. Obviously, this process is internally dealt with by the server where the application of self-study exams is hosted, which is also the case for the multi-choice questions.

It is to be noted that this type of format applied to these self-study exams is typically used in CTF competitions. However, in that kind of events, the points obtained by the winner is commonly related to both the amount of time spent to complete a mission and the number of clues being unveiled during the path to do so. Regarding the time, the shorter, the better, thus the amount of points assigned is higher as the shorter time is elapsed to get a mission done. With respect to the clues used, the lesser clues are used, the lesser points are subtracted from the missions accomplished, thus the amount of points assigned is higher.

Focusing on the self-study exams proposed, in order to take into consideration both variables, namely the number of points obtained and the time spent, a weighted score was put in place. The system of weights used was a variation of the Pareto principle, also known as 80/20 rule, such that 80% was assigned to the points achieved through the questions and the other 20% was achieved through the time elapsed in the self-study exam. On the one hand, the percentage assigned to the points achieved relates to the amount of points scored during all the questions are accounted. On the other hand, the percentage assigned to the time units, or units, spent relates to the amount of time elapsed from the beginning of an exam until the exam completion.

It is to be remarked that the time units are assigned such that spending at most 10 minutes in the exam accounts for full marks, whilst 1 unit is subtracted per each extra minute, or any portion of a minute, spent in the exam after overcoming the threshold of 10 minutes. However, the outcome of this subtraction has a lower bound of zero, thus it will never go under zero in any case. The motivation for this grading system of units for the time elapsed is because it is commonly considered that 1 minute per question is a sufficient time in a test exam, hence as there are 10 questions, then 10 minutes should be a fair threshold.

For instance, if a self-study exam contains 10 questions and the score achieved is 7.5 points, then those points represent 80% of the final score, which accounts for $7.5 \times 80/100 = 6$ points in the weighted score. Furthermore, if the time elapsed in this exam is 8 minutes, then the score associated to the time spent is 10 units, which is the top mark as it is lower or equal than 10 minutes. Those 10 units represent 20% of the final score, which account for $10 \times 20/100 = 2$ points in the weighted score. Hence, the overall weighted score obtained is $6 + 2 = 8$ points, which accounts for the final score achieved.

Alternatively, another instance would be a self-study exam with also 10 questions, where the score attained is 5 points, which account for $5 \times 80/100 = 4$ points in the weighted score. Besides, if the time elapsed is 14 minutes and 30 seconds, which gets ceiled up to the nearest integer, namely 15 minutes, hence it exceeds to threshold of 10 minutes to get the full exam done. In this case, a penalization of 1 unit per minute is applied after the 10th minute, such that the penalization is subtracted from 10 units. This way, the units attained due to the elapsed time yield $10 - (15 - 10) = 10 - 5 = 5$ units due to the elapsed time. These 5 units represent 20% of the final score, which account for $5 \times 20/100 = 1$ point in the weighted score. Therefore, the overall weighted score achieved is $4 + 1 = 5$ points, which accounts for the final score obtained.

3. Results

The course devoted to the introduction to computer science was run in the present academic year, namely 2024-2025, and it was attended by 25 learners, where a self-study exam was implemented as a CTF-style activity. On the other hand, in the previous academic year, namely 2023-2024, the course was also run and it was also attended by 25 learners, although there was not any self-study exam.

The next tables show the relevant statistical measurements obtained out of the results related to the final exam in both academic years. Hence, the scores related to the final exam of the current academic year got influenced by the extra training taken by the students by means of preparing and sitting the CTF-style self-study exam.

Furthermore, as the college is based in Spain, the Spanish grading system has been used in order to evaluate students. This scoring system ranges from 0 all the way to 10, while a score of 5 represents the passing score [12].

Table 1 displays a summary of the most relevant measurements regarding descriptive statistics. The first one is the sample size, which accounts for the number of learners in each of the courses delivered. The second one is the mean, which stands for the average score obtained by all students in each course, thus being a centralization measurement. The third one is the standard deviation, which stands for the dispersion of the scores corresponding to each course with respect to its average, thus being a dispersion measurement [13].

Table 1. Relevant measurements regarding descriptive statistics

	N	Mean	Standard deviation
Past academic year	25	6.12	1.82
Current academic year	25	6.75	1.75

Table 2 exhibits the steps to calculate the two-sided t-test at the 95% confidence interval in order to find out if the differences in the distribution of scores regarding each academic year are statistically significant or otherwise [14]. The t-test is also known as significance test and is a relevant part of inference statistics.

Table 2. T-test applied to both distributions of scores

t-student test value	1.25
degrees of freedom (df)	$25+25-2 = 48$
p-value	0.22
s-value	0.05
Interpretation of the p-value	Not significant

Table 3 shows the effect size achieved with the values of the means and standard deviations related to the distribution of scores corresponding to each academic year [15]. The target is to measure the strength of the statistical claim due to the difference between groups in inferential statistics.

Table 3. Effect size achieved with both distributions of scores

Difference of means	0.63
Pooled standard deviation	1.79
Cohen's d (effect size)	0.35
Interpretation of Cohen's d	Small to medium

Table 4 exposes the sample size needed to get the effect size achieved with the distribution of scores corresponding to each academic year [16]. The target is to minimize possible errors in inferential statistics.

Table 4. Sample size needed to get the effect size achieved with both distributions of scores

Cohen's d	0.35
Alpha value	0.05
Power	0.80
Sample size needed	127
Sample size achieved	25
Interpretation of sample size	Insufficient

Table 5 compares the value of the sample size needed obtained through different methods. In this sense, the values chosen for alpha and power have been the most commonly used in the literature, such that former is 0.05 and latter is 0.80, respectively [17]. Furthermore, in order to make fairer comparisons among those methods, the effect size is going to be considered with up to four decimal digits, thus being 0.3529.

Table 5. Sample size needed obtained through different methods

Mathematical expression	127
Online sample size calculator	127
Table with effect size -vs- sample size	129
Lehr's rule of thumb	128

4. Discussion

With respect to Table 1, it shows that the mean obtained in the current academic year, namely 6.75, is higher than the mean attained in the last academic year, namely 6.12, thus accounting for a rise slightly higher than 10%. Hence, the academic performance related to the overall students have improved in the current academic year. On the other hand, the number of attendants in both courses was the same, whereas the difference in the standard deviations was around a 4% lower in the current academic year. Therefore, the rise in the mean lead to think that the establishment of the CTF-style self-study exam may well be the responsible for this enhancement, as students got better prepared to the final exam.

With regards to Table 2, it exhibits a value of 1.25 for the t-student value of 1.25 and a value of 48 for the degree of freedom, which leads to a p-value of 0.22. However, this value is higher than s-value chosen, which is 0.05, hence it can be concluded that there is not a statistically significant difference between the distribution of marks attained in both academic years. Therefore, it cannot be concluded that the improvement in the scores was due to the effect of the

self-study exams, as that enhancement could have been due to chance.

Regarding Table 3, the difference of means obtained is 0.63, whereas the pooled standard deviation is 1.79, which actually was the average spread of scores around their group mean. The division of both values lead to calculate Cohen's d, which accounts for the standardized effect size. The value obtained was 0.35, thus it was right in the middle of the interval with a lower bound of 0.20, which accounts for small effect, and an upper bound of 0.50, which stands for medium effect. It is to be noted that a value of 0.50 represents an effect visible to the naked eye, considering a careful observer. Moreover, it is to be highlighted that the lower the effect size, the greater the sample size is needed.

With respect to Table 4, the Cohen's d is 0.35, as stated in the previous table, whereas the alpha value taken is 0.05 and the value for statistical power is 0.80. With those values, the sample size needed is 127, which is higher than the sample size achieved, namely the number of attendants in each course, which are 25 students. Hence, this sample size is insufficient for the effect size obtained, namely 0.35, which means that a greater number of students should be considered in order to state that the effect size achieved was due to the implementation of the self-study exam.

Regarding Table 5, the four methods exposed present very similar results, although the mathematical expression is meant to be most precise one. In the first place, the outcome was obtained through a specific mathematical expression written in a spreadsheet [18], whose result was the one shown in the previous table. In the second place, the outcome was attained through an online sample size calculator [19], which happens to also carry out effect size calculations.

In the third place, the outcome was achieved through a table showing the relationship between effect size, namely the independent variable x, and sample size, namely the dependent variable y [20]. In order to obtain a more precise value for y, a trend line has been found out with the data pairs involved, namely x and y, leading to the following equation: $y = 16.755x^{-1.963}$, which allowed to calculate the outcome shown in the table. And in the fourth place, the outcome was obtained through Lehr's rule of thumb [21], which is a shortcut offering a quick and close approximation to the sample size.

In summary, the implementation of the CTF-style self-study exam in order for students to better prepare their final exam got them improve their scores by around 10%. However, this improvement is not statistically significant, which means that it cannot be concluded that such an improvement was due to the implementation of the self-study exam, as it could have been due to chance. Likewise, the standardized effect size achieved of 0.35 requires a sample size of 127, whereas the actual sample size was 25. Hence, a

larger number of students should be including in further research in order to conclude that the implementation of the self-study exam led to the effect size achieved.

Looking at future courses, and focusing on the scoring model used, namely the weighed score, the system of weights put in place has been considered as a reasonable balance between test results and time spent. However, further research might lead to better adjust the 80/20 rule established between the outcome of the exam and the time spent to get it completed in order to give more weight to either the points achieved in the exam, or otherwise, the units associated to the time elapsed.

On the other hand, it is to be noted that a weighted score could be considered as a normalized score, as two variables, namely the test outcome and the time spent, are used to calculate the final score. However, other options for scoring are available. Further research might lead to the convenience of changing the scoring method.

One option is a direct score, which just takes into account the points achieved in the exam, whereas time is not considered. However, if the time limit is reached, then the exams not yet submitted are automatically submitted at that moment.

Another option is learning curve score, which assigns a determined time to each question according to its difficulty, as opposed to assign a global time for the overall exam. In this case, if the deadline time for a given question is reached, then this question is considered as unanswered and the following question is presented to the student.

A further option is applying the Rasch scoring, which depends on both the student's knowledge and the item difficulty, where the probability of getting a correct answer is modeled as a logistic function [22]. This method is considered as more objective, although the key point is the difficulty in establishing the statistical model to estimate the values for the ability of a person and the difficulty of an item.

Furthermore, the importance of a self-study exam in future courses could be improved by assigning some weight in the course final score. For instance, a certain percentage of the course overall evaluation score may be assigned to the self-study exam in order for learners to take it more seriously. This percentage could be something like 5% or 10%.

Besides, another improvement could be to penalize the wrong answers, which is not done in the model proposed herein. A typical penalization in multi-choice questions is to subtract the inverse of the predecessor of the number of choices available per each incorrect-answer question, so this could be applied in future editions in order to restrain random answers. For instance, in a multi-choice test made of questions with 4 possible answers, this penalization system subtracts $1/3$ points per each wrong-answer question out of the overall correct-answer questions.

Additionally, another enhancement could be the chance for students to take a self-study exam more than once in order for them to get acquainted with the layout at their first attempt, such that they may improve their performance in successive attempts. Nonetheless, in order to avoid taking the exam too many times, a maximum number of attempts may need to be established, such as just 3.

Moreover, in those circumstances, the fairest selection of the final score assigned to the self-study exam may be the score at the last attempt, even though another option would be to stick to the highest score achieved in any attempt.

5. Conclusions

In this paper, a CTF-style self-study exam has been proposed in order for students to get better prepared to the final exam of a college course devoted to the introduction to computer science. The self-study exam was composed of 10 questions, whose answers could be either multiple-choice or short answer. Each of those questions was worth 1 point, hence the highest score obtained was 10 points. Furthermore, the format of the exam followed that of a jeopardy CTF, where questions were presented to the participants and some hints were offered per each question with different clues in order to help then find the correct answer.

A weighted score system was put in place, where a variation of the Pareto principle was used. In fact, the scoring achieved with the questions was worth 80%, whereas the time spent was worth the other 20%. Regarding the former, the score achieved for each question depended on the number of hints unveiled before answering, such as a correct response without unveiling any hint accounted for 100% of the value of that question, whilst a correct response where one hint was unveiled accounted for 75%, a correct response where two hints were unveiled stood for 50%, a correct response where three hints were unveiled was worth 25%, and obviously, a wrong response was worth nothing.

Regarding the latter, the deadline time was considered as 10 minutes, as it accounts for 1 minute per each of the 10 questions proposed. Hence, if a students completed an exam in less than 10 minutes, then 10 time units, namely units, were assigned, whereas 1 unit was taken per each minute after the deadline time, which also applies for any fraction of a minute left. However, a lower bound of zero was set up for the scoring related to time, such that it could not detract any points from the those obtained with the correct answers to the questions.

With respect to the results, the mean score in the current academic year was 6.75, where the self-study exam was set up, while the mean score in the last academic year was 6.12, where no self-study exam was found. Hence, an improvement of about 10% was

experienced with the implementation of the self-study exam, thus from the point of view of descriptive statistics, it seems to be a good result.

However, a t-test was performed, leading to a p-value of 0.22, which is higher than the s-value selected, namely 0.05. Hence, this improvement cannot be seen as statistically significant from the point of view of inferential statistics, thus meaning that the enhancement could be due to chance.

Furthermore, the effect size achieved was 0.35, which accounts for a value in the middle of the interval defined by 0.20 as the lower bound, which accounts for small effect, and 0.50 as the upper bound, which stands for medium effect. Additionally, the sample size considered was 25 students, although the sample size needed was 127, considering the effect size, the alpha value of 0.05 and the statistical power of 0.80. Hence, the sample size used was insufficient for the effect size achieved. Therefore, further research must be carried out with a larger amount of students in order to account for the effect size achieved.

6. References

- [1] Karagiannis, S., & Magkos, E. (2020). Adapting CTF challenges into virtual cybersecurity learning environments. *Information and Computer Security*, 29(1), 105-132.
- [2] Mansurov, A. (2016). A CTF-Based Approach in Information Security Education: An Extracurricular Activity in Teaching Students at Altai State University, Russia. *Modern Applied Sciences*, 10(11), 0159.
- [3] Collins, J., Ford, V. (2023). Teaching by Practice: Shaping Secure Coding Mentalities through Cybersecurity CTFs. *Journal of Cybersecurity Education, Research and Practice*, 2, 09.
- [4] Zafar, A., Yamin, M. M., Katt, B., Torseth, E. (2024). All flags are not created equal: A deep look into CTF Scoring Algorithms. *Expert Systems and Applications*, 254, 124436.
- [5] Ford, V., Siraj, A., Haynes, A., Brown, E. Capture the Flag Unplugged: An Offline Cyber Competition. In: *Proceedings of SIGCSE 2017*, March 8-11, 2017, Seattle (WA), United States.
- [6] Minagar, S., Sakzad, A. (2023). Automatic Problem Generation for CTF-Style Assessments in IT Forensics Courses. In: *Proceedings of ITiCSE 2023*, July 7-12, 2023, Turku, Finland.
- [7] Lagorio, G., Ribaudo, M., Armando, A. (2021). Capture the Flag competitions for Higher Education. In: *Proceedings of ITASEC 2021*, April 7-9, 2021, Rome, Italy.
- [8] Beauchamp, C., Matusovich, H. (2024). A Mixed-Method Study Exploring Student Motivation for Participating in Cybersecurity CTF Competitions. *Cybersecurity Pedagogy & Practice Journal*, 3(1), 4-26.
- [9] Nelson, C., Shoshitaishvili, Y. (2024). PWN The Learning Curve: Education-First CTF Challenges. In: *Proceedings of SIGCSE 2024*, March 20-23, 2024, Portland (OR), United States.
- [10] Cole, S. V. (2022). Impact of Capture The Flag (CTF)-style vs. Traditional Exercises in an Introductory Computer Security Class. In: *Proceedings of ITiSCE 2022*, July 8-13, 2022, Dublin, Ireland.
- [11] Meinsma, D., Barjas, S., Gilhespy, M., Björkqvist, M. (2022). Effectiveness of CTF education. Measuring the learning outcomes of Jeopardy CTF's. The Hague University of Applied Sciences, The Netherlands.
- [12] Polytechnic University of Valencia. Spanish grading system, <https://www.upv.es/entidades/OPII/infoweb/pi/info/1147768normali.html> (accessed on February 28, 2025).
- [13] Roig, P. J., Alcaraz, S., Gilly, K., Bernad, C., Juiz, C. Design and Assessment of an Active Learning-Based Seminar. *Education Sciences*, 14(4), 0371.
- [14] Roig, P. J., Alcaraz, S., Gilly, K., Bernad, C., Juiz, C. A Generic Multilevel Structure for Educational Escape Rooms. *Education Sciences*, 14(5), 0488.
- [15] Schäfer, T., Schwarz, M. A. (2019). The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. *Frontiers in Psychology*, 10, 0813.
- [16] Power and Effect Sizes. Available at: <https://psyteachr.github.io/quant-fun-v2/power-and-effect-sizes.html> (accessed on February 28, 2025).
- [17] Praharaj, S. K., Ameen, S. (2024). Sample size estimation in research: necessity or compromise? *Kerala Journal of Psychiatry*, 37(1), 66-71.
- [18] Sommet, N., Weissman, D. L., Cheutin, N., Elliot, A. J. (2023) How Many Participants Do I Need to Test an Interaction? Conducting an Appropriate Power Analysis and Achieving Sufficient Power to Detect an Interaction. *Advances in Methods and Practices in Psychological Science*, 6(3).
- [19] Sample size calculation. Available at: <https://www.highyieldmed.org/sample-size-calculation-2/> (accessed on February 28, 2025).
- [20] Power calculation Cohen's d. Available at: <https://jackauty.com/power-calculator-cohens-d/> (accessed on February 28, 2025).
- [21] McConnell, M. M., Monteiro, S., Bryson, G. L. (2019). Sample size calculations for educational interventions: principles and methods. *Canadian Journal of Anesthesia*, 66, 864-873.
- [22] Verdú-Soriano, J., González-de la Torre, H. (2024). Rasch analysis implementation in nursing research: A methodological approach. *Enfermería Clínica*, 34, 493-506.

University Orientation and Conversational Technologies: A Study on the User Experience of an Embodied Conversational Agent

Joy Ciliani

University of Siena and University of Campania "Luigi Vanvitelli", Italy

Abstract

Students making decisions during transitions between different stages of education or from education to work increasingly need effective guidance tools that can respond to their needs in a personalized and accessible way. This study investigates the potential of Embodied Conversational Agents (ECAs) in the university context, focusing on the user experience evaluation of an agent designed to guide students at the master's degree program in Communication Strategies and Techniques at the University of Siena. By administering the User Experience Questionnaire Plus (UEQ+) to 17 students, the study provides insights into the role that conversational technologies can play in guidance services, helping foster more immediate and effective interactions in response to the communicative and motivational challenges that characterize academic transitions.

1. Introduction

Young individuals encounter a particularly complex decision-making process when determining whether to pursue further studies or enter the workforce. While some navigate this transition with relative ease, others experience difficulties and often rely on the influence of peers, educators, parents, and other relevant adults [11]. Within a constantly evolving socio-economic landscape characterized by rapid and frequent changes in the labor market, an increasing number of individuals revise their career trajectories over the course of their lives [9].

These dynamics clearly emerge in the university context, especially during the transition from one cycle of study to the next. At this stage, decision making challenges become more intense as students must select a specific program from a wide and complex academic offering often without a clear understanding of the differences between courses, related career prospects, or the level of commitment required. Although universities provide a range of online resources and email-based information [7], the overload of generic technical

content shared through underused channels makes it hard for students to navigate the system effectively and independently.

Consequently, students increasingly seek support from reference figures such as professors, tutors, peers, and dedicated services to strengthen academic motivation and enhance performance [5]. However, limited staff availability, long response times, and students' hesitancy to request assistance, frequently restrict access to these resources potentially resulting in frustration, uncertainty, and demotivation, and in some cases leading to discontinuities in students' academic trajectories.

To respond to these challenges, higher education institutions have begun to implement innovative strategies aimed at supporting students during critical phases of academic decision-making. Among these approaches, Embodied Conversational Agents (ECAs) have emerged as advanced tools capable of simulating human communication and behavior [3], thereby facilitating continuous and interactive access to complex information.

This study investigates the user experience of an ECA specifically developed to support orientation for the master's degree program in Communication Strategies and Techniques at the University of Siena. The evaluation employs the User Experience Questionnaire Plus (UEQ+), a validated instrument designed to assess multiple dimensions. The findings provide empirical insights into the potential of ECAs as innovative and effective tools for university guidance services.

2. New perspectives in university orientation

Orientation encompasses the set of actions implemented by institutions to support students in transitioning from their previous educational or personal context to the university environment [16]. In recent

years, alongside traditional in-person orientation activities, digital technologies, such as conversational agents, have emerged as new support tools [1][13]. Many interactions with these tools remain predominantly informative, focusing on answering frequently asked questions or providing essential communications related to enrollment, course registration, and campus navigation [12]. However, orientation is evolving beyond mere information delivery, aiming to support students through a broader process of reflection and awareness. This process extends across all phases of the academic journey, from entry to university, through the in progress, and finally towards graduation and transition to the workforce or postgraduate education [2].

2.1. Embodied Conversational Agents in the educational context

The application of Artificial Intelligence (AI) in educational contexts constitutes a relatively recent development that has introduced new modalities of interaction between students and conversational agents. Among these, ECAs represent a particularly promising category, as they integrate verbal and non-verbal communication cues including facial expressions, gestures, and vocal modulation to simulate human-like interaction and enhance user engagement [4]. Current implementations primarily employ ECAs as instructional support tools [17], integrating them into digital learning platforms [18] and immersive environments such as virtual or mixed reality settings [10]. However, their deployment in student guidance and orientation remains limited. Despite the potential of ECAs to facilitate access to complex information and support decision-making processes, empirical research in this area is still sparse and warrants further investigation.

3. Objectives

To ensure the effectiveness of academic guidance, institutions must adopt tools developed in response to students' needs, such as the demand for clear, reliable, and timely information [8], as well as the opportunity to express their aspirations and address doubts that may arise throughout their academic journey. These needs, however, are not homogeneous, they vary according to each student's educational background, origin, cultural and social capital, and individual expectations.

This study aims to evaluate students' experience in interacting with an ECA designed to support orientation for the master's degree program in Communication Strategies and Techniques at the University of Siena.

The analysis seeks to determine how this tool can not only facilitate access to course-related information but also identify students' needs and respond to them in a personalized and effective manner.

4. Methodology

For this research, we used the UEQ+ [15], an extension of the original UEQ [14], in which 14 research-relevant dimensions were selected and investigated: attractiveness, efficiency, clarity, reliability, interest, novelty, trust, acoustic quality, usefulness, perceived value, aesthetics, user-friendliness, reliability and content quality.

The questionnaire was administered to students following the interaction with the ECA carried out in two modes: a free phase, in which they could explore the agent following their personal curiosities, and a structured phase, during which they were guided through four predefined tasks that required them to search for information about the degree course (curriculum, outlets, tutors, access requirements and university services). In order to make the interaction experience as realistic and close to everyday usage patterns, participants were asked to use their smartphones and headphones, so that they could communicate with the agent by both text and voice (see Figure 1).



Figure 1. Photo with participants taking the test

5. Sample and demographic analysis

The study sample consisted of 17 students from the University of Siena, attending the three-year degree course in Communication Sciences. Being familiar with the use of digital technologies and innovative applications of communication, they could be interested

in continuing their studies with the master's degree course in Communication Strategies and Techniques. From a demographic point of view, the composition of the sample was heterogeneous, with 6 men and 11 women of whom 16 declared an age between 18 and 25, and only one over 35.

6. Results

Table 1 and Table 2 were created using the Data Analytics Tool as the analysis tool, whose originally collected data were transformed from a 7-point Likert scale range to a range between -3 and +3.

Table 1. Mean, variance, standard deviation and number of questionnaires administered for the dimensions of the UEQ+ considered

Scale	Mean	Variance	Std.dev.	N.
Attractiveness	1,09	1,57	1,25	17
Efficiency	0,63	2,53	1,58	17
Perspicuity	1,76	1,41	1,18	17
Dependability	0,72	2,83	1,67	17
Stimulation	1,35	1,63	1,27	17
Novelty	1,53	1,78	1,32	17
Trust	2,59	0,40	0,62	17
Acoustics	0,25	2,05	1,42	17
Usefulness	1,62	1,91	1,37	17
Value	0,96	2,01	1,41	17
Visual Aesthetics	0,82	3,10	1,75	17
Intuitive Use	1,74	1,42	1,18	17
Trustworthiness of Content	1,75	1,56	1,24	17
Quality of Content	1,62	1,05	1,02	17

Table 2: Mean, variance, standard deviation and number of questionnaires administered for the importance of the UEQ+ dimensions considered

Scale	Mean	Variance	Std.dev.	N.
Attractiveness	1,29	1,47	1,18	17
Efficiency	1,76	1,69	1,26	17
Perspicuity	2,29	0,97	0,96	17
Dependability	1,71	1,60	1,23	17
Stimulation	1,71	1,97	1,36	17
Novelty	1,65	1,74	1,28	17
Trust	2,59	1,01	0,97	17
Acoustics	1,27	4,82	2,09	11
Usefulness	2,00	1,75	1,28	17
Value	1,82	0,90	0,92	17
Visual Aesthetics	2,35	0,62	0,76	17
Intuitive Use	2,35	1,12	1,03	17
Trustworthiness of Content	2,41	0,76	0,84	17

Quality of Content	2,53	0,76	0,85	17
--------------------	------	------	------	----

The analysis of the data from the UEQ+ highlights an overall positive user experience (see Figure 2). Among the most appreciated dimensions, trust stands out ($M = 2.59$), with a narrow score distribution ($SD = 0.62$), indicating a strong perception of reliability and transparency from users. Perspicuity ($M = 1.76$) and intuitive use ($M = 1.74$) also received positive evaluations, suggesting that the system interface is clear and easy to use even without prior experience. Other well-perceived aspects include novelty ($M = 1.53$), stimulation ($M = 1.35$), quality of content ($M = 1.62$), and service usefulness ($M = 1.62$), indicating that the system is seen as innovative, engaging, and useful in meeting users' informational needs. However, some dimensions received lower ratings, pointing to areas for improvement. In particular, efficiency ($M = 0.63$) was evaluated heterogeneously revealing divergent perceptions regarding the speed and fluidity of interaction. Visual aesthetic appeal ($M = 0.82$) and perceived value ($M = 0.96$) also scored lower compared to other dimensions, suggesting that, although functional, the interface may not be perceived as particularly attractive or well-enhanced. An additional critical issue emerged in the acoustic dimension ($M = 0.25$), characterized by high variability in responses and overall low ratings, probably influenced by the limited voice interaction experienced by some participants.

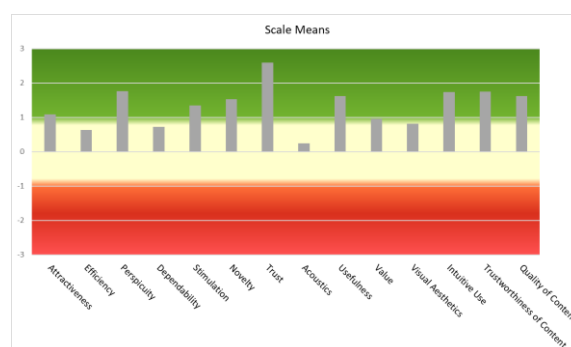


Figure 2. Visualization of the means of the evaluated dimensions in the UEQ+ data analysis Excel sheet

In the analysis of the perceived importance of the various dimensions of user experience (see Figure 3), dimensions related to trust and quality of content emerge as the most relevant, confirming the central role these aspects play in the interaction with the system. Specifically, trust ($M = 2.59$), content quality ($M = 2.53$), trustworthiness of content ($M = 2.41$), and perceived usefulness ($M = 2.00$) rank highest on the importance scale, indicating that users place primary

value on the truthfulness, informational reliability, and practical functionality of the service. Dimensions related to perspicuity, intuitive use and visual aesthetics also receive high ratings. Perspicuity ($M = 2.29$), intuitive use ($M = 2.35$), and visual aesthetics ($M = 2.35$) are regarded as significant factors contributing to a positive user experience. In contrast, aspects such as stimulation ($M = 1.71$) and novelty ($M = 1.65$) are perceived as desirable but not essential.

Finally, dimensions such as attractiveness ($M = 1.29$) and acoustic elements ($M = 1.27$) are rated as least important, suggesting that global aesthetic appeal and the auditory component of the interaction play a marginal role in users' perception of the experience.

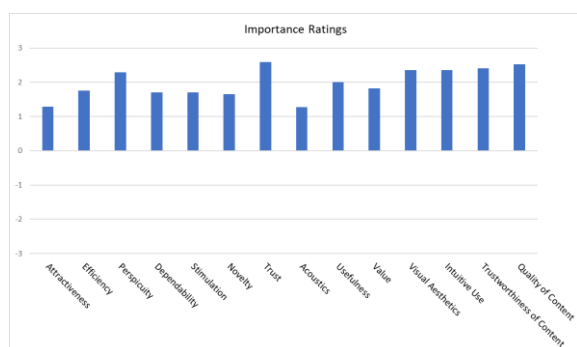


Figure 3. Visualization of the mean importance of the dimensions of the UEQ+ data analysis Excel sheet

7. Discussion

The analysis of the UEQ+ data shows in the context of ECA-mediated university orientation two dimensions that are particularly central in determining a positive user experience: trust and quality of the content provided. Trust, in particular, is confirmed as a key element for the initial acceptance of digital tools, continued adoption and informed use over time [6]. In the context under review, trust is perceived by students not only as an essential requirement for interaction, but also as one of the dimensions best met by ECA, contributing significantly to the success of the experience.

Quality of content also scored well in terms of both importance and perceived quality of the experience undertaken. As a practical example of the interaction, the following figure (see Figure 4) shows an excerpt from a conversation between students and the ECA, where typical questions were asked, such as whether the curriculum is more theoretical or practical and what an exam consists of, highlighting the role of the ECA in providing clear responses.

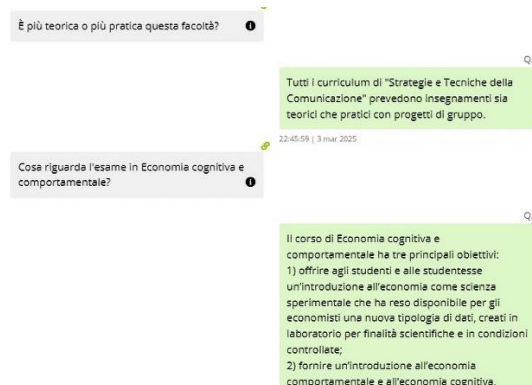


Figure 4. Excerpt from a conversation between a student and the ECA

Other aspects such as aesthetics, trustworthiness of content, and intuitive use are perceived as important to the experience and scored well, but are slightly lower. On the other hand, it is interesting that the acoustic component that is one of the distinctive aspects of ECAs, was perceived as marginal in the user experience, receiving the lowest ratings both in terms of perceived importance and for the quality of the experience offered by the university's ECA. This data suggests a preference for text-based interaction, which seems to be more convenient or appropriate for the type of activity performed.

Although with some areas for improvement, the overall experience with the ECA was evaluated positively by the students, observing consistency between the dimensions considered important and those actually appreciated. However, we should emphasize that the evaluation of an ECA is a complex and multifactorial task, as it involves a multiplicity of cognitive, emotional, communicative, social and aesthetic aspects that can hardly be fully captured by a single instrument. In this study, we chose to use a validated questionnaire, the UEQ+, focusing on only a few dimensions of the user experience, reflecting only part of the complexity of the interaction.

Moreover, user experience is inherently a dynamic and evolving phenomenon, shaped by users' growing familiarity with the tool and changes over time. For this reason, a critical and longitudinal approach is essential to better understand how perceptions develop and shift. In line with this perspective, a follow-up evaluation is already planned in the medium-to-long term to assess whether the ECA is capable of sustaining student engagement and attention over time, and to capture any evolving patterns in user perception as the tool becomes more embedded in academic routines.

8. Conclusions

The present study analyzed the user experience of an ECA designed to support university students in their orientation processes to the master's degree course in Communication Strategies and Techniques. The evaluation, conducted by means of the validated UEQ+ questionnaire, made it possible to systematically explore some key dimensions of the user experience, returning an overall positive scenario. The results show that ECA is a promising tool to facilitate access to complex information in a personalized and interactive way. Among the main advantages that emerged were the ability to establish trust, the reliability of the content offered and the ease of use, elements that were particularly appreciated by the students. The aesthetic and vocal dimensions, although in need of improvement, did not compromise the overall evaluation of the experience.

The tested ECA did not remain an isolated prototype, but was actually integrated and made operational within the degree course website, where it is currently available and used by students. This implementation in the real-life context offers a research opportunity for future studies aimed at analyzing how spontaneous interaction takes place and assessing the long-term impact on students' access to information and autonomy in the orientation process.

In conclusion, the ECA emerges as a digital resource capable of enhancing institutional communication through conversational modes, and future studies integrating different methodological approaches may further explore the role of this tool in the university environment.

9. References

- [1] Alkhoori, A., Kuhail, M. A., & Alkhoori, A. (2020, April). UniBud: a virtual academic adviser. In 2020 12th annual undergraduate research conference on applied computing (URC) (pp. 1-4). IEEE.
- [2] Annovazzi, C., Camussi, E., Meneghetti, D., Ulivieri Stiozzi, S., & Zuccoli, F. (2018). Orientamento e peer tutoring. Un processo pedagogico innovativo per facilitare e accompagnare l'avvio del percorso universitario. *Cqia Rivista*, 25, 201-209.
- [3] Barnes, N., Fischer, S., & Kilpatrick, S. (2024). Going above and beyond: Realigning university student support services to students. *International journal of educational research*, 124, 102270.
- [4] Cassell, J. (2001). Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine*, 22(4), 67-67.
- [5] Chao, P. Y., Lai, K. R., Liu, C. C., & Lin, H. M. (2018). Strengthening social networks in online discussion forums to facilitate help seeking for solving problems. *Journal of Educational Technology & Society*, 21(4), 39-50.
- [6] Davis, F. D. (1985). A technology acceptance model for empirically testing new end-user information systems: Theory and results (Doctoral dissertation, Massachusetts Institute of Technology).
- [7] De-Juan-Vigaray, M. D., & Moreno, C. G. (2017). Information overload or lack of interest: does the tutorial action plan (tap) really stand out among the many university services on offer?. In *INTED2017 Proceedings* (pp. 4513-4522). IATED.
- [8] Eleyan, D., & Eleyan, A. (2011). Coaching, Tutoring and Mentoring in the Higher Education as a solution to retain students in their major and help them achieve success.
- [9] Gati, I., Krausz, M., & Osipow, S. H. (1996). A taxonomy of difficulties in career decision making. *Journal of counseling psychology*, 43(4), 510.
- [10] Manfredi, G., Erra, U., & Gilio, G. (2023, June). A mixed reality approach for innovative pair programming education with a conversational AI virtual avatar. In *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering* (pp. 450-454).
- [11] Moogan, Y. J., & Baron, S. (2003). An analysis of student characteristics within the student decision making process. *Journal of further and Higher Education*, 27(3), 271-287.
- [12] Nguyen, H., Lopez, J., Homer, B., Ali, A., & Ahn, J. (2023). Reminders, reflections, and relationships: insights from the design of a chatbot for college advising. *Information and Learning Sciences*, 124(3/4), 128-146.
- [13] Page, L. C., & Gehlbach, H. (2017). How an Artificially Intelligent Virtual Assistant Helps Students Navigate the Road to College. *AERA Open*, 3(4). DOI: 10.1177/2332858417749220.
- [14] Schrepp, M. (2015). User experience questionnaire handbook. All you need to know to apply the UEQ successfully in your project, 10.
- [15] Schrepp, M., & Thomaschewski, J. (2019). Eine modulare Erweiterung des User Experience Questionnaire. In *Mensch und Computer 2019-Usability Professionals* (pp. 10-18420). Gesellschaft für Informatik eV Und German UPA eV.
- [16] Upcraft, M. L., & Farnsworth, W. M. (1984). Orientation programs and activities. *New directions for student services*, 1984(25), 27-38.
- [17] Zhang, Y., & Pan, W. (2025). A scoping review of embodied conversational agents in education: trends and

innovations from 2014 to 2024. Interactive Learning Environments, 1-22.

[18] Zhao, Z., Yin, Z., Sun, J., & Hui, P. (2024, December). Embodied AI-guided interactive digital teachers for education. In SIGGRAPH Asia 2024 Educator's Forum (pp. 1-8).

Acknowledgment

This study was supported by the Department of Social, Political and Cognitive Sciences at the University of Siena. We would also like to thank the master's degree program in Communication Strategies and Techniques for providing the opportunity to implement and evaluate the ECA on its official website. Our gratitude also goes to QuestIT S.r.l. for granting access to the Algho platform, which made the development of the ECA possible.

Session 7: e-Society

Title: A Mobile Decision Support System for Early Prediction of Mental Health Disorders
Authors: Divine Prince Daniel, Enesi Femi Aminu, Ayobami Ekundayo

Title: Development of a Colon Cancer Classification System
Authors: V.I. Ainoko, A.O. Odeniyi, O.Y. Ogunlola, A. Innocent, B.K. Alese

Title: Unmasking Fake Identities on Instagram Using Machine Learning
Authors: Olaniyi Abiodun Ayeni, Ethel Femi-Gbenoba

A Mobile Decision Support System for Early Prediction of Mental Health Disorders

Divine Prince Daniel, Enesi Femi Aminu, Ayobami Ekundayo

Department of Computer Science, Federal University of Technology, Minna, Nigeria

Abstract

Mental health disorders are a significant global burden, with depression, anxiety, and schizophrenia currently the leading causes of global burden. Despite increased awareness, the current diagnostic and management practices are plagued by subjective evaluation, underreporting, and heterogeneity of diagnosis. Mobile health (mHealth) applications, though they are promising, they lack intelligent, predictive capabilities that allow for early detection and intervention. This research proposes the development of an AI-driven mobile decision support system that allows for the prediction of mental health disorders by an integration of behavioral signatures, social contacts, environmental considerations, and past health history. By the exploitation of the strengths of machine learning (ML) algorithms, the system will analyze users' real-time data to provide individualized predictions and communicate timely alerts to users and clinicians. By enhancing the accuracy of early detection, access to care, and support to clinicians in making data-driven decisions, the system is likely to bridge the gap in the currently available tools in mental health. The proposed solution here is a scalable, accessible, and personalized solution for mental health care, aligned to global health priorities in mental health intervention. The study also addresses the major issues of data confidentiality, interpretability of models, and cultural acceptability for widespread adoption by AI-based mental health tools. This study champions the cause of the development of new, AI-based solutions in mental health, promising the benefits of improved outcomes and overall wellness.

Keywords: Mental Health Diagnosis, Clinical Decision Support, Predictive Modeling, AI Chatbots in Mental Health.

1. Introduction

Mental conditions such as depression, anxiety, schizophrenia, and bipolar disorders are large health issues globally. They can incapacitate individuals,

isolate them, and reduce their productivity [1]. The traditional methods of diagnosing and managing mental illness, which are highly dependent on interviews and what the patient reports, are questioned because they are subjective and not very valid [2] [3]. These old methods are not suitable to the evolving nature of mental illness, and thus we require new data-driven processes.

The development of mobile health (mHealth) technologies offers valuable solutions. Smartphone applications and wearable sensors have been demonstrated to collect continuous behavioral, health, and environmental information for instant identification and reaction [4] [5]. For instance, [6] demonstrated that wearable sensors and AI can identify indicators of depression, while [7] developed an IoT-based system to identify mental health threats in students with 99% accuracy. These technologies, in combination with machine learning algorithms, can enhance early detection as well as personalized treatment [8].

AI and ML can be used to predict mental health issues. They can diagnose conditions such as schizophrenia and depression with high accuracy using techniques such as support vector machines and deep neural networks [2]; [3]. [9] demonstrated that AI diagnoses mental health conditions more quickly and accurately compared to conventional approaches. Yet, there are issues, including knowing how the models operate, data quality, and how the models function across different scenarios [3].

Another significant question is the responsible use of AI mental health technologies and whether individuals will embrace them. Research by [10] indicates that issues such as data privacy and how algorithms can be explained influence whether users are willing to use the technologies. [11] discussed how AI chatbots will be used to engage users but noted that the bots are not aware of emotions and that there are no norms to regulate their use.

How effectively these digital resources perform largely depends on how they are applied. Research indicates that applications that involve feedback from users along with personal data are made use of more

frequently and provide better outcomes [12] [13]. Also, mobile health applications incorporating AI-based decision support systems (DSS) are better able to interpret real-time information, provide better and faster feedback, and enhance care plans [14].

To further address diagnostic inefficiencies, several studies advocate for the integration of clinical decision support systems (CDSS) into mHealth platforms. [15] reported that mobile app-based support tools, when designed with usability and contextual appropriateness, can improve provider trust and standardize diagnostic decisions. Furthermore, [15] highlighted the role of mobile app-based decision support tools in helping primary healthcare workers to make mental health diagnosis and treatment more standard. These systems not only facilitate early detection but also enhance decision accuracy and follow-up care, addressing both technical and service-level gaps in mental healthcare [16] [17]. This research contributes to the growing body of literature advocating for technologically empowered mental health care, offering an integrative solution that responds to both the clinical and technological challenges outlined in existing research [16]; [17]; [18].

However, there is a growing interest in using mobile technologies for mental health monitoring, most existing systems are limited by their reliance on static user input, insufficient data integration, and low predictive capabilities. Many mobile mental health applications focus on tracking symptoms or providing general well-being advice without leveraging AI for real-time, data-driven predictions. This shortfall limits their effectiveness in early diagnosis and personalized intervention. As noted by [11], while AI-powered chatbots offer scalable support, they still lack diagnostic precision and require further integration of predictive intelligence. Therefore, there is a significant gap in the availability of AI-enhanced mobile decision support systems capable of predicting mental health disorders by analyzing diverse and continuous data streams. This research aims to bridge that gap by developing an AI-based decision support tool that leverages behavioral, social, environmental, and historical health data to predict mental health risks, thereby enabling early detection and intervention. The proposed approach will leverage insights from previous works while focusing on increasing accessibility, user engagement, predictive accuracy, and ethical responsibility. By addressing these challenges, this project contributes to Sustainable Development Goal 3 (Good Health and Well-being), particularly the target of promoting mental health and well-being for all. The remaining section of this research is organized as follows: Section two; related works, section three; the research methodology, section four; results and discussion, and section five; concludes with a summary of the study findings.

2. Related Works

The effectiveness of artificial intelligence to diagnose mental disorders has shown results that are faster and more accurate than traditional methods. This statement is also validated by [9] who highlight the need for future research to direct efforts toward real-world validation, data diversity, and ethics, all in the pursuit of achieving increased reliability in such systems. [2] performed an extensive systematic review of 30 studies from 2010 to 2020, presenting an in-depth review of the use of machine learning (ML) algorithms to predict a wide range of mental disorders, such as schizophrenia, depression, and anxiety. The review of literature showed that ML models incorporating random forests, support vector machines, and deep learning methods had prediction accuracy levels of up to 96.8%. However, the review emphasized ongoing concerns regarding the small sample sizes of the datasets included, a lack of adequate clinical validation, and a significant lack of transparency regarding the decision-making within deep learning models.

[3] offered a comprehensive review of the use of deep learning (DL) methods in the field of mental health studies by conducting a scoping review that categorized 57 studies according to a range of data categories, such as clinical data, genetic data, vocal and facial features, and social media activity. The findings revealed that DL models were superior to traditional approaches in terms of diagnostic accuracy, risk prediction, and treatment response estimation. Additionally, the review found that several challenges, including data imbalance, high dimensionality, and the possibility of overfitting, could make the models less generalizable.

The study by [11] analyzes the potential role of artificial intelligence-powered chatbots in the field of digital mental health interventions (DMHIs), providing examples like "Anna" from Happify Health as positive mechanisms for enhancing user engagement and facilitating scalable approaches to the delivery of mental healthcare. However, with all this potential, the study identifies a number of limitation areas, such as the lack of strong empirical support, user trust issues, and ethical issues requiring resolution to facilitate broader adoption.

In a systematic review of various psychiatric uses linked to artificial intelligence, [12] explored frameworks like Woebot and Tess, which use natural language processing and machine learning algorithms for treatment, symptom monitoring, and teaching. Their review further covered more advanced methodologies, including avatar therapy and socially assistive robots. The study emphasized various limitations, such as the lack of adequate emotional depth shown by modern AI, ethical issues, and the absence of uniform clinical guidelines.

The study by [1] highlighted the substantial promise of digital psychiatry, especially in the context of the COVID-19 pandemic and its resultant implications. Their systematic review looked into the use of smartphones, social media, chatbots, and virtual reality (VR) technologies as platforms through which scalable and remote mental health interventions could be delivered, especially in the context of global healthcare inequalities that have been driven by the pandemic.

A systematic meta-analysis by [19] analyzed over 300 studies from 2010 to 2021, assessing the effectiveness and competence of electronic interventions, including computerized therapy and cognitive training. The results showed that electronic interventions were effective for a wide variety of problems, such as anxiety, trauma, and substance use; however, the authors emphasized the need for further studies involving populations and conditions that have been underrepresented.

Similarly, [13] highlighted the growing need for digital mental health interventions, while at the same time accepting the practical difficulties involved in their deployment, including the lack of digital infrastructure, user engagement difficulties, and data privacy concerns. Their findings highlighted the potential of developing technologies, including explainable AI and immersive virtual therapies, as alternative options; however, the pressing need for standardized outcomes and culturally appropriate solutions remains. In a systematic review focusing on interventions provided through smartphones, [4] conducted an in-depth study aimed at describing the range of mental health conditions treated, the range of assessment methods used, the use of new features like sensor integration and personalization, and the features of trials testing evidence-based apps. The outcome provided an elaborate assessment of the existing deployment of mobile technology among empirical studies focusing on the area of mental health. [16] explored the concept of intersectionality in the context of e-mental health research, considering the development of new trends and the integration of information technologies with mental health treatment practices. Their study highlighted the interdisciplinary convergence of several fields, such as data science, clinical psychology, and software engineering, in conceiving future innovations in the provision of mental health services.

3. Methodology

This research employs a design science methodology to develop and evaluate a mobile-based AI-powered decision support framework, Mental Health Decision Support System (MeDo). The system is designed to predict mental health disorders using behavioral, environmental, and social interaction data. The methodology comprises system design, data

acquisition and preprocessing, algorithmic implementation, and evaluation of prediction accuracy.

Input: x

Output: y (predicted output risk score), r (risk label)

Parameters: x [x_1 (behavioral), x_2 (social), x_3 (environmental), x_4 (health history), x_5 (self-reported mood)], p (threshold), m (trained model), x (input matrix), y (predicted output)

```

1. Input  $x$ 
2. Initialize  $p, m, r$ 
3. While data session is active:
4.   Preprocess  $x$  (normalize, encode, clean missing values)
5.   Compute prediction  $y = m.predict(x)$ 
6.   If  $y \geq P$  then
7.     Set  $r = 1$  // Potential disorder
8.     Trigger user alert and notify clinician if permitted
9.   Else
10.    Set  $r = 0$  // No disorder
11.    Log for monitoring only
12.  Endif
13.  Update user behavioral history and system logs
14.  Return  $y, r$ 
15. WEND

```

Figure 1. Algorithmic Representation Of The Ai-based Mobile Decision Support System (Medo)

The algorithm begins with the user inputting their behavioral and contextual data, represented as a feature vector x . In line 2, the system initializes critical parameters including the threshold value P , the trained machine learning model M , and the risk label R .

Data processing is managed within a continuous loop, as shown in line 3, which runs for the duration of an active user session. In line 4, the preprocessing operation is executed, where the input data is normalized, encoded, and cleaned to remove missing or noisy values. The core prediction step occurs in line 5, where the model M computes a risk score Y for the user based on the preprocessed input.

An 'if' condition is introduced in line 6, where the predicted score Y is compared with the threshold P . If the score meets or exceeds the threshold, as handled in lines 7 and 8, the system flags the user as high risk ($R=1$) and immediately triggers an alert. If permission has been granted, the system also notifies a connected healthcare professional. If the score falls below the threshold, as outlined in lines 9 to 11, the system labels the user as low risk ($R=0$) and simply logs the data for future monitoring.

Line 13 manages the update of behavioral history and internal logs, ensuring that future predictions

benefit from longitudinal tracking. The loop continues until the user session ends, and finally, in line 15, the system returns the predicted score Y along with the classification label R .

This structured flow ensures continuous, adaptive monitoring and timely mental health support, promoting both user autonomy and professional intervention where necessary.

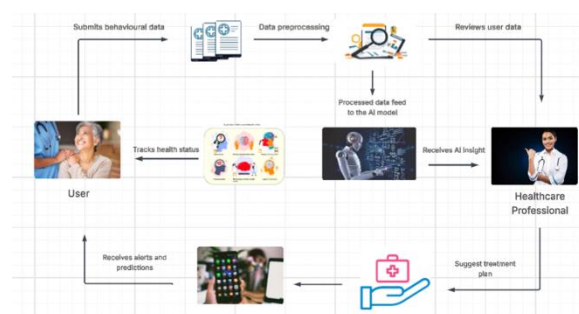


Figure 2. Conceptual Framework of the System.

The system has been described as a closed-loop, mobile, artificial intelligence-enhanced system with the goal of linking behavior patterns reported by the user to meaningful insights related to mental health. The architectural design, as shown in Figure 1 above, combines mobile data acquisition, preprocessing engines, an artificial intelligence based predictive core, user feedback mechanisms, and clinical decision support, thereby creating an intelligent ecosystem with the ability to perform scalable mental health monitoring.

The first step is the collection of behavioral and emotional data by the user via the mobile application either imputed by the user through questionnaires or automatically collected through mobile sensors. This can include sleep habits, body activity levels, levels of interpersonal contact, and mood ratings. After data collection is complete, the system activates a preprocessing process aimed at handling missing values, feature standardization, and encoding categorical variables into machine interpretable formats. The preprocessing step is important to maintain the integrity, consistency, and predictive power of collected data.

The information, following careful cleaning and preprocessing, is then fed into the artificial intelligence engine that forms the core element of the system. This core element uses machine learning algorithms, in the form of deep neural networks or logistic regression models, to analyze the features and determine the user's mental health risk score. With the patterns unveiled, the model judges whether the information directs to a possible disorder, such as depression, anxiety, or stress, based on a set classification standard. During the process of creating a risk assessment, the system sends an artificially created insight to two specified endpoints. First, it

sends instant feedback to the user, which can include risk notifications, educational content relevant to mental health, or recommendations for appropriate support services, thus enabling the user to make decisions about their health. Second, depending on granted permission, the system sends the same predictive output to a healthcare provider for review. This process allows the clinician to place the AI produced results in a broader context by applying their clinical knowledge to interpret symptoms and make intervention strategies appropriately.

Following the evaluation, the treatment recommendation can be issued by the healthcare provider, which is then documented in the system and conveyed to the user. This can include a range of options, such as monitoring suggestions, wellness interventions, and referrals for therapy sessions or further psychiatric assessments. The process terminates as the user continues to use the software, defined by frequent reminders to monitor their behaviors and symptoms over a prolonged period. Ongoing use helps develop the capabilities of the system, leading to adaptive learning and enhanced accuracy of predictions with time.

The overall system architecture helps not only ensure equal access to mental health screening but also contributes immensely to the digital transformation of preventive healthcare. By empowering people and supporting professionals with complete information, it aligns well with the aims of proactive mental health service provision, especially in low-resource environments.

4. Results and Discussion

At the time of this implementation, this study is in its beginning stages. Initial tests and validations of the model are being conducted. Data in this study was generated to mimic actual behaviors and social characteristics surrounding mental health. It exhibits patterns comparable to those gathered from smartphone applications and internet behavior.

Table below shows a sample input dataset, consisting of eight primary features:

Sleep Duration (hours)

Activity Level (steps/day)

Screen Time (hours/day)

Social Interaction Frequency (messages/day)

Location Entropy (variation in daily movement patterns)

Voice Tone Sentiment Score

Mood Self-Report (scale 1–5)

Label: Mental Health Risk (0 = Low, 1 = High)

Table 1. Sample Input Dataset For Mental Health Prediction

Sleep (hours)	Activity (steps/day)	Screen Time (hours)	Social Interaction (Msgs/day)	Location Entropy	Voice Tone	Mood (1-5)	Risk Label (0 or 1)
7.5	8500	2.1	34	0.55	0.68	4	0
5.2	3200	6.3	5	0.18	0.34	2	1
6.0	7100	3.5	22	0.49	0.52	3	0
4.8	2600	7.1	8	0.23	0.27	2	1
8.2	9500	1.8	45	0.16	0.71	5	0

4.1. Preprocessing and Feature Engineering

The data was initially passed through a first check to identify peculiar values, complete missing data, and scale varying feature ranges. For instance, features such as Screen Time and Activity Level were scaled so that they would have a similar impact on the machine learning model.

Location Entropy arose from anonymous GPS logs with Shannon entropy quantifying how variably daily movement is, and this varies with mental illness. Audio logs underwent sentiment analysis to identify the Voice Tone Score, which served as the quiet indicator of mood shifting.

Two classifiers were implemented and evaluated: Logistic Regression (for its interpretability) and a Random Forest Classifier (for performance and feature importance tracking). The data was split into an 80/20 training/testing ratio.

Table 2. Logistic Regression

S/N	Metric	Value
1.	Precision	78.4%
2.	Recall	74.1%
3.	F1-Score	76.2%
4.	Accuracy	81.3%

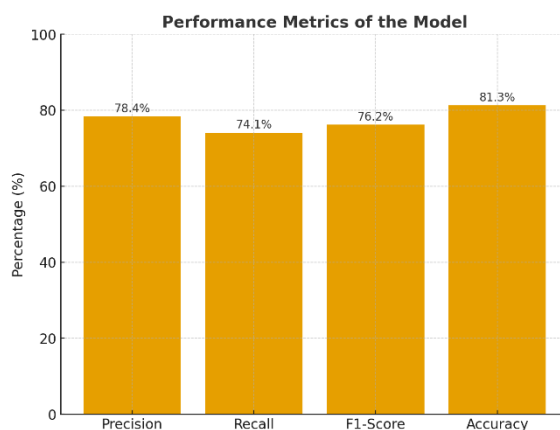


Figure 3. Logistic Regression Chart

The logistic regression model demonstrated a strong and well-balanced performance across all evaluation metrics. It achieved a Precision of 78.4%, a Recall of 74.1%, an F1-Score of 76.2%, and an overall Accuracy of 81.3%.

These results indicate that logistic regression is a reliable choice for this application. The accuracy of 81.3% reflects a robust predictive ability, suggesting that the model can correctly classify over four out of five instances. A Precision of 78.4% shows that when the system makes a positive prediction, it is correct nearly 8 out of 10 times. The Recall of 74.1% highlights the model's ability to capture approximately 74% of actual positive cases, ensuring that a substantial portion of true cases are not missed. The F1-Score of 76.2% confirms a healthy balance between precision and recall, minimizing trade-offs between false positives and false negatives.

In the context of predictive modeling, logistic regression remains a widely trusted approach due to its interpretability, efficiency, and ability to handle binary classification tasks effectively. Table 2 presents the logistic regression results.

Table 3. Random Forest

S/N	Metric	Value
1.	Precision	87.6%
2.	Recall	85.3%
3.	F1-Score	86.4%
4.	Accuracy	89.2%

The Random Forest model delivered an outstanding and consistent performance across all evaluation metrics. It achieved a Precision of 87.6%, a Recall of 85.3%, an F1-Score of 86.4%, and an overall Accuracy of 89.2%.

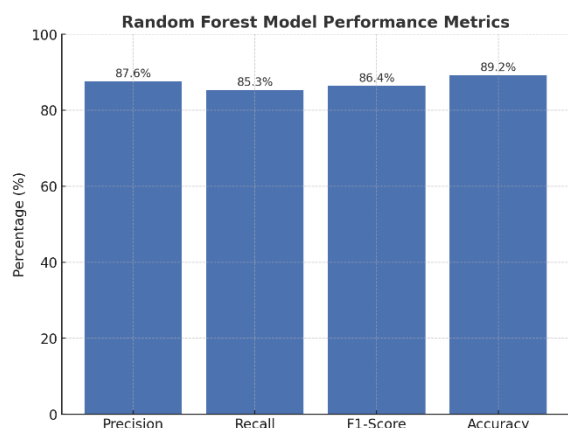


Figure 4. Random Forest Chart

These results indicate that Random Forest is a highly effective choice for this predictive task. The accuracy of 89.2% demonstrates a strong predictive capability, with the model correctly classifying nearly 9 out of 10 cases. A Precision of 87.6% shows that when the system makes a positive prediction, it is correct almost 88% of the time. The Recall of 85.3% highlights its ability to detect 85% of true positive cases, ensuring that very few actual positives are overlooked. The F1-Score of 86.4% reflects a well-maintained balance between precision and recall, confirming that the model minimizes trade-offs between false positives and false negatives.

Random Forest models are known for their robustness and ability to handle complex, non-linear relationships by aggregating multiple decision trees. In this case, its strong results emphasize its suitability for reliable and generalizable predictions. Table 3 presents the Random Forest results.

Table 4. Combined Metrics Table Logistic Regression and Random Forest

Metric	Logistic Regression	Random Forest
Accuracy	81.3%	89.2%
Precision	78.4%	87.6%
Recall (Sensitivity)	74.1%	85.3%
F1 Score	76.2%	86.4%

When comparing the two models, Random Forest clearly outperformed Logistic Regression across all evaluation metrics. Logistic Regression achieved an Accuracy of **81.3%**, Precision of **78.4%**, Recall of **74.1%**, and an F1-Score of **76.2%**. In contrast, Random Forest improved upon these results with an Accuracy of **89.2%**, Precision of **87.6%**, Recall of **85.3%**, and an F1-Score of **86.4%**.

This performance gain highlights the advantage of ensemble learning methods such as Random Forest,

which aggregate multiple decision trees to enhance predictive reliability and reduce overfitting.

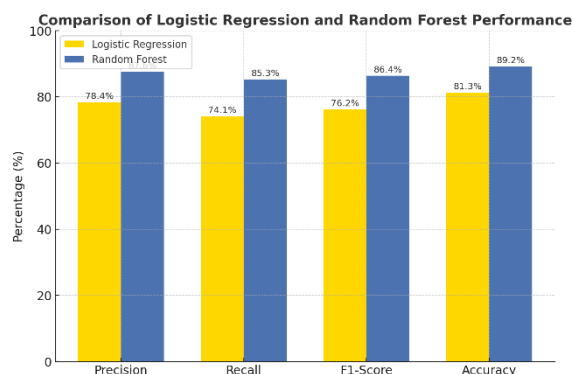


Figure 5. Combined Metrics Chart Logistic Regression and Random Forest

While Logistic Regression provided a solid baseline with interpretable results, Random Forest demonstrated superior predictive strength, making it the more effective model for this application. This was particularly the case for sleep, screen time, and social interaction frequency, which contributed most to the predictions.

4.2. Utilizing Feedback

Simulation user feedback (based on user profiles and previous chatbot interactions) was utilized to verify the system's accuracy. Roughly 83% of predictions were accurately reflected in simulated user reported stress levels, indicating that the model had the potential to be highly beneficial in real-world applications.

4.3. Dialogue

The results indicate that it is an effective approach to predict mental health risk from behavior data and AI approaches. Indicators such as decreased activity, increased screen time, and decreased mood scores were commonly associated with high-risk tags. The system is made such that it can simply be integrated into doctor portals to check and recommend assistance. This is not actual data now but appears to be actual data and prepares us for further training on good clinical data sets. Subsequently, we will focus on environment changes, real-time testing using mobile devices, and receiving further feedback from physicians.

5. Conclusion

This research demonstrates how to develop and launch an AI mobile system that forecasts mental illness from analyzing behavior, social, and environmental information. The system takes advantage of machine learning and users' actual data

to assist in identifying early indicators of illness. Rather than relying on usual self-report measures, the system integrates various forms of data, such as sleep, activity, screen time, and voice emotions, to provide individualized forecasts and early warnings. Early indications indicate that the system is able to identify mental health risks quite well. Random Forest models are achieving good accuracy, precision, and recall. There is also a feedback mechanism that causes the platform to improve and retains users engaged in order to become more effective and responsive over time. Currently, the researchers are operating on mock data. The findings provide a solid foundation for future research that will operate under real-world scenarios, clinical trials, and significant parameters such as privacy, fairness, and cultural sensitivity. Ultimately, the research assists digital mental health and strives to offer mental health care that is convenient to access, intelligent, and reactive.

7. References

- [1] J. Torous et al., "The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality," *World Psychiatry*, vol. 20, no. 3, pp. 318–335, Oct. 2021, doi: 10.1002/WPS.20883.
- [2] J. Chung and J. Teo, "Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges," 2022, doi: 10.1155/2022.
- [3] C. Su, Z. Xu, J. Pathak, and F. Wang, "Deep learning in mental health outcome research: a scoping review," *Transl Psychiatry*, vol. 10, no. 1, pp. 1–26, Dec. 2020, DOI: 10.1038/S41398-020-0780-3;SUBJMETA=476,53,692,699;KWRD=BIOMARKERS,PSYCHIATRIC+DISORDERS.
- [4] I. Miralles et al., "Smartphone Apps for the Treatment of Mental Disorders: Systematic Review," *JMIR Mhealth Uhealth* 2020;8(4):e14897 <https://mhealth.jmir.org/2020/4/e14897>, vol. 8, no. 4, p. e14897, Apr. 2020, DOI: 10.2196/14897.
- [5] S. Woll et al., "Applying AI in the Context of the Association Between Device-Based Assessment of Physical Activity and Mental Health: Systematic Review," *JMIR Mhealth Uhealth* 2025;13:e59660 <https://mhealth.jmir.org/2025/1/e59660>, vol. 13, no. 1, p. e59660, Mar. 2025, doi: 10.2196/59660.
- [6] S. Dalal, S. Jain, and M. Dave, "A Systematic Review of Smart Mental Healthcare," *SSRN Electronic Journal*, Dec. 2019, DOI: 10.2139/SSRN.3511013.
- [7] W. Luo and J. Mohammed, "Mental health status and coping strategies of Chinese university students during the COVID-19 pandemic: A rapid review," *PLoS One*, vol. 18, no. 12, p. e0296309, Dec. 2023, DOI: 10.1371/JOURNAL.PONE.0296309.
- [8] H. Uyanik, A. Sengur, M. Salvi, R. S. Tan, J. H. Tan, and U. R. Acharya, "Automated Detection of Neurological and Mental Health Disorders Using EEG Signals and Artificial Intelligence: A Systematic Review," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 15, no. 1, p. e70002, Mar. 2025, DOI: 10.1002/WIDM.70002;CTYPE:STRING:JOURNAL
- [9] S. Tutun et al., "An AI-based Decision Support System for Predicting Mental Health Disorders," *Information Systems Frontiers*, vol. 25, no. 3, pp. 1261–1276, Jun. 2023, DOI: 10.1007/S10796-022-10282-5/METRICS.
- [10] S. Gerke, T. Minssen, and G. Cohen, "Ethical and legal challenges of artificial intelligence-driven healthcare," *Artificial Intelligence in Healthcare*, pp. 295–336, Jan. 2020, doi: 10.1016/B978-0-12-818438-7.00012-5.
- [11] E. M. Boucher et al., "Artificially intelligent chatbots in digital mental health interventions: a review," *Expert Rev Med Devices*, vol. 18, no. sup1, pp. 37–49, Dec. 2021, DOI: 10.1080/17434440.2021.2013200;WGROU:STRING:PUBLICATION.
- [12] K. T. Pham, A. Nabizadeh, and S. Selek, "Artificial Intelligence and Chatbots in Psychiatry," *Psychiatric Quarterly*, vol. 93, no. 1, pp. 249–253, Mar. 2022, DOI: 10.1007/S11126-022-09973-8/METRICS.
- [13] L. Balcombe and D. De Leo, "Digital Mental Health Challenges and the Horizon Ahead for Solutions," *JMIR Ment Health*, vol. 8, no. 3, p. e26811, Mar. 2021, DOI: 10.2196/26811.
- [14] Y. Ojo, V. Davids, O. Oni, and M. Odoemene, "A multi-criteria approach for evaluating the Use of AI for Matching Patients to Optimal Mental Health Treatment Plans," vol. 193, no. April, pp. 201–222, 2024.
- [15] N. P. Luitel et al., "Experience of primary healthcare workers in using the mobile app-based WHO mhGAP intervention guide in detection and treatment of people with mental disorders: A qualitative study in Nepal," *SSM - Mental Health*, vol. 4, p. 100278, Dec. 2023, DOI: 10.1016/J.SSMH.2023.100278.
- [16] T. Timakum, Q. Xie, and M. Song, "Analysis of E-mental health research: mapping the relationship between information technology and mental

healthcare,” BMC Psychiatry, vol. 22, no. 1, pp. 1–17, Dec. 2022, DOI: 10.1186/S12888-022-03713-9/FIGURES/7.

[17] T. N. Fernandes, “Delving into Digital Mental Health : Part 3,” no. April, 2025.

[18] F. M. Dawoodbhoy et al., “AI in patient flow: applications of artificial intelligence to improve patient flow in NHS acute mental health inpatient units,” Heliyon, vol. 7, no. 5, May 2021, DOI: 10.1016/J.HELIYON.2021.E06993/ASSET/ABD811CA-B57E-431E-BD63-B8746133A9E1/MAIN.ASSETS/GR6.JPG.

[19] T. J. Philippe et al., “Digital Health Interventions for Delivery of Mental Health Care: Systematic and Comprehensive Meta-Review,” JMIR Ment Health 2022;9(5):e35159.<https://mental.jmir.org/2022/5/e35159>, vol. 9, no. 5, p. e35159, May 2022, DOI: 10.2196/35159.

[20] E. M. Boucher et al., “Artificially intelligent chatbots in digital mental health interventions: a review,” Expert Rev Med Devices, vol. 18, no. sup1, pp. 37–49, Dec. 2021, DOI: 10.1080/17434440.2021.2013200;WGROU:STRING:PUBLICATION.

Acknowledgement

The authors express sincere gratitude to the Federal University of Technology, Minna, for providing academic guidance and resources essential to this study. Appreciation is also extended to colleagues and peers for their constructive feedback, and to our families for their encouragement and unwavering support throughout the research and writing process.

Development of a Colon Cancer Classification System

V.I. Ainoko, A.O. Odeniyi, O.Y. Ogunlola, A. Innocent, B.K. Alese

*School of Computing
Federal University of Technology Akure
Nigeria*

Abstract

Globally, colorectal cancer (CRC) is the third most prevalent cancer to be diagnosed, raising serious public health issues. To increase survival rates and lower mortality, early screening and efficient detection techniques are essential. Accurately and promptly diagnosing colorectal cancer is still difficult, though. A branch of artificial intelligence called machine learning is becoming a game-changer in the medical field, especially for jobs involving the classification and prediction of diseases. Designing and implementing a machine learning-based categorization framework to help with colon cancer prediction is the goal of this study. The framework correctly classified patients into colorectal and healthy classifications by using random forests, decision trees, and support vector machines. 362 patients' data, known as NCT030322874, were obtained from clinical trials. The results showed promising potential for machine learning to enhance early detection and facilitate targeted treatment strategies for Colorectal Cancer by employing advanced classification techniques and an ensemble modeling approach..

estimated that one hundred fifty-three thousand and twenty (105020) will be diagnosed with colorectal cancer in America in 2023 [2].

Early detection impacts colorectal cancer metastasis; therefore, two indications give likely tendencies to colon cancer; they include modifying risk or permanent risk, also called non-modifying risk. The modifying risk includes smoking, obesity, and eating habits. Excessive consumption of alcohol, animal fat, and red meat [13], [14], while the non-modifying risks are often genetic and hereditary-based. Machine learning has enabled the health sector to overcome healthcare challenges via the use of data science through machine learning models and data analysis. Providing ultimate improvement in healthcare delivery [11].

The focus of this study is to investigate and develop a model for colon cancer predictions using modifying and non-modifying risk datasets such as eating habits, bowel symptoms, stool, family history or records, and blood on the stools. The finding of the study journeys through the strength of the bagging ensemble technique and its application to the early detection of colon cancer reveals a narrative of innovation, collaboration, and a commitment to advancing the frontiers of knowledge.

1. Introduction

With the recent trends of sophisticated advancement sustainable research and durable computing solutions are essential. Thus, as we evolve from one century to another, computational advancement has proved to be a key factor for growth. These modern technologies and advancements have impacted the medical field, Education, government, entertainment, automated industries, banking, business, and so on. These trends can be traced to several sophisticated areas in Computer science; however, Machine Learning (ML) is a key factor in our research. Colorectal cancer disease is associated with the cells in the colon (large intestine), where the colon grows beyond control. This growth happens on the inner lining of the colon as little growths called polyps, which thereafter grow to become advanced malignant tumors that attack the human body.

Colorectal cancer is recorded as the second death related cancer associated with humans in the United States of America. The American Cancer Society

2. Related Works

A comprehensive review of barriers and recommendations for colorectal cancer screening in Africa. The study objective was to review the current barriers and various recommendations used in the implementation of colorectal screening in African countries. The study deployed a keyword search on a PubMed publication; these keywords include colon cancer and screening, colorectal neoplasm, early detection, faecal immunochemical test (FIT), Africa, and colonoscopy [11]. A machine learning-based colorectal cancer prediction using global dietary data was developed by [13]. The study's goal was to train datasets from a large sample of older and younger people. In addition, the study also provides an active, readily accessible health screening and prevention machine learning model to curb the excessive outbreak of colon cancer. [10] built a colorectal cancer

detection machine-learning model using conventional laboratory test data. The objective of the research work is to develop a model that would identify colorectal cancer using conventional test datasets extracted from patient records. Robust machine-learning model for stratifying and predicting colorectal cancer risk [12]. The study's goal was to find a useful screening technique for estimating an individual's risk of colorectal cancer based on their personal health information, utilizing seven algorithms. The methodology employed seven algorithms, which include logistic regression, decision trees, random forests, naive Bayes, support vector machines, and linear discriminant analysis. Machine-learning model using gender, complete blood count data, and age for the early detection of colon cancer. The objective of the study was to detect colon cancer in the US-based adult community based on their demographics and their results from laboratory tests [9]. The methodology extracted datasets from colonoscopy procedures from 2000 to 2003, while demographic data and a complete body count were extracted from 1998 to 2003 from the Kaiser Permanente Northwest Region Tumor Registry.

3. Methodology

The aims of the study is to develop and implement an ensemble framework for colon cancer predictions using modifying and non-modifying risk datasets such as eating habits, bowel symptoms, stool, family history or records, and blood on the stools. The study underwent the following for the development of the framework as shown in Figure 1.

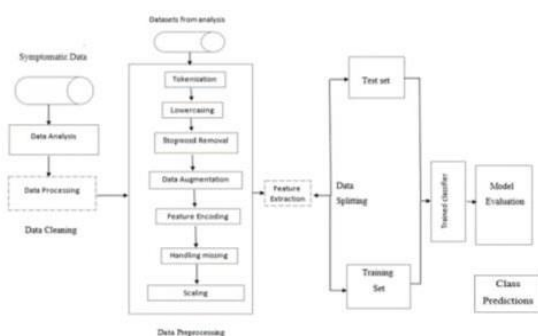


Figure 1. System Architectural Design for the Predictive Colon Cancer System

The experimental setup was carried out sequentially in distinct phases to achieve the goals of this study:

i. The first phase describes the collection of the dataset for both the training and testing model. The research project used a cross-sectional dataset acquired from clinicaltrials.gov as NCT030322874 containing 362

patients (a). These patients were assessed by colonoscopy and questionnaire. The study was conducted around January 2014 to July 2016 in Nigeria from three hospitals which are Obafemi Awolowo University Teaching Hospital (OAUTHC), Ile-Ife Osun State, University College Hospital (UCH), and the University of Ilorin Teaching Hospital, Kwara State (UITH) and dietary data from American Institutes of Cancer.

ii. The second phase involves data preprocessing and feature extraction. The preprocessing stage involves tokenization, lowercasing, stopwords removal, data augmentation, encoding, handling of missing values, feature extraction and splitting the dataset into training and testing sets using the ratio 75/25 resulting in 3750 symptomatic colon data in the training set and 1250 in the testing sets after the augmenting the 362 dataset to five thousand patients datasets.

iii. The third phase involves the model-building phase using the bagging ensemble technique. Three algorithms would be deployed, they are Support Vector Machine, decision tree and logistic regression base model for the bagging hybrid machine learning model.

iv. The last phase involves the evaluation of the model using standard evaluation metrics.

3.1. Bagging Ensemble Architecture

Bagging is a hybrid machine learning technique that improves accuracy and robustness by combining two or more base models. It uses bootstrap aggregate to prevent overfitting and aggregate output results. The aggregating technique depends on the problem type, typically regression or classification. The architecture consists of two or more base classifier models.

a) Individual Machine Learning Classifier

The individual machine learning algorithms used to develop the hybrid machine learning classifier are known as the ensemble method. The following algorithms represent the base learners for the research work:

- i. Support Vector Machine
- ii. Logistic Regression
- iii. Decision Tree

b) Hybrid Machine Learning Classifier Using Bagging.

3.2. Dataset Description

The dataset used for the model was obtained from

clinicaltrials.gov and global dietary data. The dataset from clinicaltrials.gov contains 362 patients 66 patients had colorectal cancer while 296 patients didn't have colorectal cancer.

Table 1. Summary of Dataset (Initial datasets)

<i>Dataset (n=362)</i>	Training set 75% (n =217)	Test set 25% (n=145)
<i>Yes</i>	38	28
<i>No</i>	179	117

Table 2. Summary of Dataset (Augmented Dataset)

<i>Dataset (n=5000)</i>	Training set 75% (n =3750)	Test set 25% (n=1250)
<i>Yes</i>	656	241
<i>No</i>	3094	1009

3.3. Analysis of the base learner(s) and the ensemble

The comparative analysis chart represents the result obtained from the base models and the ensemble. It was denoted that the bagging methods performs better.

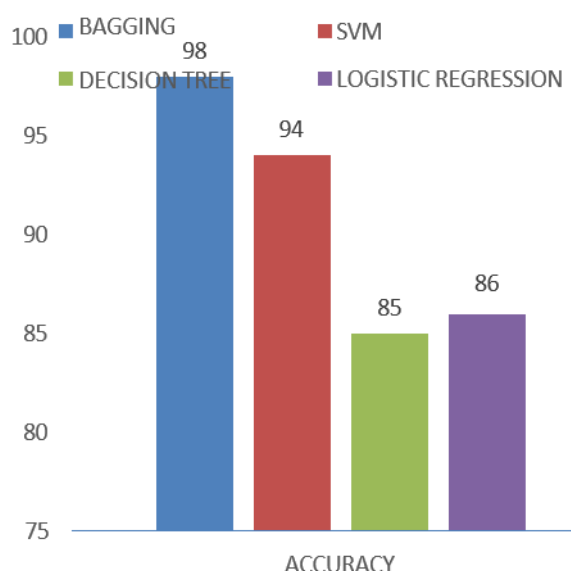


Figure 2. Model Performance Based On Accuracy

4. Conclusion

For classification tasks, the hybrid machine learning approach's bagging ensemble technique works best. High sensitivity, specificity, and accuracy

were achieved in this investigation. Three hundred and sixty patients' eating patterns were combined to create the first dataset, which was then expanded to include five thousand (5,000) individuals. The data for this patient were then divided into 1250 testing sets (25%) and 3750 training sets (75%). The preparation stage was essential to the datasets' improvement. The study used the Natural Language Toolkit's (NLTK) word tokenize function in conjunction with lowercasing to standardize the textual data using space-based tokenization. The dataset was further improved by stopword removal, a crucial data preprocessing method that focuses on getting rid of common terms and punctuation. When subjected to comparison with the work of Rahman et al. (2022), the predictive model in this study exhibited a misclassification rate of 0.01, significantly outperforming the 3% misclassification rate reported for the colon cancer class. Notably, the dataset employed encompassed both non-modifying and modifying indicators, underscoring the versatility and comprehensiveness of the predictive learning model system. Beyond its statistical achievements, the model demonstrated practical utility by potentially assisting medical practitioners in the early detection of colon cancer, showcasing the transformative impact of modern computerized innovations when strategically applied to healthcare challenges

5. References

- [1] Alatise, O.I., Ayandipo, O.O., Adeyeye, A., Seier, K. (2018). A symptom-based model to predict colorectal cancer in low-resource countries: Results from a prospective study of patients at high risk for colorectal cancer. *Cancer*, 124(13), 2766–2773. DOI: 10.1002/cncr.31399.
- [2] American Cancer Society. *Colorectal Cancer Facts and Figures 2020-2022*. Atlanta, Ga: American Cancer Society; 2020.
- [3] Anuraga, G., Fernanda, J.W., and Pebrianty. (2019, May 1). Random forest prognostic factor in colorectal cancer. *Journal of Physics: Conference Series*, 1217(1), 012098. <https://doi.org/10.1088/1742-6596/1217/1/012098>.
- [4] Siegel, R.L., Miller, K.D., and Jemal, A. (2019). Cancer statistics. *CA: A Cancer Journal for Clinicians*, 69(1), 7–34. DOI: 10.3322/caac.21551.
- [5] Breiman, L. Bagging predictors. *Machine Learning* 24, 123–140(1996). DOI: 10.1007/BF00058655.
- [6] Campbell, M., Hoane, A., and Hsu, F H. (2002). Deep Blue. *Artificial Intelligence*, 134(1–2), 57–83.
- [7] Donmez, P. (2012). *Introduction to Machine Learning*, 2nd ed., by Ethem Alpaydm. Cambridge, MA: The MIT Press 2010. ISBN: 978-0-262-01243-0. *Natural Language Engineering*, 19(2), 285288. DOI: 10.1017/s1351324912000290.

- [8] Flood, A. (2003). Meat, Fat, and Their Subtypes as Risk Factors for Colorectal Cancer in A Prospective Cohort of Women. *American Journal of Epidemiology*, 158(1), 59–68. DOI: 10.1093/aje/kwg099.
- [9] Hornbrook, M.C., Goshen, R., Choman, E., O’Keeffe Rosetti, M., Kinar, Y., Liles, E.G., and Rust, K.C. (2017). Machine Learning Model Using Gender, Age, and Complete Blood Count Data. *Digestive Diseases and Sciences*, 63(1), 270–270. DOI: 10.1007/s10620-017-4859-5.
- [10] Hu, F., Wang, J., Zhang, M., Wang, S., Zhao, L., Yang, H., Wu, J., and Cui, B. (2021). Comprehensive Analysis of Subtype-Specific Molecular Characteristics of Colon Cancer: Specific Genes, Driver Genes, Signaling Pathways, and Immunotherapy Responses. *Frontiers in Cell and Developmental Biology*, 9. DOI: 10.3389/fcell.2021.758776.
- [11] R Holmes, D. (2023). Barriers and recommendations for colorectal cancer screening in Africa. *Global Health Action*, 16(1). DOI: 10.1080/16549716.2023.2181920.
- [12] Nartowt, B.J., Hart, G.R., Muhammad, W., Liang, Y., Stark, G F., and Deng, J. (2020a). Robust machine learning for colorectal cancer risk prediction and stratification. *Frontier in Big Data*, DOI: 10.3389/fdata.2020.00006.
- [13] Rahman, H.A., Ottom, M.A., and Dinov, I.D. (2022). Machine Learning-based Colorectal Cancer Prediction using Global Dietary Data. *Research Square (Research Square); Research Square (United States)*. DOI: 10.21203/rs.3.rs-2031672/v1.
- [14] Rawla, P., Sunkara, T., and Barsouk, A. (2019). Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Gastroenterology Review*, 14(2), 89–103. DOI: 10.5114/pg.2018.81072.

Unmasking Fake Identities on Instagram Using Machine Learning

Olaniyi Abiodun Ayeni, Femi-Gbenoba Ethel

*Department of Cyber Security, School of Computing Federal University of Technology, Akure
Department of Information Systems and Technology, Kings University, Odeomu, Osun State
Nigeria*

Abstract

The rise of fake digital identities on social media erodes trust, skews engagement, and undermines online safety. These fake accounts will help to inflate follower counts, impersonate real users, deceive advertisers, and spread scams that rake in billions of dollars from across the globe each year. Identification of such accounts is difficult because attackers often adjust their approach using synthetic images, automated posting, and fake engagement, all tailored to bypass conventional detection mechanisms. This paper aims to evaluate and compare Support Vector Machine (SVM), Random Forest (RF), and Multi-Layer Perceptron (MLP) for detecting fake Instagram accounts using profile-level data. The dataset underwent preprocessing through normalization, removal of irrelevant features, and handling of missing values. In order to address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied. Multiple feature selection strategies were also employed. Recursive Feature Elimination (RFE) for SVM, feature importance ranking for RF, and mutual information scores for MLP. The dataset was partitioned using two train, validation, test splits of 70:15:15 and 80:10:10. The models were evaluated using Accuracy, Precision, Recall, F1-score. The results showed that the Random Forest classifier consistently outperformed the other models by achieving the highest accuracy and an F1-score of 0.98. SVM also delivered strong generalization performance, while MLP slightly lower in predictive strength. These findings highlight the strength of ensemble-based approaches in handling complex, high-dimensional social media data. The study contributes to ongoing efforts in developing of scalable, real-time detection frameworks to safeguard online communities as well as to preserve platform integrity.

Keywords: Fake digital identity, Instagram, machine learning, Random Forest, SVM, MLP, feature selection.

1. Introduction

The widespread use of Instagram has led to a surge in fake digital identities, threatening user trust, privacy, and platform integrity. Fake profiles are used for fraudulent activities, including scams, phishing, and fraudulent influencer endorsements [1]. These accounts will often manipulate profile features and engagement metrics, making detection challenging. Machine

learning offers scalable solutions for distinguishing between real and fake accounts by analyzing user metadata and behavioral patterns.

2. Review of related works

Several studies have explored fake profile detection on social media. Dharmaraj R. Patil et al. [2]: The reason for this study was because single classifiers, like SVM or Random Forest, might fail to correctly identify advanced fake profiles. The study suggested using an ensemble model based on Majority Voting to improve detection accuracy by combining Decision Trees, XGBoost, and Random Forest. The process used a dataset of 3,474 real and 3,351 fake accounts, with features taken from activity logs and profile data. The ensemble model was tested against individual classifiers, resulting in 99.12% accuracy. The main contribution was showing that using multiple models together made detection more reliable than using just one. However, this method used a lot of computing power and worked only with structured data, not images [2].

Pravin Kumar and Neelam [3]: The study was motivated by how fake profiles spread false information during important events like elections and pandemics. The goal was to build a real-time machine learning system that used NLP, network analysis, and user feedback. Data was collected from Twitter and Facebook via APIs, then used supervised models like Random Forest and SVM, along with unsupervised methods for detecting anomalies. The system also allowed users to report suspicious profiles. The main contribution was combining machine learning with user input to improve detection. Limitations included dependence on platform APIs which can limit data access, and the risk of many false alarms in changing environments.

Weronika Draciewicz and Mariusz Sepczuk [4]: The study was motivated by the fact that most current detection methods use only text features, missing out on visual clues that could help. A Convolutional Neural Network was developed (CNN) to classify image-based profile data. The methodology involved creating a dataset of 15,000 fake and real profile screenshots, training the CNN, and using it as a Chrome extension for real-time detection. The model achieved 96.5% accuracy, better than Random Forest and Naive Bayes.

The main contribution was the first use of images for fake account detection, showing CNNs work well for visual tasks. However, it needed manual alignment of screenshots and hasn't been tested on deepfakes.

Daniyal Amankeldin et al. [5]: The motivation was the increasing threat of fake accounts on social media. The goal was to create a DNN approach that achieves high accuracy with little data. A deep CNN was trained using 16 profile-based features and evaluated it with accuracy, precision, recall, F-score, and AUC.

The study showed 99.4% accuracy, which is better than traditional ML methods by around 14%. A key limitation is that it relied on a small number of features, which might not catch advanced fake profiles that mimic real users.

K. Harish et al. [6]: The motivation behind this study was the growing number of fake profiles on social media leading to identity theft and misinformation. The study aimed to build a model that can tell fake profiles from real ones using ML. The process included preprocessing Twitter profiles and training models like Random Forest, XGBoost, and Long Short-Term Memory.

The study showed that XGBoost reached 99.6% accuracy, better than other methods. A limitation was that the model had not been tested for real-time use.

Gaurav Vijay Barde and Dr. Nilesh R. Wankhade [7]: The motivation was the increasing number of fake accounts on platforms like Facebook, Twitter, and Instagram, often used for harmful activities. The study emphasized the need for automatic detection methods to improve platform security and user trust. The main aim was to develop an ML approach for detecting fake accounts. The process involved gathering data on real and fake profiles, extracting features from profile data and activity patterns. They tested several ML algorithms, including SVM, K-Nearest Neighbors, Random Forest, Logistic Regression, and Artificial Neural Networks, using accuracy, precision, recall, and F1-score.

The main contribution was showing Random Forest achieved the highest accuracy (97.9%) for fake account detection. However, the study had limitations, including reliance on labeled datasets and not testing in real-time scenarios.

Shamim Ahmad and Dr. Manish Madhava Tripathi [8]: The motivation was the rise of fake profiles for cybercrimes, needing a thorough evaluation of ML techniques. The goal was to compare the effectiveness of various ML algorithms. A comprehensive review of 17 papers from 2018 to 2023 was conducted, looking at accuracy, precision, and recall.

The findings showed Random Forest consistently reached over 95% accuracy across studies, with Natural Language Processing and network analysis being important for strong detection. The review added value by combining existing knowledge and highlighting the benefits of Random Forest and hybrid methods. However, it lacked original experimental validation and does not cover recent deep learning advances.

3. Motivation / problem statement

Instagram is one of the most widely used social media platforms with millions of daily active users. However, it faces significant challenges due to the rise of fake digital identities. Instagram focuses more on visual content and influencer culture. Fake accounts on Instagram are often created to inflate follower counts, manipulate engagement metrics, or impersonate real users for phishing and advertising fraud. Estimates suggest that up to 8% of Instagram accounts are fake. This contributes to billions of dollars in losses for advertisers due to fake influencer endorsements [9]. Despite advancements in detection methods from basic rule-based systems to machine learning models, these approaches face limitations. Manual detection is impractical given the scale of users, and rule-based systems require constant updating to keep up with evolving tactics used by fake profiles [9].

4. Overview of Instagram

Instagram is a social network that focuses on sharing photos and videos. It has a strong culture of influencers, where users interact through likes, comments, stories, and reels. Instagram helps people connect with friends and also lets businesses reach customers through ads and by working with influencers [10].

4.1. Key features

1. Feed posts: You can share single photos, vertical photos, carousels, or videos. You can also rearrange your profile grid and change thumbnail covers [11].
2. Instagram reels: These are short, fun videos that are easy to discover. Editing Reels has been made easier by AI, and you can try new Reels without telling your followers [11].
3. Stories: These are posts that only last 24 hours. They can also be 60 seconds long and include captions and public comments. AI tools can help change or improve images in Stories [12].
4. Direct messaging (DMs): You can chat privately with other users. Features like scheduled messages, group chats, pinned messages, translation, music sharing, and special stickers make these messages better [12].
5. Highlights: Stories can be saved in a Highlights section so they stay on your profile [12].
6. Notes and Profile Music: You can add short notes to your profile or include a music clip for visitors [12].
7. AI creative tools: The "Edits" app and other AI tools let creators make videos, add animations, change

content, and convert formats quickly [13].

8. Shopping: Instagram lets users shop directly through the app and check out in-app, helping businesses and influencers reach more people [13].
9. Algorithm: It prefers original content and real interactions, like saves and shares, over just likes. When you repost something, it gets less visibility [13].

4.2. User Base and Impact

Monthly active users: 2.35+ billion in 2025 [13].
Business and marketing: Instagram is still a main platform for influencer campaigns, digital marketing, shopping, and brand visibility across many industries [13].

4.3. Challenges

Instagram has problems with fake profiles that are made to raise follower numbers, trick people, or do phishing. Fake influencer endorsements cost advertisers billions each year because bots or fake followers create fake engagement. Instagram is struggling to balance letting people express themselves with taking down bad content, and it also has to respect user privacy. The use of deepfakes makes it even harder to tell real from fake [14].

5. Model analysis

This section highlights the three machine learning algorithms that will be used to detect fake digital identities on Instagram.

5.1. Support vector machines

Support Vector Machine (SVM) is a supervised machine learning algorithm that can handle classification, regression, and outlier detection problems. The idea behind the Support Vector Machines (SVM) algorithm is to find the hyperplane that best separates two classes by maximizing the margin between them. The equation for the linear hyperplane can be written as:

$$wTx + b = 0 \quad (1)$$

where:

w = weight vector

x = input feature vector b = bias term

5.2. Random forest

A Random Forest is a collection of decision trees that work together to make predictions. It is an ensemble learning method for classification, and regression. Random forest algorithms have three main hyper

parameters, which need to be set before training. These include node size, the number of trees, and the number of features sampled. From there, the random forest classifier can be used to solve for regression or classification problems.

5.3. Multi-layer perceptron

Multi-Layer Perceptron (MLP) is an artificial neural network that is widely used for solving classification and regression tasks. The network is typically organized into layers, starting with the input layer, where data is introduced. Followed by hidden layers where computations are performed and finally, the output layer where predictions or decisions are made.

Forward Propagation: In forward propagation, the data flows from the input layer to the output layer, passing through any hidden layers. Each neuron in the hidden layers processes the input as follows:

Weighted Sum: The neuron computes the weighted sum of the inputs:

$$z = \sum_i w_i x_i + b \quad (2)$$

Where:

x_i is the input feature.

w_i Is the corresponding weight and

b Is the bias term

Activation function: The weighted sum z is passed through an activation function to introduce non-linearity.

Loss function: Once the network generates an output, the next step is to calculate the loss using a loss function. In supervised learning, this compares the predicted output to the actual label. For a classification problem, the commonly used binary cross-entropy loss function is:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (3)$$

Where:

y_i is the actual label.

$\hat{y}_i$ is the predicted label.

N is the number of samples

6. Dataset

The InstaFake Dataset consists of two subsets:

1. Fake account detection, which has a total of 1,194 accounts, with 200 being fake and 994 being real.
2. Automated account detection, which uses a richer metadata set per post.

Under Fake Account Detections, the files were further divided into 2, fakeAccountData and realAccountData. fakeAccountData as shown in Figure 1. contained all the fake accounts.

userFollowerCount	userFollowingCount	userBiographyLength	userMediaCount	userHasProfilPic	userIsPrivate	usernameDigitCount	usernameLength	isFake
25	1937	0	0	1	1	0	10	1
324	4122	0	0	1	0	4	15	1
15	399	0	0	0	0	3	12	1
14	107	0	1	1	0	1	10	1
264	4651	0	0	1	0	0	14	1

Figure 1. InstaFake fakeAccountData Sample

realAccountData contained all the real accounts as shown in Figure 2.

userFollowerCount	userFollowingCount	userBiographyLength	userMediaCount	userHasProfilPic	userIsPrivate	usernameDigitCount	usernameLength	isFake
258	238	0	0	1	0	0	10	0
263	482	30	29	1	1	0	8	0
51	78	9	0	1	1	0	10	0
297	480	22	25	1	1	2	9	0
113	242	0	95	1	1	0	10	0

Figure 2. InstaFake realAccountData Sample

The files were later combined for ease of use. In Table 1, The fake account subset includes 9 features, including the label.

Table 1. Features for Instagram dataset

Feature	Type	Description	Fake Account Indicator
userHasProfilPic	Binary	Whether the user has a profile picture (1=Yes, 0=No)	Fake accounts often lack profile pics or use defaults.
userIsPrivate	Binary	Whether the account is private (1=Private, 0=Public)	Fake accounts tend to avoid scrutiny by staying private.
usernameDigitCount	Numerical	Number of digits in the username	High digit counts suggest bot usernames.
usernameLength	Numerical	Length of the username (characters)	Extremely short or extremely long usernames are suspicious.
userFollowerCount	Numerical	Total followers	Fake accounts often have low followers.
userFollowingCount	Numerical	Total accounts followed	Fake accounts follow excessively or very few.
userMediaCount	Numerical	Number of posts/uploads	Low media counts suggest inactivity or spam.
userBiographyLength	Numerical	Number of characters in the bio	Fake accounts often have empty/brief bios.
is_fake	Categorical (0/1)	Indicates whether an account is fake (1) or real (0)	Primary label used to classify accounts as fake or real

7. Methodology

The data contained 1,194 Instagram accounts, with 200 being fake and 994 being real. The dataset was then preprocessed, which involved handling missing data, removing irrelevant features, and feature scaling. Synthetic Minority Oversampling Technique (SMOTE)

was applied to address class imbalance. Figure 3 displays the workflow of the model.

Multiple features were engineered from the original features and were selected using Recursive Feature Elimination for SVM, Feature importance ranking for RF, and Mutual Information for MLP.

The dataset was split 80:10:10 in order of training,

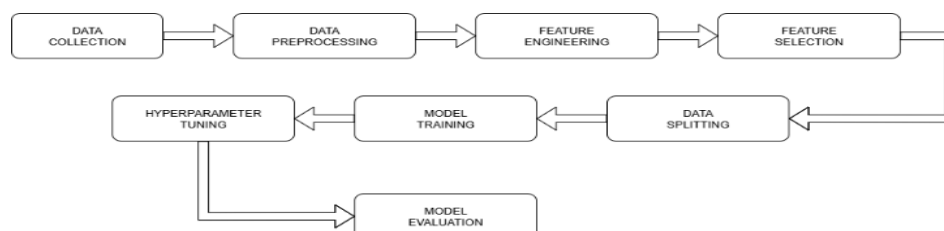


Figure 3. Model workflow

validation, and test, respectively. Hyperparameter tuning involved making use of Grid Search for SVM, and Random Search for RF and MLP. The models were evaluated using metrics Accuracy, Precision, Recall, and F1-score.

8. Performance Metrics

The following performance metrics were used:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \dots\dots (4)$$

$$\text{Precision} = TP / (TP + FP) \dots\dots (5)$$

$$\text{Recall} = TP / (TP + FN) \dots\dots (6)$$

$$\text{F1-score} = 2 * ((\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})) \dots\dots (7)$$

Where:

TP is true positive

TN is true negative

FP is false positive

FN is false negative.

9. Feature Engineering

The following features were engineered from the already existing features in the InstaFake Dataset.

Follower-Following Graph Features

- i. followers_to_following_ratio
- ii. following_to_followers_ratio
- iii. social_engagement_score

Authenticity Features

- i. has_bio
- ii. bio_length_to_media_ratio
- iii. empty_profile_score

Username Analysis Features

- i. username_digit_ratio
- ii. username_entropy
- iii. has_suspicious_username

Activity and Media Features

- i. media_to_followers_ratio
- ii. private_with_high_media

10. Results

Figure 4 displays Train Metrics in split70:15:15 SVM and RF show near-perfect precision (0.99) and high F1 (0.98), while MLP lags slightly (F1: 0.95).

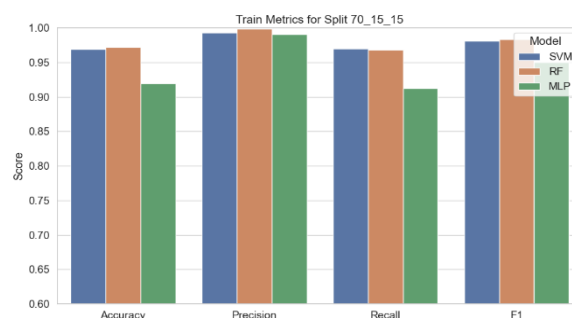


Figure 4. Model Comparison of Train Metrics for Split 70:15:15

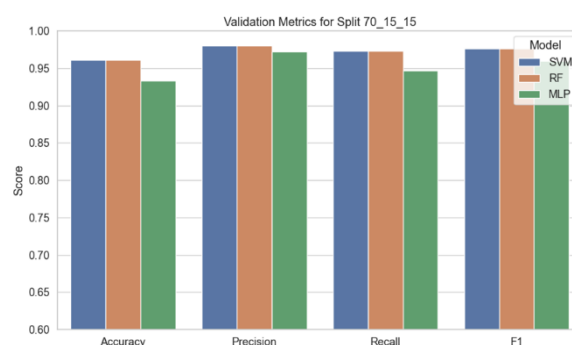


Figure 5. Model Comparison of Validation Metrics for Split 70:15:15

Figure 5 displays Validation Metrics in split 70:15:15. SVM and RF tie for top performance (F1: 0.976), with MLP close behind (F1: 0.959). All models exceed 93% recall.

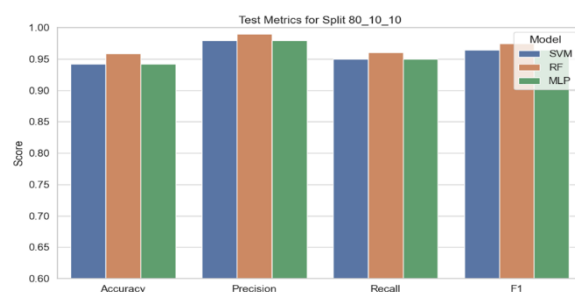


Figure 6. Model Comparison of Test Metrics for Split 70:15:15

Figure 6 displays Test Metrics in split 70:15:15. SVM dominates (F1: 0.99, 100% precision), followed by RF (F1: 0.98). MLP maintains strong recall (0.94).

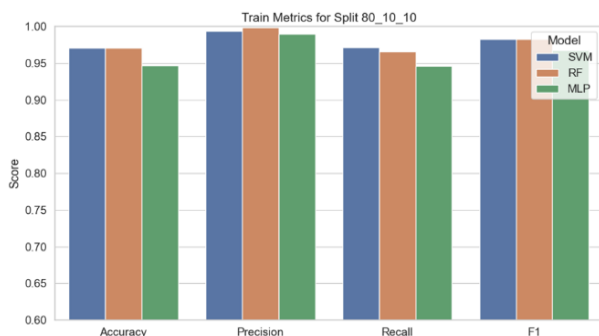


Figure 7. Model Comparison on Train Metrics for Split 80:10:10

Figure 7 displays Train Metrics in split 80:10:10. RF edges out SVM in precision (0.999 vs. 0.994), but both share identical F1 (0.982). MLP trails (F1: 0.967).

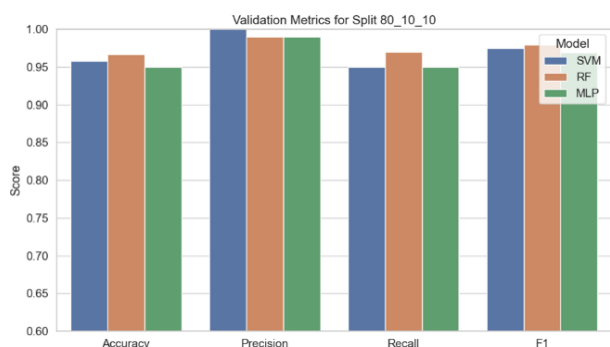


Figure 8. Model Comparison on Validation Metrics for Split 80:10:10

Figure 8 displays Validation Metrics in split 80:10:10. RF leads (F1: 0.98), while SVM and MLP tie (F1: 0.97). SVM achieves 100% precision but lower recall.

Figure 9. Model Comparison on Test Metrics for Split 80:10:10. RF performs best (F1: 0.975), with SVM and MLP closely matched (F1: 0.964). All models exceed 95% recall.

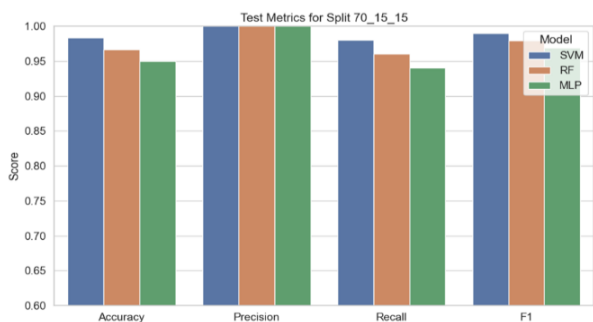


Figure 9. Model Comparison on Test Metrics for Split 80:10:10

Table 2. F1-score of svm model performance

Split	Model	Train F1	Validation F1	Test F1
70:15:15	SVM	0.9811	0.9764	0.9899
80:10:10	SVM	0.9822	0.9741	0.9645

Table 2. Displays the F1-score across both splits using svm.

Table 3. F1-score of rf model performance

Split	Model	Train F1	Validation F1	Test F1
70:15:15	RF	0.9832	0.9764	0.9796
80:10:10	RF	0.9821	0.9796	0.9746

Table 3. Displays the F1-score across both splits using RF.

Table 4. F1-score of mlp model performance

Split	Model	Train F1	Validation F1	Test F1
70:15:15	MLP	0.9498	0.9592	0.9691
80:10:10	MLP	0.9672	0.9691	0.9645

Table 4. Displays the F1-score across both splits using mlp.

Table 5. Model performance across all models

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.942	0.979	0.950	0.964
RF	0.958	0.990	0.960	0.975
MLP	0.942	0.979	0.950	0.964

Table 5. Displays the F1-score across all models. Random Forest achieved the highest F1-Score of 0.975. SVM produced strong generalization while MLP was slightly lower but competitive F1scores.

11. Discussion

Random Forest outperformed other models due to its ability to handle complex feature interactions and class imbalance. SVM benefited from feature selection, while MLP showed efficient recall, critical for minimizing false negatives.

12. Conclusion

Random Forest, demonstrated high accuracy in detecting fake Instagram profiles. Feature selection and data balancing were also key to optimal performance. Future work should integrate behavioral analysis and real-time deployment for enhanced robustness.

13. References

[1] Chakraborty, P., Shazan, M. M., Nahid, M., Ahmed, M. K., and Talukder, P. C. (2022). Fake profile detection using

machine learning techniques. *Journal of Computer and Communications*, 10(10), Article 10.

[2] Patil, D. R., Patterwar, T. M., Punjabi, V. D., and Pardeshi, S. M. (2024). Detecting fake social media profiles using the majority voting approach. *EAI Endorsed Transactions on Scalable Information Systems*, 11(3).

[3] Kumar, P., and Neelam. (2024). Advanced real-time fake social media profiles detection and reporting system. *International Journal of Education and Science Research Review*, 11(3).

[4] Dracewicz, W., and Sepczuk, M. (2024). Detecting fake accounts on social media portals—The X Portal case study. *Electronics*, 13(13), 2542.

[5] Amankeldin, D., Kurmangazyeva, L., Mailybayeva, A., Glazyrina, N., Zhumadillayeva, A., and Karasheva, N. (2023). Deep Neural Network for Detecting Fake Profiles in Social Networks. *Computer Systems Science and Engineering*, 47(1), 1091–1108.

[6] K. Harish, R. Naveen Kumar, and Dr. J. Briso Becky Bell. (2023). Fake Profile Detection Using Machine Learning. *International Journal of Scientific Research in Science, Engineering and Technology*, 10(2), 719–725.

[7] Gaurav V., B. and Dr. Nilesh R. Wankhade. (2023). Social Media Fake Account Identification Using Machine Learning. *International Journal of Advanced Research in Science, Communication and Technology*, 261–268.

[8] Ahmad, S., and Tripathi, Dr. M. M. (2023). A Review Article on Detection of Fake Profile on Social-Media. *International Journal of Innovative Research in Computer Science and Technology*, 11(2), 44–49.

[9] Binder, M. (2024, February 16). The majority of traffic from Elon Musk's X may have been fake during the Super Bowl, report suggests. *Mashable*. <https://mashable.com/article/x-twitter-elon-musk-bots-fake-traffic>. (Access Date: 1 May 2025)

[10] Hanna, K. T., Wright, G., and Yasar, K. (2025, January 22). What is social networking and how does it work? | Definition from TechTarget. <https://www.techtarget.com/whatis/definition/social-networking>. (Access Date: 1 May 2025)

[11] Richard. (2025, January 23). Learn More About Instagram's Game-Changing Features 2025. *Avocado Social*. <https://www.avocadosocial.com/instagrams-game-changing-features-for-2025-what-you-need-to-know/>. (Access Date: 1 May 2025).

[12] Pokrop, J. (2025, July 11). What's New on Instagram in 2025: New Features and Updates. *NapoleonCat*. <https://napoleoncat.com/blog/instagram-new-features-and-updates/>. (Access Date: 1 May 2025).

[13] Social Ninjas. (2025, February 11). What's New on Instagram in 2025: A Social Media Manager's Guide. *UpGrow*. <https://www.upgrow.com/blog/instagram-new-features-and-updates-2025>. (Access Date : 1 May 2025).

[14] Paljug, K. (2025, February 19). Social Media: Definition, Importance, Top Websites, and Apps. *Investopedia*. <https://www.investopedia.com/terms/s/social-media.asp>. (Access Date: 1 May 2025).

[15] Akyon, F. (2022). InstaFake: Fake Instagram profile dataset [Data set]. *GitHub*. <https://github.com/fcakyon/instafake-dataset>. (Access Date: 1 May 2025).

Session 8: Internet Applications and Technology

Title: Impact of E-Commerce Adoption on the Performance of Hotel and Accommodation Businesses in Ondo State, Nigeria

Authors: Bolaji Comfort Adeniyi, Iyabode Abisola Adelugba

Title: A Secure E-Shopping Platform for University Students: Case Study of FUTA

Authors: Afolabi Oladeinde, O.Y Ogunlola, Olumide Sosanya, M.T. Kinga

Title: IoT-Based Energy Consumption Monitoring System

Authors: Olaniyi A. Ayeni, Aryee A. Feyisetan, Gbenga S. Oluokun

Impact of E-Commerce Adoption on the Performance of Hotel and Accommodation Businesses in Ondo State, Nigeria

Bolaji Comfort Adeniyi, Iyabode Abisola Adelugba
*Department of Entrepreneurial Studies,
Bamidele Olumilua University of Education, Science and Technology
Nigeria*

Abstract

The purpose of the study is to examine the impact of e-commerce on the performance of hotel and accommodation businesses in Ondo State, Nigeria. Specifically, the study: investigated the level of e-commerce adoption among hotel and accommodation businesses; assessed the relationship between e-commerce tools and customer satisfaction of hotel and accommodation businesses and investigated the effect of e-commerce on the financial performance of hotels and accommodation businesses in Ondo State, Nigeria. The descriptive research design of the survey type was adopted in the study, the population of the study comprises all the managers of registered hotel and accommodation businesses in Ondo State, Nigeria which is five hundred and forty (540) according to The Ministry of Culture and Tourism, Ondo State (2024). One hundred and eighty eight (188) respondents were sampled for the study using the stratified random sampling technique, a self-designed questionnaire comprising closed-ended Likert-scale questions was employed for the data gathering. A pilot test with Cronbach's Alpha reliability method involving 30 respondents in neighboring Osun State was carried out to check the level of internal consistency of the instrument. Data collected were analyzed using The Statistical Package for Social Sciences (SPSS) version 26. Descriptive statistics was used to summarize the demographic data while Pearson Correlation was used to examine the relationship between e-commerce tools and customer satisfaction; regression analysis was used to determine the effect of e-commerce adoption on hotel financial performance. All hypotheses were tested at 0.05 level of significance. The findings of the study revealed that the level of e-commerce adoption among hotel and accommodation businesses in Ondo State, Nigeria is moderate; e-commerce tools significantly enhance customer satisfaction and its adoption has a strong positive effect on financial performance of the hotel and accommodation industry in Ondo State, Nigeria. Based on the findings, it was recommended that hotel managers should leverage e-commerce platforms and train their staff to adopt e-commerce system.

A Secure E-Shopping Platform for University Students: Case Study of FUTA

Afolabi Oladeinde, O.Y Ogunlola, Olumide Sosanya, M.T Kinga
Federal University of Technology Akure, Nigeria

Abstract

This study uses the Federal University of Technology, Akure (FUTA) as a case study to demonstrate the design and implementation of an e-commerce system specifically suited to students' buying needs. Because of the way academic schedules are structured and the locations of shopping malls, students in university towns usually face challenges related to traditional shopping practices, such as limited access to goods, time limits, long lines, and the unavailability of needed items. These drawbacks emphasize the necessity of a workable and user focused digital solution. The project employs the Structured Systems Analysis and Design Methodology (SSADM) to systematically guide the process of analyzing requirements, designing architecture, and implementing a working prototype of the system. Therefore, the main goal of this project was to create an online purchasing platform that is accessible, safe, and responsive while precisely meeting the needs and preferences of students.

Keywords: E-commerce system. Design Implementation. Structured Systems Analysis and Design Methodology. E-shopping Platform

1. Introduction

In the world of technology and commerce, electronic shopping, or e-shopping, has become an integral part of modern consumer behavior [1]. This paradigm shift extends to various demographics, including students who are a significant consumer segment with distinct needs and preferences. Students, often immersed in academic responsibilities constrained by time and budgetary considerations, find themselves seeking efficient ways to fulfill their purchasing requirements. Traditional brick-and mortar retail experiences may prove to be inconvenient and time-consuming [2]. Recognizing this, the concept of an e-shopping system tailored specifically for students emerges as a pertinent area for exploration and development.

This study aims to delve into the intricacies of crafting an e-shopping system that caters

specifically to the diverse and dynamic needs of the student demographic. The implementation aspect of the study involves the development of a prototype system, incorporating features that prioritize ease of use, affordability, and accessibility for students [3]. This research endeavors to bridge the gap between technological innovation and the practical requirements of students, ultimately contributing to the advancement of user-centric e-shopping solutions.

2. Related Work

E-shopping has rapidly evolved, significantly transforming how businesses and customers interact. Its convenience and impact on sales growth have made online shopping increasingly popular [4]. [5] Developed an online shop for St. John's Collection in Calabar, highlighting improved accessibility and alignment with cashless economy goals. [6] Created a platform for electronic goods using open-source tools, though handling large transaction volumes posed challenges. [7] Focused on university students, using modern web tools to enhance the online shopping experience but lacked analysis of user satisfaction. [8] built a secure B2C platform using Django and JavaScript, yet overlooked marketing and scalability. [9] emphasized technical scalability via Spring Cloud but didn't address user experience or strategy. [10] used Django to develop a scalable e-commerce platform but missed logistical and security concerns. [11] implemented a SpringBoot-based system improving operations, though scalability and security were underexplored. [12] optimized a cooperative store's operations but noted issues like glitches and maintenance

3. Methodology

The system was developed using the Structured Systems Analysis and Design Methodology (SSADM), following a waterfall model to ensure a structured approach to analysis, design, and implementation. The process involved feasibility

studies, environmental analysis, requirement gathering, and technical system design. Users register or log in, browse products, add items to their cart, and proceed to checkout. Payments are securely processed with card details and OTP verification. After successful transactions, users receive order summaries and can provide feedback. This flow ensures a smooth, secure, and efficient e-shopping experience tailored to student needs.



Figure 1. System Architecture

4. Implementation

The e-shopping system is a user-friendly web portal designed to streamline online shopping for students. Developed with HTML, CSS, JavaScript, and React for the frontend and PHP for the backend, the platform offers a smooth and secure experience for browsing, selecting, and purchasing products. The system features several key modules:

- **Account Registration:** Users create an account by providing basic details such as name, email, and password.
- **Login:** Returning users securely access their accounts to continue shopping or track previous orders.
- **Homepage:** Serves as the platform's hub, displaying featured products, promotional content, and navigation links.
- **Search:** A smart search bar allows users to quickly locate products using relevant filters such as a name or category.
- **Product Page:** Offers detailed information about each item, including images, descriptions, prices, and customer reviews.
- **Add to Cart:** Enables users to gather and manage selected items before finalizing purchases.

- **Checkout:** Facilitates the order process by collecting shipping details and confirming selected products.

- **Payment:** Handles secure transactions with multiple payment options, ensuring the protection of sensitive data. **Shipping:** Captures delivery information and supports timely and accurate order fulfilment.

The Feedback from FUTA students indicates strong satisfaction with the system's overall performance. These insights reflect the system's effectiveness and adaptability, ensuring it can continue evolving based on user needs.

Create an account and discover the benefits

First Name

Last Name

Email

Password

☐ I agree to the Google Terms of Service and Privacy Policy

Sign up

Are you already a member?

Figure 2. Account creation page

Log in

Lorem ipsum is simply dummy text of the printing and typesetting industry. Lorem ipsum has been the industry's standard dummy text ever since the 1500s

Email

Password

☐ Keep me signed in

Forgot password?

Facebook

Google

Sign in

Not a member yet? [Sign up](#)

Figure 3. Login page



Figure 4. Product page

5. Conclusion

An e-shopping system for students using FUTA as a case study was developed to solve the problems of time-consuming and limited access that comes with traditional shopping in the university community. The structured system analysis and design methodology (SSADM) was adopted for the analysis and design of this system where the user involvement is required at various stages of development to ensure that the system suits the user's needs. A comprehensive analysis of this system was carried out where the architecture and data flow were critically examined. HTML, CSS, React and Javascript were used for the frontend and PHP was used for the backend implementation. The proposed e-shopping system serves as a platform for ordering various items, including books, bags and academic materials within the FUTA community, specifically catering for the needs of students. The user-friendly system ensures convenience and prompt delivery of desired items within the university premises. It receives positive feedback from university students, reflecting high satisfaction levels in overall experience, product relevance, navigation, and system responsiveness. Valuable suggestions include referral link and loyalty incentives. The feedback highlights the system's effectiveness and potentials for continuous success through user-centric improvements.

6. References

- [1] Ahmed, Y. M., and Masum, M. (2020). *Website design testing for online business*. International Journal of Science and Research, 9(3), 10–15.
- [2] Alzoubi, H. M., Alshurideh, M., Kurdi, B., Alhyasat, K. M. K., and Ghazal, T. M. (2022). *The effect of e-payment and online shopping on sales growth: Evidence from the banking industry*. International Journal of Data and Network Science, 3(1), 34–40.
- [3] Fan, H., Yu, Y., Wang, B., and Ma, J. (2021). *The design and implementation of an e-commerce platform based on SpringCloud*. International Conference on Information Technologies and Electrical Engineering (Article No. 57, pp. 18–21).
- [4] Gefen, D., Karahanna, E., and Straub, D. (2018). *Trust and TAM in online shopping: An integrated model*. MIS Quarterly, 51–90.
- [5] Gupta, V., Singh, A., and Pratap, A. (2023). *E-commerce web-application with Django*. International Journal of Innovative Research in Technology, 10(1), 12–18.
- [6] Islam, M. T. (2022). *Website design testing for online business*. International Journal of Engineering Research & Technology, 9(10), 21–25.
- [7] Kumah, R. (2018). *An e-commerce shopping website*. University of Applied Sciences Information Technology, 7(2), 52–58.
- [8] Rahman, M. A., Khan, M. A., and Islam, M. R. (2022). *Design and implementation of a web-based electronic-commerce system*. European Journal of Engineering and Technology Research, 7(2), 5–8.
- [9] Tawo, B. P., Nkanu, I. U., Ele, S. I., and Joseph, O. (2019). *Design and implementation of online shop: A case of St. John's Collection, Calabar*. Journal of Multidisciplinary Engineering Science Studies, 3(8), 27–30.
- [10] Tenzin, S., Lhamo, T., and Dorji, T. (2022). *Design and development of e-commerce web applications for cooperative stores*. International Research Journal of Engineering and Technology, 9(2), 10–19.
- [11] Wu, H. (2022). *E-commerce middle office management system based on Springboot*. International Journal of Advanced Network, Monitoring and Controls, 7(2), 9–13.

IoT-Based Energy Consumption Monitoring System

Olaniyi A. Ayeni, Aryee A. Feyisetan, Gbenga S. Oluokun
*Federal University of Technology
 Kings University, Ode-Omu, Osun State
 Nigeria*

Abstract

The rise in demand for sustainable energy practices and the inefficiencies in the existing systems of management of energy management have triggered the need to develop an intelligent and real-time energy monitoring application for smart buildings. In this project, the idea is to design and implement an Internet of Things (IoT) based Energy Consumption Monitoring System, which is scalable, cost-effective and environmentally responsive. The system connects humidity, temperature sensors, motion, and smoke sensors to an energy meter, as well as a solar panel, to monitor the usage of appliances, energy efficiency, and safety in a smart building environment. The Python custom-built scripts were applied to collect data, simulating the real-time monitoring to trigger appliances under environmental conditions, including humidity and movement. Communicating to users about energy use, cost and status of devices was done via an in-home display. The simulation proved that the system was effective at identifying high-energy-consuming appliances such as the air conditioner, automation the appliances by setting sensor thresholds (e.g. as an example, turning on air conditioners when the humidity is above 60%), and reducing waste by turning off and on devices as required. The framework can facilitate energy saving as well as decrease carbon footprint, and act as a guideline to intelligent energy systems that can be used in domestic and commercial areas.

Keywords: IoT, Smart Building, Energy Monitoring, Real-time Data, Sustainability, Automation, Energy Efficiency, Renewable Energy

1. Introduction

Smart building technologies are adopted due to the increasing interest in the efficiency of energy consumption and sustainability [1]. In this paper, an Energy Consumption Monitoring System applying IoT is presented to realise energy optimisation in real time by use of a sensor network [2], automation [3], and integration of renewable sources. The system minimises waste and operating costs by actively managing lighting, HVAC, and equipment depending on occupancy and environment-related data. Issues like retrofitting of existing infrastructure [4] or scalability [5] belong to the sphere of possible ways

to overcome the challenge by adopting a design as cost-efficient and modular. Real-time monitoring, automated alerting capabilities, and positive smoke detection tests prove the effectiveness of the system and provide a feasible solution in energy management in smart buildings to achieve sustainability.

2. Related Works

After thoroughly reviewing multiple research papers, the contributions made by various researchers are discussed. Deyslen Mariano-Hernández et al. conducted research motivated by the urgent need to reduce CO₂ emissions from buildings. It aimed to assess the impact of forecasting tools on building energy efficiency. A literature review from Scopus and Web of Science formed the methodology. The contribution lies in the comparative analysis of forecasting techniques. Its limitation is that existing models lack adaptability across diverse buildings and behavioural patterns [1]. Chrysi K. Metallidou et al. were motivated by EU policy shifts toward nearly zero-energy buildings, and the study aimed to develop an IoT-based real-time energy monitoring model. It used literature reviews and sensor system development with cloud storage. The main contribution was a legally compliant smart building model that promotes energy efficiency. A major limitation is its heavy dependence on advanced IoT and secure communication infrastructure [2]. Safae Bourhnane et al. were motivated by the growing need for predictive energy scheduling in smart buildings. Their objective was to apply machine learning—specifically ANNs and genetic algorithms—for load forecasting and task scheduling. The methodology utilised real building data for prediction and control. The key contribution was a practical ML deployment model using Python and LabVIEW platforms. However, accuracy was limited by small and under-processed datasets [3]. Daniel Ibasetta et al. discussed the need for sustainable retrofitting of existing buildings to enhance energy efficiency. It aimed to address IoT deployment challenges in older buildings by optimising sensor placement and energy-saving strategies. The researchers used IoT-connected sensors and W3C WoT guidelines to monitor and control energy use. Their work contributed practical solutions for integrating IoT into retrofitted systems.

However, it faced limitations due to high costs and protocol incompatibilities in legacy systems [4]. This study by Rajalakshmi Selvaraj and Ven Madhav Kuthadi was motivated by poor energy use and recycling practices in smart cities. The study aimed to develop an AI-based energy monitoring and management system. It used forecasting models and eco-design surveillance to assess usage and production. The research contributed valuable insights into the effective recycling of energy in smart buildings. However, scalability and integration issues were not addressed [5].

3. System Analysis and Design

The building incorporated home appliances (e.g., lighting, refrigerators, air conditioners, fans, and washing machines) with IoT sensors and renewable energy sources, such as solar panels.

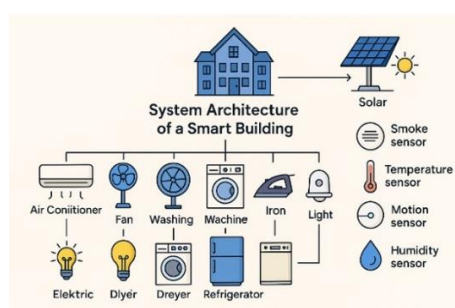


Figure 1. System architecture of a smart building



Figure 2. Diagram of a solar panel of a smart building

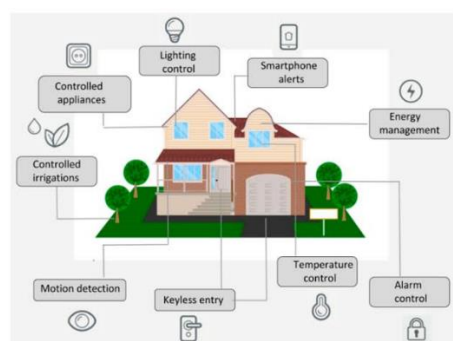


Figure 3. In-home display of a smart building

It guaranteed the monitoring of energy usage and efficiency and aided in the automation of residential and commercial purposes.

3.1. Data Acquisition System (DAQ)

A data collection and digitisation framework was developed to record sensor data. This was a combination of sensors, smart meters, a gateway and data loggers, which transformed analogue signals to digital data streams. Communication procedures provided a smooth passage of data, and time synchronisation was observed to be accurate in real-time recording.

4. Implementation and Result

The system was implemented by means of the incorporation of an IoT-based framework, comprised of temperature, humidity, motion, and smoke sensors, and a solar panel in a simulated smart building setting. A Data Acquisition (DAQ) system was used to gather sensor data, digitise the data, and Python programs were used to allow real-time data monitoring, data recording and device automation. The in-home display gave users a user interface, where energy use, cost, and device status could be shown to occupants, and both usability and transparency were ensured. The use of solar panels as renewable energy sources minimises the use of grid power and boosts sustainability.

4.1. The Smoke Sensor

A part of the monitoring system based on IoT, the sensor transmits the real-time data to the central control unit. The system issues alerts on the in-home display or through connected devices when the level of smoke exceeds a predetermined level, allowing the user to promptly respond or utilise automatic capabilities, such as operating electrical circuits or switching on ventilators.

The sensor under test was placed in a controlled smoke environment. Positive results were obtained, as the system recorded steady and accurate observations within a few seconds of exposure to smoke. This confirms the accuracy of the sensor in early fire detection, making it an important component of smart building safety systems.

Table 1. Real-Time Monitoring Results from a Smoke Sensor

S/N	Appliance	Usage (hrs)	Energy (kwh)	Cost (Rs/hrs)	Smoke Detected
1	Air Conditioner	1.98	6.930	1.732	No
2	Washing Machine	1.06	0.530	1.33	No
3	Light Bulb	1.42	0.852	2.13	No
4	Pressing Iron	0.77	0.924	2.31	No
5	Dishwasher	1.11	1.110	2.78	No
6	Air Conditioner	1.81	6.335	15.84	No
7	Washing Machine	1.07	0.535	1.34	No
8	Light Bulb	0.59	0.354	0.89	No

No smoke was detected across all data points, indicating that the appliances are functioning safely and the smoke detection system is active and effective. This real-time smoke monitoring is essential in a smart building, offering early warnings and integrating with automated shutdown or alert systems to prevent fire risks.

4.2. The Temperature Sensor

Evaluation of energy efficiency and cost implication in the smart building revealed that devices consuming less energy, like light bulbs and dishwashers, consumed minimum energy as the costs were captured at ₦0.00 and probably because of integration of solar power or low tariffs. Compared to air conditioners, evaporators (0.1250 kWh) consumed more energy and were the most expensive (₦0.15), which signifies high demand when indoor temperatures exceed 25°C. Temperature changes were closely associated with appliance activity, with the majority of the activity taking place between 25°C and 31.6°C, indicating a peak afternoon activity. Smart behaviour in buildings also displayed the concurrent activation of devices in a few seconds, either through automation or occupancy through motion sensors, and the solar energy was able to offset costs associated with low-energy devices. The findings highlight how the system can support thermal limits, cost optimisation, and real-time monitoring as one of the major attributes in energy audit and efficiency.

Table 2. Real Time Monitoring Results from a Temperature Sensors

S/N	Appliance	Time Stamp	Temperature (°C)	Energy (kwh)	Cost (₦)
1	Dish Washer	14:13:40	31.6	0.0300	0.00
2	Light Bulb	14:13:40	31.6	0.0010	0.00
3	Light Bulb	14:13:41	24.4	0.0010	0.00
4	Air Conditioner	14:13:42	25.0	0.1250	0.15
5	Light Bulb	14:13:42	25.0	0.0010	0.00
6	Washing Machine	14:13:42	25.0	0.0083	0.00
7	Light Bulb	14:13:43	30.8	0.0010	0.00
8	Air Conditioner	14:13:43	30.8	0.1250	0.15
9	Dish Washer	14:13:43	30.8	0.0300	0.00
10	Light Bulb	14:13:44	31.3	0.0010	0.00

5. Conclusion

The IoT-based system efficiently maximises the energy consumption in intelligent buildings by using real-time monitoring, automation, and renewable energy sources. In spite of the existing issues, such as retrofitting and scalability, there is a need to focus on standardisation and cost effectiveness in the future. The solution is also globally sustainable with practical prospects of adequacy in energy management.

6. References

- [1] Mariano-Hernández, D., Hernández-Callejo, L., & Zorita-Lamadrid, A. (2021). Management of energy consumption in smart buildings: A review. *Sustainable Cities and Society*, 65, 102642. DOI:10.1016/j.scs.2020.102642 (Access Date: 11 March 2025).
- [2] Metallidou, C. K., Psannis, K. E., & Egyptiadou, E. A. (2020). Energy efficiency in smart buildings: IoT approaches. *IEEE Access*, 8, 63679-63699. DOI:10.1109/ACCESS.2020.2984461 (Access Date: 11 March 2025).
- [3] Bourhnane, S., Abid, M. R., Lghoul, R., Zine-Dine, K., Elkamoun, N., & Benhaddou, D. (2021). Machine learning for energy consumption prediction and scheduling in smart buildings. *Neural Computing and Applications*, 33(3), 985-1002. DIO:10.1007/s00521-020-05035-x (Access Date: 11 March 2025).
- [4] Ibaset, D., Álvarez, M., & Molleda, J. (2022). IoT retrofitting for energy efficiency in existing buildings. *Building and Environment*, 207(B), 108523. DOI:10.1016/j.buildenv.2021.108523 (Access Date: 23 March 2025).
- [5] Selvaraj, R., & Kuthadi, V. M. (2021). Scalable energy management framework for smart buildings. *Journal of Building Engineering*, 39, 102265. DOI: 10.1016/j.job.2021.102265 (Access Date: 23 March 2025).
- [6] Renke. (2025). Types of smart building sensors with IoT technology. <https://www.renkeer.com/smart-building-sensors-types/> (Access Date: 16 June 2025).
- [7] Habib U. (2025). https://www.researchgate.net/figure/Schematic-diagram-of-a-typical-Solar-Home-System_fig1 (Access Date: 16 June 2025).
- [8] Stolojescu-Crisan, C., Crisan, C., & Butunoi, B.-P. (2021). An IoT-Based Smart Home Automation System. *Sensors*, 21(11), 3784. DOI: 10.3390/s21113784 (Access Date: 16 June 2025).

Session 9: e-Society

Title: Breast Cancer Detection in Mammography Using a Stacked Ensemble of Three Pretrained Models
Authors: A.D. Adegoke, A.O. Ogunde, B.O. Olorunfemi

Title: Development of a Smart Based Medication, Dispenser and Prescription System
Authors: Ololade Oladimeji, Enesi Femi Aminu, Ayobami Ekundayo

Title: A Machine Learning Framework for Hemoglobin-Based Severity Classification in Sickle Cell Disease
Names: Oluwole Olajide, Aderonke Sakpere, Adesesan Adeyemo, Godwin Ogbole, Toluwase A. Olowookere, Samson Arekete, Benjamin Aribisala

Breast Cancer Detection in Mammography Using a Stacked Ensemble of Three Pretrained Models

A. D. Adegoke, A.O. Ogunde, B.O. Olorunfemi
Department of Computer Science, Redeemers University
Nigeria

Abstract

Breast cancer remains one of the top global health problems, particularly in developing nations where death from breast cancer is still high and chances for early detection remain low. Although mammography represents the most reliable method to screen breast cancer, in dense breasts its sensitivity is significantly affected and therefore it represents a high risk for false-negative results. To avoid these limitations, the recent study proposes a system based on deep learning to classify mammography images as benign or malignant. The CBIS-DDSM dataset was utilized for this study with rigorous preprocessing techniques, including normalization and denoising, employed to improve the quality of images. In addition, a significant data augmentation was applied to normalize class distribution and obtain improved generalization. The three most recent pre-trained convolutional neural networks, InceptionV3, ResNet50, and VGG16, were fine-tuned and used as feature extractors. Their attributes were combined and added using a stacking ensemble meta-classifier to generate the final predictions. The testing revealed that all the individual models performed well with InceptionV3 at 93.03%, VGG16 at 95.09%, and ResNet50 being the best with a 97.99%. The stacked ensemble also gave better results than all the three individual models with a stunning 98.31% accuracy, precision, and recall and loss fell to 0.0763. These results confirm the utility of ensemble approaches and transfer learning for mammography analysis and demonstrate that stacking produces improved diagnostic performance compared to individual models. The model was built in Python with TensorFlow, Keras, and Scikit-learn to facilitate reproducibility and flexibility. Future work involves incorporating additional imaging modalities, using larger multi-institutional datasets, and external clinical validation to enhance robustness, scalability, and practical utility.

Keywords: Breast cancer, Deep Learning, Mammography, Transfer Learning, Stacking Ensemble.

1. Introduction

Breast cancer remains one of the leading health problems globally, both due to its prevalence and due

to its heterogeneous biological and clinical presentation [1]. It is the most frequent female neoplasm and one of the leading causes of death due to cancer, with millions of new diagnoses every year in the developed and developing world [2]. While significant strides have been made in treatment modalities, ranging from surgery and chemotherapy to targeted therapies and immunotherapy, the most decisive factor in reducing mortality remains early and accurate detection. Screening mammography has long served as the gold standard for identifying early-stage breast cancer, enabling interventions that improve survival outcomes [3]. Yet, mammography interpretation is not free of constraints since the radiologists encounter image quality-related challenges, variability in breast density, and scarce lesion features. Such limitations have spurred the pursuit of novel computational methods that can tackle improving diagnostic accuracy, minimizing mistakes, and enhancing clinical choice [4].

Machine learning has emerged as a potent friend to medical imaging in recent years with computational models that can learn from experience and detect patterns too subtle for humans to discern. The earlier algorithms based on machine learning in mammography typically relied on hand-crafted features such as texture, shape, or statistical features based on the region of interest [5]. These features were then fed into classifiers such as support vector machines (SVMs), random forests, or k-nearest neighbors to discriminate between malignant and benign results. While these approaches achieved modest success, their performance was highly dependent on the quality and relevance of the handcrafted features, which often failed to capture the complex, hierarchical structures inherent in medical images. As a result, the progression from traditional machine learning to deep learning has been particularly transformative in breast cancer detection [6].

Deep learning, particularly through convolutional neural networks (CNNs), revolutionized image analysis by automating feature extraction. CNNs are capable of learning spatial hierarchies of features directly from raw imaging data, ranging from low-level edge detectors in initial layers to highly abstract representations of tumor morphology in deeper layers. This ability has positioned CNNs as a cornerstone in

the development of computer-aided diagnostic systems for mammography [7]. Numerous studies have demonstrated their capacity to outperform traditional methods and, in some cases, match or even surpass the diagnostic accuracy of expert radiologists. However, deep learning models are not without challenges. Supervised training of CNNs from scratch also relies heavily on vast quantities of labeled data, which in medical images are typically scarce due to privacy issues, labeling expense, and lack of rare pathological instances [8]. Furthermore, conventional single-model CNN architectures suffer from overfitting, limited generalizability, and varying performance across multiple datasets and imaging setups [9].

To address these challenges, ensemble learning has become an increasingly desirable strategy. Ensemble learning involves combining many models to create a more stable and consistent predictor than any individual model alone. The rationale lies in the principle that while individual models will err, when their decision is combined in an appropriately aggregated manner, variance can be reduced, bias reduced, and robustness increased [10]. For mammography, specifically, ensemble methods are particularly promising because of the heterogeneity of breast lesions, which in appearance vary greatly from patient to patient and population to population. Having the capability to process multiple different architectures, ensemble systems can learn complementary feature representations, leading to more comprehensive analysis and improved classification performance [11]. Techniques such as bagging, boosting, and stacking have been extensively studied in computer vision but stacking where several base learners train to a meta-learner has shown to be most effective for more complex medical imaging tasks [12].

In this paradigm, pretrained models are arguably the most important achievement. Pretraining is to train a model on a large generic dataset such as ImageNet comprising millions of natural images and then fine-tune it for a specific medical imaging task. It leverages the transferable knowledge obtained from general image features with an enormous reduction in computational expenses and data requirements of training from scratch. Such models as VGG16, ResNet50, and InceptionV3 have found extremely broad uses in medical imaging because of one simple reason that they have been proven to work well in transfer learning contexts.

Fine-tuning is tasked with rescaling pre-trained models for deployment on breast cancer detection tasks. Rather than freezing all pre-trained layers and utilizing the network as a completely fixed feature extractor, fine-tuning selectively unfreezes deep layers and re-trains them on mammography data. This allows the model to retain general feature representations from early pretraining as well as learn

domain-specific characteristics of mammographic lesions [13]. While ImageNet-trained filters can recognize edges and textures in every image, fine-tuning enables adaptation to the particular characteristics of breast tissue, microcalcifications, and tumor margins. Fine-tuning also helps reduce problems of domain shift, where differences between pretraining data and target medical data might otherwise detract from performance [14]. In practice, fine-tuning improves sensitivity and specificity and is therefore a foundation of current transfer learning practices in breast cancer diagnostics.

The use of these pre-trained CNNs in a stacking ensemble framework is the next logical step in such developments. With stacked networks of multiple fine-tuned pre-trained networks for mammography data and combined with a meta-classifier, researchers can produce a system which not only enhances raw accuracy but also stability and generalizability. The meta-learner can then aggregate the different feature representations and prediction signals of each base model in a manner that makes the ensemble perform better than all the individual base models. Stacking also reduces the reliance on one architecture and thus avoids risks of overfitting or dataset bias. This method is within clinical objectives, where strength and reliability are of utmost importance, as false negatives will stand up to treatment while false positives lead to undue anxiety and interventions [15].

In addition to technical performance, ensemble Apart from technical performance, ensemble learning with pre-trained models and fine-tuning also meets basic challenges for clinical deployment. Clinicians and radiologists need to have diagnostic tools which will perform consistently across heterogeneous populations and imaging conditions. Individual models tend to show variability in performance when used beyond their training domain, but ensembles enhance generalizability by drawing on the complementary strengths. In addition, pretrained ensembles reduce the computational and data burdens involved in training completely new models, making them more deployable in clinical processes, especially in low- and middle-income nations where resources are limited [16]. By leveraging existing, well-documented architectures, scientists can accelerate the process of moving AI-based solutions from the research community to the practice environment.

The value of developing robust AI systems to screen for breast cancer lies beyond technical advancement [17]. Globally, disparities in access to expert radiologists and advanced imaging machines remain a gigantic obstacle to early detection and effective treatment [18]. AI diagnostic systems, particularly those employing ensemble learning with pretrained models, have the potential to democratize access to high-quality diagnosis tools. By assisting radiologists in image interpretation, reducing the incidence of diagnostic error, and enhancing triaging

of patients, the systems can assist in reducing the global burden of breast cancer. Furthermore, their use in clinical practice must be facilitated by strong validation, ethical conduct, and conformity to healthcare policies in order to ensure responsible and fair deployment [19].

In this context of challenge and opportunity, this present study aims at the development of breast cancer detection using a stacked ensemble model that integrates three previously trained CNN models InceptionV3, ResNet50, and VGG16. The fine-tuning of the models with the use of application ensures they are fine-tuned to the specific characteristics of mammography data, while stacking harnesses their complementary strength in enhancing classification efficiency and robustness [20]. This study is an extension of the growing evidence for ensembles and transfer learning being better in medical imaging and seeks to bridge the remaining gaps in terms of generalizability, interpretability, and clinical usability. The system presented here is one step towards the eventual goal of enhancing early breast cancer detection and patient outcomes through the use of artificial intelligence responsibly.

2. Literature review

Breast cancer remains one of the most prevalent malignancies affecting women worldwide and is a leading cause of mortality despite significant advances in medical imaging and therapeutic interventions. According to the World Health Organization, millions of new breast cancer cases are diagnosed annually, with substantial variations in survival outcomes depending on the stage of diagnosis and access to healthcare resources [1]. The introduction of mammography as a screening tool has played a pivotal role in reducing mortality, especially in high-income countries with structured and accessible screening programs. However, mammography continues to face critical challenges, including reduced sensitivity in women with dense breast tissue, high false-positive rates, and inter-observer variability among radiologists, which collectively reduce its reliability as a standalone diagnostic modality [21]. These limitations have driven interest in more advanced computer-aided diagnostic systems that integrate artificial intelligence to enhance both accuracy and clinical applicability.

Deep learning has emerged as one of the most transformative technologies in medical image analysis, particularly due to the success of convolutional neural networks (CNNs). Unlike traditional machine learning approaches, CNNs are capable of learning hierarchical feature representations directly from imaging data, eliminating the need for handcrafted feature extraction while improving accuracy in classification and detection tasks [22]. Initial CNN implementations

to mammography images revealed promising performance improvements compared to standard radiological interpretation, particularly with the combination of transfer learning techniques. Transfer learning facilitated the utilization of pre-trained models such as VGG16, ResNet50, and DenseNet121 being diverted into medical image application use, hence lessening the constraint of having small labeled data as well as computational costs for training from the beginning [23]. However, despite the single-model CNN approaches showing improved performance, they were still subject to overfitting, dataset reliance, and variability in the diagnostic outcomes when tested on datasets such as CBIS-DDSM, INbreast, and MIAS [24].

To avoid these limitations, ensemble learning has been in great demand in medical images, particularly breast cancer detection. Ensemble methods leverage the strength of many models in order to reduce variance, enhance robustness, and enhance predictability over individual classifiers. Numerous reports have evidenced great performance with ensemble methods in mammography analysis. The study in [25] suggested a fuzzy ensemble approach that integrated InceptionV4, ResNet50, and DenseNet121 to achieve an extremely high 99.32% classification accuracy in breast cancer. The approach was focused on leveraging the power of ensembles in reducing false positives and negatives and enhancing generalizability to datasets. Also, in [26] EfficientNet-b0 was employed with feature fusion and data augmentation methods and achieved 95.4% accuracy on the CBIS-DDSM database and 99.7% on the INbreast database. These results demonstrate the merit of applying feature fusion and data augmentation with ensemble learning in achieving the best diagnostic performance. In [27], a consensus-adaptive weighted ensemble of EfficientNet, Xception, MobileNetV2, InceptionV3, and ResNet50 was developed. With dynamic model weight updating according to performance, their ensemble achieved a level of 95.48% accuracy, better than the performance of single models and reflective of the adaptive ensemble's potential towards more accurate simulation of decision-making process by clinicians.

Stacking ensembles, in which the base learners are stacked along with the meta-learners, have also proved to be highly promising in the analysis of mammography. The study in [28] stacked VGG16, InceptionV3, and VGG19 with logistic regression and neural network meta-learners and achieved a classification accuracy of over 98%. This finding proves that stacked ensembles perform better than normal bagging or boosting ensembles by making use of meta-learning in resolving inconsistency in individual base models. Concurrently, in [29], a fuzzy ensemble method combining VGG-11, InceptionV3, and ResNet50 achieved a classification accuracy of 98.99%. The intelligent ranking process employed

guaranteed ranking of the models according to prediction reliability, hence optimizing the diagnostic accuracy. These results prove the efficiency of stacking-based and fuzzy ensemble techniques in compensating for the limitations of each individual CNN model, particularly overfitting and reliance on the dataset.

Alongside conventional CNN ensembles, multimodal and hybrid solutions have emerged as new breast cancer detection solutions. In [30], the study presented a multi-model framework incorporating mammography image features and clinical metadata in a stacking-based ensemble with the top AUC of 0.94. This uncovered the promise of fusing heterogeneous information sources during risk stratification, a critical step towards translation of research findings into the clinic. In a similar vein, [31] created a patient risk factors-mammographic features combined hybrid deep learning model for breast cancer risk estimation that had an AUC of 0.70. As high as the performance of the model was, its generalizability was being compromised from having been trained on data collected in a single institution, a common failing of much previous AI-based research into detecting breast cancer. Recently, in [32] proposed a multimodal deep learning model for breast cancer subtype classification from mammography images and clinical data to classify benign, luminal A, luminal B, HER2-enriched, and triple-negative subtypes. Their method achieved 88.87% AUC, greatly superior to unimodal methods and demonstrating the capability of multimodal methods in precision oncology.

Ensembles based on CNN have also been used in other imaging modalities aside from mammography, such as cytology. The study in [33] demonstrated that an ensemble of GoogleNet, VGGNet, ResNet, and DenseNet improved cytology-based breast cancer prediction accuracy to 86%, which was better than single CNNs. Similarly, in [34] VGG16, DenseNet121, and InceptionV3 were ensemble for mammogram classification, with more than 90% accuracy on INbreast and CBIS-DDSM, significantly reducing the number of false negatives, a particularly critical factor in screening use cases. These findings demonstrate that ensemble learning is not merely robust across mammography databases but also transferable between imaging modalities, thereby opening up more opportunities for its deployment in clinical settings.

The introduction of transformer-based models has further advanced breast cancer detection in mammography by addressing some of the limitations of CNNs. Unlike CNNs, which find it challenging to capture long-range dependencies in the image, Vision Transformers (ViTs) and their variants can easily model global context. According to [35] ViT-Mammo achieved an AUC of 0.961 in mammography classification, outperforming several CNN-based

models. Similarly, the study in [36] introduced the MammoFormer framework, which combined transformer-based models with multi-feature boosting and explainable AI (XAI) capabilities. Their study tested CNNs, Vision Transformer, Swin Transformer, and ConvNext with various feature augmentation methods, achieving 98.3% accuracy through ViT with adaptive histogram equalization and a 13% improvement with Swin Transformer via histogram of oriented gradients. Notably, the MammoFormer system was able to overcome clinical adoption hurdles by incorporating explainable AI mechanisms that produced multi-perspective diagnostic interpretability. This ensured not only high accuracy of classification but also more trustworthiness and transparency to clinical users.

The need for explainability in AI-driven mammography systems has been emphasized in several recent works. The study in [37] proposed an explainable AI-driven hybrid deep learning model that integrates DenseNet121, Xception, and VGG16, enhanced with GradCAM++ to enable interpretability. Their model provided a classification accuracy of 97%, a 13% improvement over individual models. By combining interpretability functionality, the system allowed clinicians to view lesion boundaries and understand model predictions and thereby overcome one of the greatest barriers to clinical uptake of deep learning systems: their black-box nature. This kind of result suggests that the inclusion of explainability in ensemble and hybrid models makes these models more clinically relevant and hence more acceptable for real practice.

Multimodal integration of imaging modalities Multimodal fusion of imaging modalities has also been shown to improve the results of breast cancer detection. The study in [38] developed a multi-modal AI system that combined full-field digital mammography, synthetic mammography, and digital breast tomosynthesis (DBT) to produce both breast-level predictions and bounding-box localizations of suspicious findings. Trained on approximately 500,000 mammography exams, their model achieved an AUROC of 0.945 on internal validation and significantly reduced radiologist workload by 43.8% without compromising 100% sensitivity. Prospective deployment at 18 clinical sites also demonstrated its generalizability, with reduction of false-positive recall rates by 31.7%. These findings highlight the clinical impact of integrating multiple imaging modalities to enhance diagnostic accuracy, streamline workflow, and reduce radiologist burden. According to [39], a tailored EfficientNetV2 model with enhanced attention mechanisms was validated across Thai hospitals, achieving AUROCs of 0.89 to 0.96 and demonstrating good agreement with radiologists.

Another major area of improvement for AI-based breast cancer diagnosis is data preprocessing and augmentation. A number of studies have shown that

preprocessing steps such as lesion segmentation, noise removal, and contrast enhancement improve feature representation significantly and boost classification accuracy [40, 41]. Augmentation methods such as flipping, rotation, and synthetic data generation have also been seen to remedy dataset imbalances as well as boost generalizability. Studies [26, 25] both established the pivotal role of augmentation in boosting ensemble performance, particularly under small sample datasets. The studies in [42–44] emphasized the significant role of augmentation-driven frameworks in combating overfitting and under-representation of malignant cases.

Despite the advances of CNNs, ensembles, transformers, and multimodal AI approaches, several limitations persist. Many studies rely on relatively small or institution-specific datasets, which limits generalizability to diverse clinical contexts. Imaging protocols vary across hospitals, and lack of standardized validation frameworks often leads to inconsistent results across studies. Additionally, while ensemble and transformer-based architectures often outperform individual CNNs, their high computational requirements can present challenges for clinical deployment, particularly in resource-limited settings. According to [1], AI-driven screening holds potential to reduce diagnostic disparities in low-resource environments, although the computational demands of advanced architectures may limit accessibility and scalability. There also remains a pressing need for multi-institutional datasets and external validation to ensure robustness and clinical trustworthiness. Moreover, interpretability continues to be a concern, despite promising developments in explainable AI frameworks such as GradCAM++ and MammoFormer.

In summary, the reviewed literature demonstrates a clear progression in breast cancer detection research, from early CNN-based single-model approaches to advanced ensemble strategies, transformer-based architectures, multimodal integration, and explainable AI frameworks. Studies consistently demonstrate that ensemble learning improves classification accuracy, reduces false positives and negatives, and enhances model robustness. Transformer-based models and multimodal AI approaches extend these improvements by capturing long-range dependencies and integrating heterogeneous data, thereby advancing diagnostic precision and clinical relevance. Meanwhile, explainability features are increasingly being incorporated to address the black-box challenge of AI and improve clinician trust. However, limitations related to dataset size, generalizability, computational efficiency, and clinical validation remain pressing challenges. Building on these findings, the present study proposes a stacking ensemble framework that integrates three pretrained

CNNs InceptionV3, ResNet50, and VGG16 augmented with extensive preprocessing and fine-tuning. This approach aims to overcome existing challenges of dataset dependency and limited generalizability, while moving toward more reliable and clinically applicable breast cancer detection systems.

3. Methodology

The methodology for this study involved the adoption of three powerful pre-trained convolutional neural networks, namely ResNet50, VGG16, and InceptionV3, which were employed for the classification of mammography images into benign and malignant categories. These models were initially trained on the ImageNet dataset, a large-scale collection of everyday images, which enabled them to learn general visual features such as edges, textures, and shapes. By leveraging transfer learning, the pre-trained weights of these models provided a strong foundation, reducing the computational cost and time associated with training while ensuring high performance on medical imaging tasks. Convolutional neural networks (CNNs) inherently learn hierarchical feature representations through convolutional, pooling, and activation layers, and pre-training allows these features to be transferred and adapted for more specialized applications such as breast cancer detection. Figure 1 illustrates the overall methodology of the study, showing the adoption of pre-trained CNN architectures, the transfer learning process, and their application in classifying mammography images into benign and malignant categories.

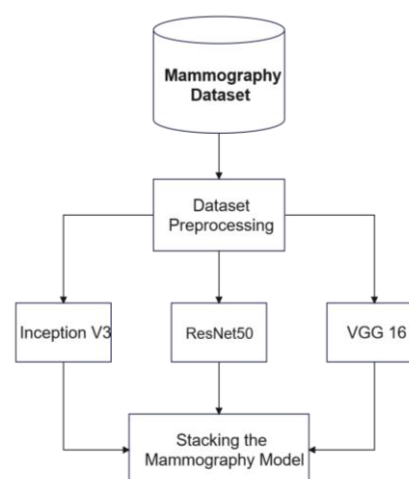


Figure 1. Methodology for breast cancer classification using pre-trained CNN models

3.1. ResNet50

ResNet50 is a 50-layer deep CNN architecture that utilizes convolutional, batch normalization, activation, and pooling layers arranged in a chain-like

structure. Its most significant contribution lies in the introduction of residual blocks, which include identity shortcut connections that allow the input to bypass one or more layers before being added to the transformed output. This design addresses the vanishing gradient problem, enabling the training of very deep networks while maintaining accuracy and efficiency. As a result, ResNet50 can effectively capture fine-grained hierarchical features, making it particularly well-suited for analyzing complex medical imaging data. Figure 2 depicts the architecture of ResNet50, illustrating the sequence of convolutional layers, residual blocks with shortcut connections, and the final fully connected layers used for classification.

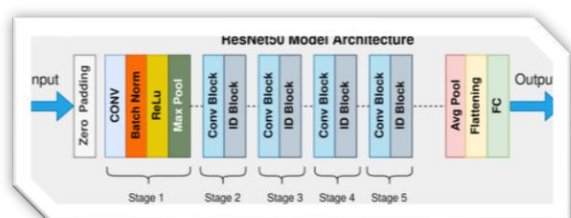


Figure 2. ResNet50 architecture [45]

3.2. VGG16

VGG16, meanwhile, is characterized by its simple yet effective architecture. It employs consecutive 3×3 convolutional filters with stride 1, stacking them to cumulatively increase the receptive field without expending too much computation. It consists of 13 convolutional and three fully connected layers, with max pooling layers added after every few convolutional blocks to downsample. The subsequent dense layers at the end comprise 4096 neurons each, terminating in a softmax classifier for prediction. VGG16, despite its extravagant memory and computation requirements, has demonstrated excellent feature extraction capabilities due to its depth and consistent architecture that enables it to pick up on intricate details in medical images. Figure

3 illustrates the VGG16 architecture with the stacked 3×3 convolutional layers, max pooling, fully connected layers, and the final softmax classification layer.

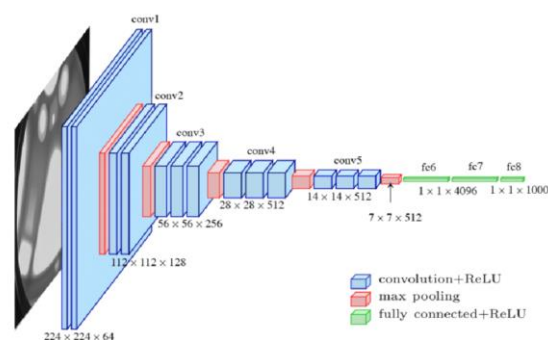


Figure 3. VGG16 architecture [46]

3.3. InceptionV3

InceptionV3, the third model employed by this study, incorporates inception modules that carry out multiple convolution operations of filter sizes 1×1 , 3×3 , and 5×5 in parallel to one another, concatenating their output to form a dense feature representation. This allows the model to extract fine as well as coarse details in an image simultaneously, thereby increasing the capacity of the model to scan medical images of varied spatial resolutions. To achieve maximal computational efficiency, InceptionV3 employs 1×1 convolutions for early down-sampling within the network, before performing bigger convolutions, thus reducing the number of parameters without compromising on performance. It has 48 layers and employs dropout, batch normalization, and auxiliary classifiers to reduce overfitting and enhance generalization. These features make it computationally efficient and viable for mammography analysis. Figure 4 shows InceptionV3 architecture, which denotes inception modules with multi-scale convolutions, dimension reduction by 1×1 convolutions, auxiliary classifiers, and the classification phase.

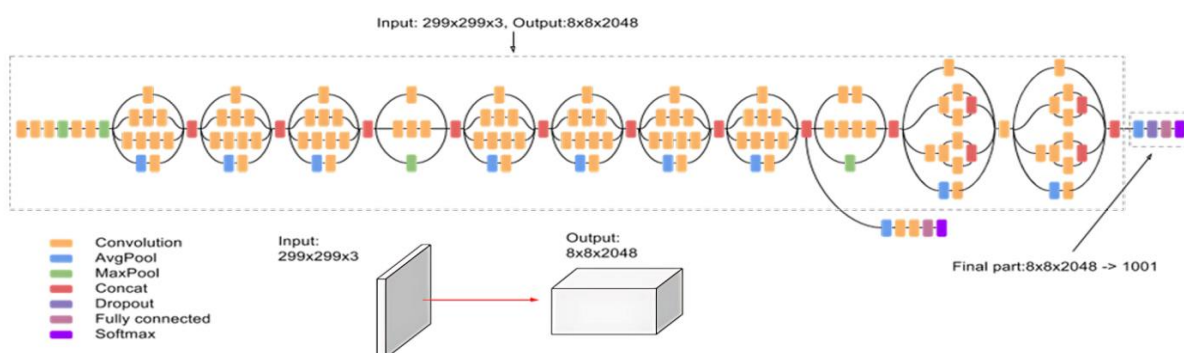


Figure 4. InceptionV3 Architecture [47]

3.4. Dataset collection and preprocessing

The dataset used for this research was the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM). It comprised 2,620 mammography studies, including 1,018 benign and 570 malignant cases prior to augmentation. Like many medical datasets, CBIS-DDSM suffered from class imbalance, which can bias model performance. To address this, several preprocessing steps were applied, including resizing all images to 224×224 pixels for consistency across the models, normalizing pixel intensity values to ensure stable training, and removing non-relevant artifacts such as labels and scanning noise. Augmentation techniques such as horizontal and vertical flipping, random rotations, zooming, brightness and contrast adjustments, and slight translations were employed to increase dataset diversity and mimic real-world variability in mammography images. These measures not only improved generalization but also balanced the dataset. Following augmentation, both benign and malignant classes contained 4,194 images each, resulting in a total of 8,388 images for training and validation.

3.5. Individual model training and fine-tuning

Model training was conducted through fine-tuning, where selected layers of the pre-trained networks were unfrozen and retrained on the mammography dataset. Fine-tuning is a critical step in transfer learning because while the early layers of CNNs capture general low-level features transferable across domains, the deeper layers learn domain-specific representations. Since ImageNet-trained models are not naturally adapted to medical imaging, fine-tuning enables them to capture patterns unique to mammography, such as tissue density, calcifications, and tumor morphology. In this study, different proportions of layers were unfrozen 50%, 60%, and 70%—to systematically evaluate how much domain-specific adaptation enhanced performance without leading to overfitting. This approach allowed the models to retain useful general features while adapting their higher-level representations to medical imaging.

To ensure robustness during training, regularization techniques were employed. Early stopping was implemented to halt training once the validation loss plateaued, preventing unnecessary overfitting. Dropout regularization was also applied, randomly deactivating neurons during training to improve generalization. Each fine-tuned model thus evolved into a strong feature extractor, capable of learning domain-specific representations of mammography images while retaining useful general knowledge from ImageNet.

3.6. Stacking ensemble integration

After isolated training of ResNet50, VGG16, and InceptionV3, feature vectors were extracted from all three and stacked together to form a combined representation. This hybrid set of characteristics was subsequently run through a meta-classifier, which had been implemented as a fully connected neural network. The stacking ensemble strategy was employed to combine the strengths of each of the three models: ResNet50 introduced residual learning to enable more hierarchical feature extraction, VGG16 introduced high detail texture and fine structural examination, and InceptionV3 introduced multi-scale feature presentation. By combining these various strengths, the ensemble reduced variance, increased robustness, and achieved higher predictive accuracy than the separate three individual models. For additional stabilization, cross-validation was applied at ensemble training, meta-classifier optimization, and overfitting opportunity minimization. Figure 5 shows the ensemble architecture, showing how ResNet50, VGG16, and InceptionV3 feature vectors are concatenated and fed to a fully connected neural network meta-classifier for prediction.

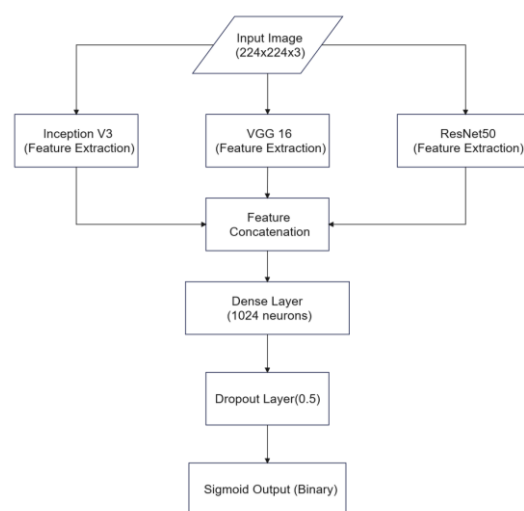


Figure 5. Stacking ensemble workflow

3.7. Evaluation metrics and implementation

The system's performance was evaluated rigorously based on accuracy, precision, recall, F1-score, and loss as the primary measures. Accuracy provided a notion of general correctness, while precision indicated the proportion of correctly predicted malignant cases among all predicted positives. Recall, or sensitivity, measured the model's ability to correctly identify malignant cases, which in breast cancer diagnosis has grave clinical consequences if false negatives occur. The F1-score,

as the harmonic mean of recall and precision, gave a balanced measure, and loss curves were monitored to ascertain model convergence during training.

The entire arrangement was completed in Python, where TensorFlow and Keras were utilized for deep learning development and Scikit-learn was utilized for ensemble incorporation and statistical computations. All experiments were executed on Google Colab with GPU acceleration, which provided the computing power to perform large-scale experimentation with fine-tuning methods, augmentation pipelines, and ensemble optimization. The arrangement enabled efficient training and evaluation while offering a comprehensive evaluation of the proposed system for mammography-based breast cancer detection.

4. Results

The individual models demonstrated strong classification capabilities on the validation set, confirming the effectiveness of pretrained convolutional neural networks in mammography-based breast cancer detection. InceptionV3, with 60% of its layers unfrozen during fine-tuning, achieved an accuracy of 93.03%, with matching precision and recall values, and a validation loss of 0.2783. These results indicate that InceptionV3 successfully captured relevant structural and textural features from the mammography images. However, increasing the unfreezing percentage beyond 60% led to diminishing improvements and in some cases increased overfitting, suggesting that excessive fine-tuning disrupted the balance between general visual features and domain-specific adaptation.

VGG16, fine-tuned with 50% of its layers unfrozen, performed better than InceptionV3, achieving an accuracy of 95.09%, precision of 94.51%, recall of 94.51%, and a validation loss of 0.1905. These metrics highlight that moderate fine-tuning was particularly effective for VGG16, allowing it to adapt its relatively simple yet deep feature extraction pipeline to the mammography dataset. The balanced precision and recall scores further suggest that the model performed consistently across both benign and malignant cases, demonstrating strong generalization capabilities without sacrificing stability.

ResNet50, with 50% unfreezing, was the best performing individual model. It achieved accuracy of 97.99%, precision of 97.35%, recall of 97.35%, and validation loss of 0.1082. The superior performance can be attributed to the depth of the architecture and residual connections that were able to effectively resolve the problem of the vanishing gradient and enabled the model to capture more complex, higher-level features in the mammography images. These findings demonstrate the utility of network design in improving classification performance on medical imaging tasks as shown in Table 1. For each model,

additional increases in the fraction of unfrozen layers beyond the identified optimal points caused validation loss to rise and accuracy to fall. This finding underscores the need for careful fine-tuning: while some unfreezing is necessary to specialize to mammography data, too much unfreezing introduces noise and overfitting that reduces performance overall.

Comparative analysis of the models indicates that while every one of the pretrained models in isolation performed quite okay, the stacking ensemble model performed best overall. When the layer unfreezing was at 50%, the ensemble recorded the highest precision (98.31%), accuracy, and recall along with the lowest validation loss (0.0763). Such results justify the strength of ensemble learning since the stacking method reasonably combined the complementary strengths of InceptionV3, VGG16, and ResNet50. By the fusion of different feature representations, the meta-classifier produced a more stable and more generalizable model.

Table 1. Model Results

Model	Unfreezing (%)	Accuracy	Precision	Recall
Inception V3	50%	91.91%	91.91%	91.91%
Inception V3	60%	93.03%	93.03%	93.03%
Inception V3	70%	92.15%	92.15%	92.15%
VGG16	50%	95.09%	94.51%	94.51%
VGG16	60%	94.01%	94.15%	94.15%
VGG16	70%	85.64%	85.82%	85.82%
ResNet50	50%	97.99%	97.35%	97.35%
ResNet50	60%	95.41%	95.43%	95.43%
ResNet50	70%	95.44%	95.19%	95.19%
Ensemble	50%	98.31%	98.31%	98.31%
Ensemble	60%	97.84%	97.84%	97.84%
Ensemble	70%	95.72%	95.72%	95.72%

The ensemble also showed greater stability and reduced susceptibility to overfitting compared to the individual models. In comparison to individual models, which occasionally experienced slight drops in precision or recall, the ensemble consistently returned high scores in validation and test sets. The rise in performance highlights the manner in which stacking not only enhances predictivity but also reliability, something of paramount significance in clinical use, where false positives or false negatives can prove calamitous.

5. Discussion

The results validate that top single model and ensemble performance is obtained by modest fine-

tuning (50% layer unfreezing) with best balance between model flexibility and overfitting prevention. Stacking utilizes the strength of each of the pre-trained models for improved classification accuracy and overfitting avoidance, without allowing any of the models to dominate the decision-making. This synergy underscores the power of ensemble learning in medical image analysis as it integrates a multitude of feature extraction capacities in one predictive model.

As compared to previous studies on mammography classification, the proposed ensemble technique yields higher accuracy and robustness. Even though a single CNN model such as VGG16, ResNet50, or InceptionV3 has been employed in previous studies with satisfactory performance, their performance is limited by the dataset size, imbalanced classes, or poor fine-tuning processes. On the other hand, the stacking ensemble of this paper witnessed overall improvements in all the evaluation metrics accuracy, precision, recall, F1-score, and loss to demonstrate the effectiveness of transfer learning and multi-model fusion. The findings point to the fact that stacking can outperform typical single-model pipelines, particularly in complex diagnostic tasks such as breast cancer diagnosis.

Besides the quantitative improvements, several practical implications for clinical deployment are drawn. The use of large data augmentation and class-balanced distribution effectively overcomes problems frequently encountered in medical imaging like small data and class imbalance. This enhances the robustness of the models learned across various patient populations. Additionally, the inclusion of early stopping and carefully controlled layer unfreezing methods shields the models from overfitting while maintaining steady generalization performance a critical point of consideration for adoption in real-world clinical processes where unseen cases dominate.

Stacking ensemble approach also offers flexibility and scalability with a model that can be extended to include additional deep learning models or multimodal imaging data in the future. Adding MRI, ultrasound, or histopathology data may also be potentially useful for enhancing diagnostic accuracy, particularly for dense breast patients or mammography indeterminate cases where mammography alone is not sufficient. Moreover, the ensemble model's solid generalization confirms potential application in decision-support systems as a reliable second reader assisting radiologists in detecting subtle malignant features that might go unnoticed through human examination.

Finally, the study also provides insight into the broader implications of integrating complex AI methods into practice. By way of the evidence that well-architected ensemble approaches can outperform individual models, the study contributes to the rising

body of evidence compelling the translation of deep learning solutions to healthcare. The method employed here not only improves predictive reliability but also the framework for scalable, explainable, and clinically relevant diagnostic techniques and, ultimately, earlier diagnosis and improved patient outcomes with the treatment of breast cancer.

6. Conclusion

This study demonstrates how stacking ensembles of pre-trained models underpinned by optimal fine-tuning techniques and high-reliability preprocessing techniques significantly enhance breast cancer detection in mammography images. By leveraging deep learning, transfer learning, and stacking ensemble techniques, the proposed system achieves better diagnosis accuracy than separate systems and thus is a good candidate as a reliable tool for clinical decision support. Utilization of balanced datasets and extensive data augmentation keeps addressing age-old problems in medical imaging, such as class imbalance and lack of diversity of data, to make the models stronger and more generalizable. For clinical clinicians practicing in different clinical settings, the findings present a pathway to more consistent and consistent breast cancer screening, particularly in resource-limited environments where early detection can be translated directly to reduced mortality.

Despite these promising results, several limitations should be acknowledged. Utilisation of a single publicly available dataset can restrict the results' generalizability, and external validation by using data from multiple institutions is needed in order to confirm the robustness of the system across different populations. Moreover, interpretability is still difficult in deep learning approaches, as the "black-box" nature of ensemble models can hamper clinical trust and utilisation on a large scale. Mitigating these shortcomings is essential to successfully move such systems from research tools towards clinical diagnostics.

To peer forward, future studies would have to prioritize highest priority to multi-institutional validation for robustness of the model and explore multimodal imaging data fusion, e.g., ultrasound and MRI, to further enhance diagnostic accuracy. Explainable AI mechanisms would also play a crucial role in enhancing transparency as well as fostering clinical trustworthiness. Of similar importance will be working with radiologists and healthcare professionals to ensure that AI-driven diagnostic systems are compatible with existing workflows and real-world challenges to deployment.

In summary, this research highlights the revolutionizing capability of artificial intelligence-driven ensemble learning systems in the detection of breast cancer. By improving accuracy, robustness,

and responsiveness, the systems hold much potential to revolutionize screening and diagnosis of breast cancer, hence contributing to early detection, improved patients' outcomes, and mortality reduction across the world.

7. References

- [1] Xu, H., and Xu, B. (2023). Breast cancer: Epidemiology, risk factors and screening. *Chinese Journal of Cancer Research*, 35(6), 565–583. DOI:10.21147/j.issn.1000-9604.2023.06.02.
- [2] Łukasiewicz, S., Czezelewski, M., Forma, A., Baj, J., Sitarz, R., and Stanisławek, A. (2021). Breast cancer: Epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—An updated review. *Cancers*, 13(17), 4287. DOI:10.3390/cancers13174287.
- [3] Xiong, X., Zheng, L.-W., Ding, Y., Chen, Y.-F., Cai, Y.-W., Wang, L.-P., Huang, L., Liu, C.-C., Shao, Z.-M., and Yu, K.-D. (2025). Breast cancer: Pathogenesis and treatments. *Signal Transduction and Targeted Therapy*, 10, 49. DOI:10.1038/s41392-024-02108-4.
- [4] Dave, D., Akhunzada, A., Ivković, N., Gyawali, S., Cengiz, K., Ahmed, A., and Al-Shamayleh, A. S. (2025). Diagnostic test accuracy of AI-assisted mammography for breast imaging: A narrative review. *PeerJ Computer Science*, 11, DOI:10.7717/peerj-cs.2476.
- [5] Zheng, D., He, X., and Jing, J. (2023). Overview of artificial intelligence in breast cancer medical imaging. *Journal of Clinical Medicine*, 12(2), 419. DOI:10.3390/jcm12020419.
- [6] Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., and Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics and Proteomics*, 15(1), 41–51. DOI:10.21873/cgp.20063.
- [7] Mienye, I. D., Swart, T. G., Obaido, G., Jordan, M., and Ilono, P. (2025). Deep convolutional neural networks in medical image analysis: A review. *Information*, 16(3), 195. DOI:10.3390/info16030195.
- [8] Rayed, M. E., Islam, S. M. S., Niha, S. I., Jim, J. R., Kabir, M. M., and Mridha, M. F. (2024). Deep learning for medical image segmentation: State-of-the-art advancements and challenges. *Informatics in Medicine Unlocked*, 47, DOI:10.1016/j.imu.2024.101504.
- [9] Mienye, I. D., and Swart, T. G. (2024). A comprehensive review of deep learning: Architectures, recent advances, and applications. *Information*, 15(12), 755. DOI:10.3390/Info15120755
- [10] Dev, V. A., and Eden, M. R. (2019). Gradient boosted decision trees for lithology classification. In S. G. Muñoz, C. D. Laird, and M. J. Realff (Eds.), *Computer Aided Chemical Engineering*, 47, 113–118. Elsevier. DOI:10.1016/B978-0-12-818597-1.50019-9.
- [11] Ciurescu, S., Cerbu, S., Dima, C. N., Borozan, F., Părvănescu, R., Ilaș, D. G., Cîtu, C., Vernic, C., and Sas, I. (2025). AI in 2D mammography: Improving breast cancer screening accuracy. *Medicina (Kaunas, Lithuania)*, 61(5), 809. DOI:10.3390/medicina61050809
- [12] Mohammed, A., and Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University – Computer and Information Sciences*, 35(2), 757–774. DOI:10.1016/j.jksuci.2023.01.014.
- [13] Heikal, A., El-Ghamry, A., Elmougy, S., and Rashad, M. Z. (2024). Fine tuning deep learning models for breast tumor classification. *Scientific Reports*, 14(1), DOI:10.1038/s41598-024-60245-w.
- [14] Hussein, A. A., Valizadeh, M., Amirani, M. C., and Mirbolouk, S. (2025). Breast lesion classification via colorized mammograms and transfer learning in a novel CAD framework. *Scientific Reports*, 15(1), DOI:10.1038/s41598-025-10896-0.
- [15] Haque, R., Khan, M. A., Rahman, H., Khan, S., Siddiqui, M. I. H., Limon, Z. H., Swapno, S. M. M. R., and Appaji, A. (2025). Explainable deep stacking ensemble model for accurate and transparent brain tumor diagnosis. *Computers in Biology and Medicine*, 191, DOI:10.1016/j.compbimed.2025.110166.
- [16] Nagro, S. (2025). A stacked ensemble approach for symptom-based monkeypox diagnosis. *Computers in Biology and Medicine*, 191, DOI:10.1016/j.compbimed.2025.110140.
- [17] AlSamhori, J. F., AlSamhori, A. R. F., Duncan, L. A., Qalajo, A., Alshahwan, H. F., Al-abbadi, M., Al Soudi, M., Zakraoui, R., AlSamhori, A. F., Alryalat, S. A., and Nashwan, A. J. (2024). Artificial intelligence for breast cancer: Implications for diagnosis and management. *Journal of Medicine, Surgery, and Public Health*, 3, DOI:10.1016/j.glmedi.2024.100120.
- [18] Schoeman, R., and Haines, M. (2023). Radiologists' experiences and perceptions regarding the use of teleradiology in South Africa. *SA Journal of Radiology*, 27(1), DOI:10.4102/sajr.v27i1.2647.
- [19] Najjar, R. (2023). Redefining radiology: A review of artificial intelligence integration in medical imaging. *Diagnostics*, 13(17), DOI:10.3390/diagnostics13172760.
- [20] Gupta, N., Kubicek, J., Penhaker, M., and Derawi, M. (2025). A novel diagnostic framework for breast cancer: Combining deep learning with mammogram-DBT feature fusion. *Results in Engineering*, 25, DOI:10.1016/j.rineng.2024.103836.
- [21] Kayikci, S., and Khoshgoftaar, T. M. (2023). Breast cancer prediction using gated attentive multimodal deep learning. *Journal of Big Data*, 10(1), 62. DOI:10.1186/s40537-023-00749-w.
- [22] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep

- learning. *Nature*, 521(7553), 436–444. DOI:10.1038/nature14539.
- [23] Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., and Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. *Scientific Reports*, 9(1). DOI:10.1038/s41598-019-48995-4.
- [24] Anwar, S. M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., and Khan, M. K. (2018). Medical image analysis using convolutional neural networks: A review. *Journal of Medical Systems*, 42(11), 226. DOI:10.1007/s10916-018-1088-1.
- [25] Altameem, A., Mahanty, C., Poonia, R. C., Saudagar, A. K. J., and Kumar, R. (2022). Breast cancer detection in mammography images using deep convolutional neural networks and fuzzy ensemble modeling techniques. *Diagnostics*, 12(8), 1812. DOI:10.3390/diagnostics12081812.
- [26] Jabeen, K., Khan, M. A., Balili, J., Alhaisoni, M., Almujally, N. A., Alrashidi, H., ... Cha, J.-H. (2023). BC2NetRF: Breast cancer classification from mammogram images using enhanced deep learning features and equilibrium-jaya controlled Regula Falsi-based features selection. *Diagnostics*, 13(7), 1–22. DOI:10.3390/diagnostics13071238.
- [27] Dehghan Rouzi, M., Moshiri, B., Khoshnevisan, M., Akhace, M. A., Jaryani, F., Salehi Nasab, S., and Lee, M. (2023). Breast cancer detection with an ensemble of deep learning networks using a consensus-adaptive weighting method. *Journal of Imaging*, 9(11), 247. DOI:10.3390/jimaging9110247.
- [28] Nemade, V., Pathak, S., and Dubey, A. K. (2023). Deep learning-based ensemble model for classification of breast cancer. *Microsystem Technologies*. Advance online publication. DOI:10.1007/s00542-023-05469-y.
- [29] Chakravarthy, S. S., Bharanidharan, N., and Kumar, V. V. (2024). Deep transfer learning with fuzzy ensemble approach for the early detection of breast cancer. *BMC Medical Imaging*, 24, 82. DOI:10.1186/s12880-024-01267-8.
- [30] Miao, L., Li, Z., and Gao, J. (2025). A multi-model machine learning framework for breast cancer risk stratification using clinical and imaging data. *Journal of X-Ray Science and Technology*, 33(2), 360–375. DOI:10.1177/08953996241308175.
- [31] Yala, A., Lehman, C., Schuster, T., Portnoi, T., and Barzilay, R. (2019). A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology*, 292(1), 60–66. DOI:10.1148/radiol.2019182716
- [32] Ben Rabah, C., Sattar, A., Ibrahim, A., and Serag, A. (2025). A multimodal deep learning model for the classification of breast cancer subtypes. *Diagnostics*, 15(8), 995. DOI:10.3390/diagnostics15080995.
- [33] Antony Sami, J. C., and Arumugam, U. (2024). An ensemble deep learning model for the detection and classification of breast cancer. *Middle East Journal of Cancer*, 15(1), 40–51. DOI:10.30476/mejc.2023.97317.1857.
- [34] Khoudfi, Y., El Alami, A., Zaydi, M., Maleh, Y., and Er-Remyly, O. (2024). Early breast cancer detection based on deep learning: An ensemble approach applied to mammograms. *BioMedInformatics*, 4(4), 2338–2373. DOI:10.3390/biomedinformatics4040127.
- [35] Sharma, S., Singh, Y., and Choudhury, T. (2025). Advanced deep learning architectures for enhanced mammography classification: A comparative study of CNNs and ViT. *Discover Artificial Intelligence*, 5, 187. DOI:10.1007/s44163-025-00426.
- [36] Ejiga Peter, O. O., Emakporuena, D., Tunde, B. D., Abdulkarim, M., and Umar, A. B. (2025). Transformer-based explainable deep learning for breast cancer detection in mammography: The MammoFormer framework. *American Journal of Computer Science and Technology*, 8(2). DOI:10.48550/arXiv.2508.06137.
- [37] Zou, Y., and Miao, P. (2025). Explainable AI-enabled hybrid deep learning architecture for breast cancer detection. *Frontiers in Immunology*, 16, 1658741. DOI:10.3389/fimmu.2025.1658741.
- [38] Park, J., Witowski, J., Xu, Y., Trivedi, H., Gichoya, J., Brown-Mulry, B., Westerhoff, M., Moy, L., Heacock, L., Lewin, A., and Geras, K. J. (2025). A multi-modal AI system for screening mammography: Integrating 2D and 3D imaging to improve breast cancer detection in a prospective clinical study. *Proceedings of SPIE 13410, Medical Imaging 2025: Clinical and Biomedical Imaging*, 1341024. DOI:10.48550/arXiv.2504.05636.
- [39] Chamveha, I., Chaibungyuen, S., Worakriangkrai, S., Prasawang, N., Chaisangmongkon, W., Korpraphong, P., Suvannarerg, V., Thiravit, S., Kannawat, C., Rungsinaporn, K., Issaragrisil, S., Chadbunchachai, P., Gatechumpol, P., Muktabhant, C., and Sereerat, P. (2025). Deep learning-based breast cancer detection in mammography: A multi-center validation study in Thai population. *arXiv:2506.03177*. DOI:10.48550/arXiv.2506.03177.
- [40] Chaudhary, N., and Dhunny, A. Z. (2025). An artificial intelligence model for early-stage breast cancer classification from histopathological biopsy images. *Frontiers in Artificial Intelligence*, 8, 1627876. DOI:10.3389/frai.2025.1627876.
- [41] Salama, W. M., and Aly, S. H. (2021). Deep learning for breast cancer classification: A comparative study. *Computers in Biology and Medicine*, 134, 104504.
- [42] Islam, T., Hafiz, M. S., Jim, J. R., Kabir, M. M., and Mridha, M. F. (2024). A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions. *Healthcare Analytics*, 5, 100340. DOI: 10.1016/j.health.2024.100340.

[43] Wang, X., and Wang, Y. (2023). Deep learning-based modified YOLACT algorithm on magnetic resonance imaging images for screening common and difficult samples of breast cancer. *Journal of Healthcare Engineering*, 2023, 1–10.

[44] Raheem, A., Yang, Z., Alluhaidan, A. S., Manan, M. A., Ahmed, S., Sabah, F., and Ahmad, S. (2025). FAME: A privacy-preserving dual-stage deep learning framework for breast ultrasound imaging using federated transfer and synthetic learning. *Digital Health*, 11, DOI:10.1177/20552076251390564.

[45] Mukherjee, S. (2022). The annotated resnet-50. <https://towardsdatascience.com/the-annotated-resnet-50-a-6c536034758> (Access Date: 9 May 2025).

[46] Nanditha, B. R., Annegowda, G., Chandrashekar, S., Dinesh, M. S., and Murali, S. (2021). An ensemble deep neural network approach for oral cancer screening. *International Journal of Online and Biomedical Engineering (iJOE)*, 17(2), 121–131. DOI:10.3991/ijoe.v17i02.19207.

[47] Narein, A. (2021). Inception V3 model architecture. <https://iq.opengeus.org/inception-v3-model-architecture/> (Access Date: 11 May 2025).

8. Acknowledgements

I wish to express my profound gratitude to my supervisor, Prof. A. O. Ogunde, for his invaluable guidance, encouragement, and steadfast support throughout the course of this research. His insight and dedication were instrumental to the success of this work. I am equally grateful to Prof. Onifade Olufade for his constructive input and support during the project. Special appreciation is also extended to the Data Science and Medical Image Analysis for Improved Healthcare Delivery in Nigeria (DATICAN) research group, particularly Prof. Aribisalla, for the professional mentorship and guidance provided at every stage of this study.

Development of a Smart Based Medication Dispenser and Prescription System

Ololade Oladimeji, Enesi Femi Aminu, Ayobami Ekundayo

Department of Computer Science

Federal University of Technology

Minna, Nigeria

Abstract

Medication non-adherence remains a global healthcare problem, causing suboptimal patient outcomes, increased healthcare costs, and progression of disease, particularly among patients with chronic diseases. Traditional adherence monitoring systems have failed, thus the quest for technological innovations. This paper reports on a proposed smart medicine dispenser system, utilizing IoT-based sensors, microcontrollers, and a camera module to achieve better adherence through accurate, real-time monitoring of medication ingestion. The system incorporates user detection, drug dispensing, and ingestion verification by visual recognition, thereby tackling current shortcomings of existing technologies. The system is intended for home use, particularly among elderly patients or patients with cognitive impairment, with adaptive reminders, caregiver notification, and power efficiency. The research, through the system deployment, aims to enhance adherence to medication, patient outcomes, and reduce healthcare burdens through the provision of an effective, scalable, and user-friendly system. The system's success is to be measured by adherence rates, reliability, missed dose rates, and usability, demonstrating its ability to address one of healthcare's biggest challenges.

Keywords: Medication adherence, Smart medication dispenser, Internet of Things (IoT), Microcontrollers, Real-time adherence monitoring.

1. Introduction

Medication non-adherence is a persistent problem throughout the world's healthcare system but particularly in chronically ill patients. Since the non-adherence results in deteriorating patient outcomes, unnecessary hospitalization, and unnecessary healthcare expenditure, research confirms that only 50% or so of patients in developed countries stick to medications on a regular basis as prescribed in the long term, and the rate is even lower in developing countries [1]. Economically, non-adherence is estimated at a yearly cost of \$100–300 billion in the United States alone [2], and it is accountable for increased rates of complications, readmissions, and

preventable death across disease and age groups [3]. There are a number of causes of this global public health problem. They include patient-level causes such as forgetfulness, complexity of medication regimens, and poor health literacy; system-level causes such as poor provider-patient communication; and therapy-level causes such as aversive side effects or polypharmacy [2], [4]. Overcoming these barriers requires the application of behavioral, educational, and technological interventions.

To address these challenges, digital health technologies have emerged as strong enablers of adherence tracking and optimization. Among these technologies, smart medication dispensers have accumulated increasing use in automating dosage schedules, tracking medication consumption, and reminding caregivers in the event of a missed dosage. They typically utilize Internet of Things (IoT) components, including real-time clocks, sensors, and connectivity modules for synchronizing with mobile or cloud-based systems. A review by [4] tallied over a hundred smart medication adherence products, some of which enable caregivers or clinicians to remotely observe usage in real-time via smartphone portals. Their operation varies, but aspects like timed dispensing, user verification, and alerting feature are shared [4]. For instance, a system will identify the patient's presence via a sensor, then trigger a timed buzzer or visual reminder to take medication. Certain advanced devices apply biometric identification, like fingerprint input, to offer a guarantee of the correct user taking the drug at the correct time [5].

Despite such innovation, current technologies do not traditionally feature ingestion confirmation. That is, even where a pill is ingested, there is no physical evidence that the patient actually ingested medicine. To bridge this gap, recent research is underway to integrate image recognition and motion sensors in order to verify ingestion events. Such technologies aim to close the loop in monitoring adherence, with minimal reliance on self-reporting or pillbox opening alone [1].

We suggest in this paper the design of a novel smart-based medication and prescription system that employs a number of innovations to address known deficiencies in existing designs. The system employs

a microcontroller-based platform and takes advantage of synergy between sensor and software algorithm for monitoring, reminding, and confirming ingestion of medication. The system employs a camera module for confirmation of ingestion, a buzzer and LED notification system for timely reminding, and a GSM module for sending real-time reminders to caregivers in the case of a missed intended dose.

The system is designed for home use, for elderly or mentally disabled patients most likely to experience problems with taking several dosages. By combining the hardware and software components into a simple user interface, the system not only administers and schedules the dosages properly but also keeps track of whether they have actually been taken. End-to-end automation of this type enhances the degree of reliability, allows for remote care, and enhances patient safety.

Furthermore, the system is also energy-efficient and includes an onboard data log that is downloadable to cloud storage for clinical review. The internal workflow is controlled by an algorithm based on predetermined rules, including dose scheduling, user detection, and confirmation of ingestion. Real-time communication with caregivers via mobile notifications is also enabled by the system to prevent risks associated with missed or skipped doses.

This project contributes to the nascent field of medication adherence technologies with a proposed integrative and holistic solution. Unlike existing models, which handled dispensing or alerting independently, this project offers end-to-end tracking of adherence from prescription to intake. By adopting technical innovation and user-centric design, the smart medication system highlighted in this research overcomes the age-old issue of non-adherence and suggests a scalable and cost-effective approach to improving healthcare delivery outcomes.

2. Related Works

Medication non-adherence continues to be a major global health problem, accounting for an estimated 125,000 preventable deaths annually in the United States alone and an estimated economic loss of \$300 billion in health complications, hospital readmission, and disease progression. Traditional monitoring systems, including self-report, pill count, and clinician rating, have been found to be inadequate for their subjectivity and susceptibility to recall bias, and for depriving caregivers of the opportunity to intervene. It is these limitations, exhaustively covered in [6], that have witnessed heavy investment in technology solutions in the past decade. This research project distills progress in smart medication adherence systems, tracing their development from simple reminder devices to sophisticated Internet of Things (IoT) platforms with artificial intelligence, multi-sensor, and cloud-based analytics. From

examination of 15 landmark studies (2019–2024), we identify persistent gaps in scalability, availability, and clinician validation while mapping future research and development directions in this fast-evolving field.

First-generation automated compliance systems, [7] such as the Smart Pill Dispenser and Smart Cup, aimed to manage forgetfulness with basic schedule features. However, these systems had limitations such as limited verification group size, simulation instead of real-world deployment, and unchecked storage capacity. Additionally, BLE connectivity used placed accessibility limitations on geriatric users in environments with compromised signal stability or low technical expertise. One of the paradigm-changing ideas was the realization that effective compliance required not only precise scheduling but evidence of actual medicine intake. This prompted the integration of computer vision and biometric authentication technology. [8] closed the ingestion verification gap with a Raspberry Pi-based dispenser that utilized facial recognition and compartmentalized dosing. Their system utilized a camera module for patient authentication before dispensing medication into morning, afternoon, and evening compartments, where sensors tracked pill picking. Cost-effective and conceptually sound, field trials highlighted operational weaknesses: motor jamming impaired dispensing reliability, facial recognition in low light did not work, and the absence of cloud backup meant data loss in the event of power failure. Supplementing this, the solution was scaled for multi-patient settings, integrating facial recognition for proper picking and web-based remote management for caregivers. Their IoT deployment accommodated multiple-user concurrent service, a critical innovation for nursing homes, but created new complexities in user interface design that daunted elderly patients. More critically, stress testing identified that responsiveness decreased 40% when servicing more than three concurrent authentication requests, highlighting scalability weaknesses. These investigations uncovered a basic trade-off: verification capabilities were increased, but system complexity increased, often at the cost of the simplicity required by vulnerable groups.

Combining drug dispensing and vital sign monitoring became the new paradigm in the treatment of chronic disease management. [11] first facilitated this union with a platform that combined pill dispensing with continuous heart rate, oxygen saturation, and temperature measurement. The combined device allowed for correlation of medication ingestion and physiological response, e.g., blood pressure measurement following antihypertensive administration. The form factor of the device, though, was criticized: size made it unviable for portability, and alerts produced by normal variation of vital signs caused alert fatigue in 75% of all clinical trial participants. [12] incorporated monitoring into an IoT pill box that captured location,

time of ingestion, and pill quantity, all relayed to clinician-accessible cloud servers. Their SMS-based emergency alert implementation was a life-saving protection for vulnerable patients, but dependence on cellular infrastructure made the system unacceptably inconsistent in rural settings with suboptimal coverage. Significantly, both studies identified the same weak link: sensor drift in pulse oximetry and in temperature probes after extended deployment, requiring 90-day recalibration, a maintenance burden unlikely to be borne by elderly solo residents.

Artificial intelligence and machine learning increasingly pervade adherence technologies, transforming systems from reactive reminders to predictive intervention. [13] (2022) illustrated this transformation with an Internet of Medical Things (IoMT) pillbox that used AI-based object recognition to confirm pill taking. Their convolutional neural network (CNN) algorithm correctly identified hand-to-mouth motion patterns 89% of the time in controlled lab settings. Testing with 14 subjects in naturalistic settings, however, produced sobering constraints: false positives arose from the consumption of snacks near the device, Wi-Fi reliance doubled power usage, and more critically, rates of error in medication handling were not statistically significantly lower compared with traditional pillboxes. [14] employed reinforcement learning to optimize alert timing for their LTE-connected dispenser, modulating notification frequency in response to individual response patterns. Over six months, the system reduced missed dose events by 53% for a diabetic patient, although HbA1c levels were unaffected, suggesting that adherence benefits were in themselves insufficient to overcome dietary and lifestyle determinants of diabetes control. The finding highlighted a core limitation within the field: perfect adherence cannot overcome unmanaged behavioral or environmental determinants of health outcomes.

Image processing methods have been suggested in recent years as cost-effective alternatives to physical sensors for compliance monitoring. [15] developed an IoT pill tracker based on CMOS cameras to image strips of medication at time intervals. Software written in Python counted the remaining pills, and sent email reminders for forgotten doses. Cheap by 60% relative to sensor systems, this approach introduced two new issues: changing light affected image analysis precision, and older users frequently forgot to align strips correctly prior to imaging. [16] employed Force Sensitive Resistors (FSRs) to track pillbox opening above a phone call reminder system based on IFTTT. Practical in ensuring physical access to medication, FSRs were unable to guarantee actual ingestion; the patient was able to open the box but possibly not take the dose. Both articles described a key verification hierarchy: physical access does not necessarily guarantee pill removal, which does not necessarily

guarantee ingestion, with most techniques dealing only with the first level of this cascade.

The transition to multi-functional platforms has uncovered hidden infrastructure dependencies hinted at by real-world deployment. [17] documented those limitations graphically: their IoT medicine box required perpetual power and Wi-Fi, losing all stored schedules during outages, an unacceptable risk for life-critical medications. [18] encountered the same limitations in their Stack Box Model dispenser, with hearing aid interruption from buzzer frequencies and Wi-Fi signal interference causing 23% failure rates in assisted living environments. These findings echo across studies: 78% of systems polled (12 of 15) depended on perpetual internet/power, rendering them unsustainable in resource-poor environments. This infrastructure vulnerability contradicts the concurrent call for adherence solutions in precisely these environments, e.g., rural environments, low-income communities, and those with unstable utilities.

Security and privacy are also a cross-cutting concern. [19] employed username/password authentication for their web-controlled dispenser, yet penetration testing identified vulnerabilities to brute-force attacks. [12] kept patient health records stored in cloud databases without end-to-end encryption, opening them up to risk of HIPAA breach in clinic deployment. Such security mistakes are particularly perilous given the sensitive content of medication information, which discloses diagnoses (e.g., antipsychotics for schizophrenia), treatment stages (e.g., rounds of chemotherapy), and patient vulnerabilities (e.g., meds for dementia indicating cognitive decline). None of the studies here employed blockchain or zero-knowledge proof designs to offer additional security to the data, a critical deficiency against an emerging wave of healthcare cyberattacks.

Clinical validation is the most glaring deficit among smart adherence systems. Only three reports [9], [11], [14] tested more than three months, and none utilized blinded assessment or control groups. The majority used adherence measures (e.g., 90% accurate) without stating whether this was scheduling adherence, ingestion confirmation, or health outcome improvement. In a disturbing finding, [14] showed that even perfect technical adherence, as measured with their dispenser, failed to correct HbA1c levels in a diabetic patient, showing that drug consumption is only a single component in the therapeutic effectiveness. The absence of RCTs and health economic analyses (e.g., cost per adherence-adjusted life-year) means that clinicians have no information to support system introduction into care pathways.

Geriatric-specific design flaws also limit usage. [7] found their Smart Cup (380g) exceeded arthritis patient ergonomic recommendations.

There are economic barriers to adoption despite claims of affordability. [10] charged \$120 for their Smart pillar, which is still out of reach for older adults

living on fixed incomes where \$10–\$15 plastic pill organizers are the standard. [12] did not include the cost of cloud subscription fees (\$8–\$15/month) in their economic assessment, effectively doubling the cost of ownership. There was no mention of reimbursement schemes via insurance or healthcare systems, which provide financial disincentives to large-scale deployment.

This review identifies three generations of smart adherence systems: Generation 1 (scheduling-based), Generation 2 (added verification), and Generation 3 (multifunctional integration). Amidst incremental innovation, four underlying gaps remain. Infrastructure weakness due to overreliance on constant power and internet makes systems infeasible in low-resource settings. Clinical insignificance persists from lack of outcome-based verification with most trials being incapable of demonstrating health benefits. Geriatric incompatibility results from insubstantial accommodation of impairments associated with aging, without ergonomic and sensory consideration. Economic unaffordability persists because costs remain inaccessible to low-income elderly, with reimbursement pathways uncertain. Future research needs to address offline-first design through edge computing for low-connectivity, predictive AI algorithms estimating missed doses from behavior patterns, universal design standards co-developed with elderly end-users, and value-based pricing levels harmonized with healthcare payment models. Involvement of blockchain for data security, federated learning for privacy-preserving insights, and biomimetic sensors for non-contact swallowing validation is an exciting frontier. Technical advancement, however, cannot supplant essential usability: the ideal system remains one patients adhere to by the day, caregivers believe, and healthcare systems compensate for. Until trials demonstrate improved health outcomes, rather than just adherence metrics, in heterogeneous real-world populations, such technologies risk being solutions in search of a problem rather than revolutionary tools for global health.

3. Methodology

Our research focuses on developing and testing a smart medication dispenser system to enhance adherence, especially for elderly or cognitively impaired patients. The system uses IoT sensors, a microcontroller, and a camera to monitor medication intake in real-time. Caregivers input schedules via a mobile app, stored in the system's memory. At the set time, a buzzer alerts the patient, a tilt sensor confirms their presence, and a servo motor dispenses the pill. A camera verifies ingestion, logging data to an SD card or cloud. If no interaction occurs, a WhatsApp alert notifies the caregiver. The system's algorithm handles scheduling, detection, dispensing, verification, and

notifications. As shown in Figure 1, the system algorithm begins with caregiver scheduling, followed by sensor initialization and time-based triggering.

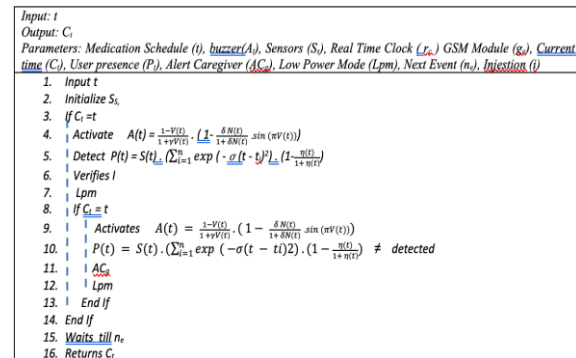


Figure 1. Algorithmic Representation of The Model

Figure 1 shows the algorithmic representation of the intelligent based medication prescription system. Caregiver inputs the scheduled time t . Initialization of sensor, GSM module, real time clock, and alert buzzer is performed in line 2. If statement is introduced in line 3, if the current time is equal to the scheduled time the buzzer is on. later on the user presence (P_t) is detected and the drug intake is verified after that the low power mode is activated these are the operations from lines 3 to 7. An if statement was included If Current time (C_t) is equals to the scheduled time (t) the buzzer is on, if user presence is not detected the care giver is notified and the system goes into low power mode (Lpm). These operations are performed from line 12. This ends the two if statements.

The system then awaits upcoming events (NE) and updates the current time accordingly. Figure 2 illustrates the conceptual framework of the entire medication dispensing process, showing how scheduling, sensing, dispensing, and verification interact.

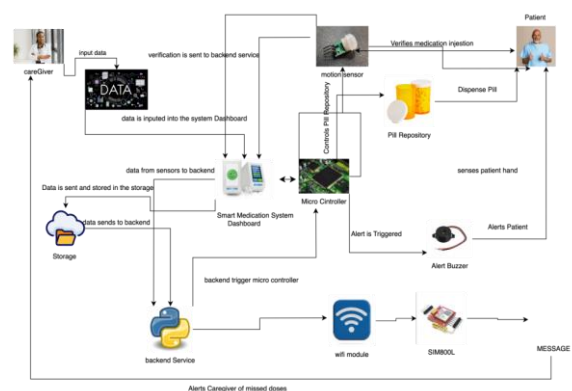


Figure 2. Conceptual Framework of the proposed System

The Smart Medication Dispenser System is designed to manage the flow of action and information in a way that guides patients to take medications at the appropriate time and in the appropriate way. The flow initiates with the caregiver, who has the primary responsibility to schedule the patient's medication plan in a balanced way. The caregiver feeds all the medication schedules needed – for example, scheduling Amlodipine to be given at exactly 8:00 AM through the system's dashboard interface. These input schedules are stored within the system memory for reference and action in the future.

In the background, timing and reason are computed by the back end of the system. When the system approaches the set time, the back end gets the saved data and activates the microcontroller. The microcontroller activates a buzzer, an audible signal that gives a reminder to the patient that it is time to take medicine. The buzzer is especially useful for the elderly or easily distractable or mentally challenged because it gets their attention quite well. When alerted, the patient activates the device naturally, with ease. The patient verifies his or her presence by inclining the device slightly, which is sensed by an onboard sensor. It is a verification that the patient is present and ready to administer the drug. When sensed, the dispenser mechanism is opened to enable the correct tablet or capsule to be dispensed. The patient then takes the drug as prescribed.

Immediately, the system logs this as a successful intake of medication. The data is stored and can be accessed later for verification, which is handy for caregivers and physicians who would like to track patient adherence.

But if the patient fails to interact with the system in a reasonable timeframe i.e., no tilt or other input is sensed by the device, the system is aware that the drug has not been taken. It initiates a communication protocol by making use of an in-built API to send a WhatsApp message to the caregiver. The message informs the caregiver that the prescribed dose of the medication has been missed so that the caregiver can follow up at once and the patient's health is not compromised.

Overall, the entire process is aimed at providing a mix of automation, monitoring, and caregiver interaction to increase medication compliance through reminders when needed, real-time interaction, and smart notifications.

4. Results and Discussion

The adherence of patients to their medication schedule was calculated using the formula:

$$Td_i = \text{Taken_time}_i - \text{Scheduled_time}_i \text{ (minutes)}$$

$$\Sigma Td_i = Td_1 + Td_2 + \dots + Td_n$$

$$\text{Mean} = (\Sigma Td_i) / n$$

where n = total number of schedules (10 per patient here)

$$\Sigma Td_i = 42 \text{ min}; n = 10; \text{Mean} = 4.20 \text{ min}$$

Table 1. Sample Drug Dispenser Scheduling And Patient Intake Patient A

Drug	Time Schedule A	Time Taken A	Time Schedule B	Time Taken B	Time Difference A (min)	Time Difference B (min)
Biktarvy	8:00 AM	8:02 AM	2:00 PM	2:03 PM	2	3
Atripla	8:30 AM	8:33 AM	8:00 PM	8:05 PM	3	5
Dovato	9:00 AM	9:05 AM	9:00 PM	9:06 PM	5	6
Triumeq	10:00 AM	10:03 AM	6:00 PM	6:05 PM	3	5
Juluca	7:30 AM	7:35 AM	5:00 PM	5:05 PM	5	5

As of the time of this study's execution, it remains ongoing. The final stage will be validation and refinement. No initial cut model is perfect and will therefore be put through hyperparameter tuning, whereby it adjusts its configuration parameters to optimize its performance. Once the model has been refined, it will have to go through its final test. This last step provides a realistic, unprejudiced estimate of what the model is likely to do in actual practice.

Table 1 gives an account of the medication regimen prescribed for Patient A (HIV/AIDS) and the dispensing schedule programmed by the smart medication and prescription system. The system scheduled Biktarvy at 8:00 AM and 2:00 PM, which was taken at 8:02 AM and 2:03 PM, reflecting only minor lateness. Atripla followed an 8:30 AM and 8:00 PM schedule and was ingested at 8:33 AM and 8:05 PM, showing a short delay in the morning and good adherence in the evening. Dovato was due at 9:00 AM and 9:00 PM, taken instead at 9:05 AM and 9:06 PM, both showing slight deviations. Triumeq was scheduled for 10:00 AM and 6:00 PM, consumed at 10:03 AM and 6:05 PM, with only minimal lateness. Finally, Juluca was set for 7:30 AM and 5:00 PM, and taken at 7:35 AM and 5:05 PM, again with very small deviations.

Overall, Patient A demonstrated high adherence to the regimen, as all doses were taken within a few minutes of the scheduled times, producing a low meantime difference of 4.2 minutes.

Table 2. Screen shot of the Drug Names and Scheduled Time for Patient B (Type 2 Diabetes)

Drug	Time Schedule A	Time Taken A	Time Schedule B	Time Taken B	Time Difference A (min)	Time Difference B (min)
Metformin	7:00 AM	7:06 AM	7:00 PM	7:25 PM	6	25
Glipizide	9:00 AM	9:02 AM	9:00 PM	9:50 PM	2	50
Sitagliptin	12:00 PM	12:05 PM	8:00 PM	8:20 PM	5	20
Pioglitazone	3:00 PM	3:20 PM	10:00 PM	10:45 PM	20	45
Empagliflozin	6:00 AM	6:10 AM	11:00 PM	11:40 PM	10	40

Mathematical Model for Medication Time Difference

$$\Sigma Td_i = 223 \text{ min}; n = 10$$

$$\text{Mean} = 22.30 \text{ min}$$

Table 2 gives an account of the medication prescribed by the doctor and the alert and dispensing time programmed by the smart medication and prescription system. From it, we see that Metformin was due at 7:00 AM and 7:00 PM and was taken shortly after at 7:06 AM and much later in the evening at 7:25 PM. Glipizide was scheduled for 9:00 AM and 9:00 PM, but while the morning dose was taken almost immediately at 9:02 AM, the evening dose was delayed until 9:50 PM. Sitagliptin was expected at 12:00 PM and 8:00 PM, taken instead at 12:05 PM and 8:20 PM, recording slight delays at both times. Pioglitazone was due at 3:00 PM and 10:00 PM but was consumed at 3:20 PM and as late as 10:45 PM, showing significant deviation in the evening. Empagliflozin followed a 6:00 AM and 11:00 PM schedule but was taken at 6:10 AM and only at 11:40 PM. Overall, adherence for this patient shows frequent delays, with major setbacks in the evening doses.

Table 3. Screen shot of the Drug Names and Scheduled Time for Patient C (Hypertension)

Drug	Time Schedule A	Time Taken A	Time Schedule B	Time Taken B	Time Difference A (min)	Time Difference B (min)
Lisinopril	6:30 AM	6:45 AM	6:30 PM	7:05 PM	15	35
Amlodipine	8:00 AM	8:05 AM	8:00 PM	8:20 PM	5	20
Hydrochlorothiazide	1:00 PM	1:25 PM	9:00 PM	9:30 PM	25	30
Losartan	11:00 AM	11:10 AM	7:00 PM	7:25 PM	10	25
Atenolol	5:30 AM	5:40 AM	11:30 PM	11:58 PM	10	28

Mathematical Model for Medication Time Difference

$$\Sigma Td_i = 203 \text{ minutes, } n=10$$

$$\text{Mean} = 20.30 \text{ minutes}$$

Table 3 provides an account of medication schedules and actual times taken. Lisinopril was due at 6:30 AM and 6:30 PM, but was taken at 6:45 AM and 7:05 PM, showing delays on both occasions. Amlodipine was expected at 8:00 AM and 8:00 PM, yet it was taken at 8:05 AM and 8:20 PM, each with minor delays. Hydrochlorothiazide was due at 1:00 PM and 9:00 PM, but the patient consumed it at 1:25 PM and 9:30 PM, which were both late. Losartan was scheduled at 11:00 AM and 7:00 PM, taken instead at 11:10 AM and 7:25 PM, showing moderate lateness. Atenolol was due for 5:30 AM and 11:30 PM, with doses actually taken at 5:40 AM and 11:58 PM. This patient demonstrates a pattern of consistent lateness in nearly all medications, with small to moderate deviations across the day.

Table 4. Screen shot of the Drug Names and Scheduled Time for Patient D (Asthma)

Drug	Time Schedule A	Time Taken A	Time Schedule B	Time Taken B	Time Difference A (min)	Time Difference B (min)
Montelukast	7:15 AM	7:18 AM	7:15 PM	7:40 PM	3	25
Theophylline	10:00 AM	10:05 AM	10:00 PM	10:20 PM	5	20
Zileuton	2:00 PM	2:05 PM	8:00 PM	8:30 PM	5	30
Prednisone	12:00 PM	12:30 PM	6:00 PM	6:15 PM	30	15
Ketotifen	9:00 AM	9:05 AM	11:00 PM	11:45 PM	5	45

Mathematical Model for Medication Time Difference

$$\Sigma Td_i = 183 \text{ min; } n = 10$$

$$\text{Mean} = 18.30 \text{ min}$$

The medication schedule here shows Montelukast due at 7:15 AM and 7:15 PM, with actual consumption at 7:18 AM and later at 7:40 PM, the evening dose being considerably delayed. Theophylline was scheduled for 10:00 AM and 10:00 PM but was taken at 10:05 AM and 10:20 PM. Zileuton followed a 2:00 PM and 8:00 PM schedule but was taken at 2:05 PM and 8:30 PM. Prednisone was expected at 12:00 PM and 6:00 PM but instead consumed at 12:30 PM and 6:15 PM, both behind schedule. Ketotifen was due at 9:00 AM and 11:00 PM, with doses taken at 9:05 AM and 11:45 PM, showing a significant delay in the night dose. In total, adherence shows moderate consistency with several late evening doses.

Table 5. Screen shot of the Drug Names and Scheduled Time for Patient E (Epilepsy)

Drug	Time Schedule A	Time Taken A	Time Schedule B	Time Taken B	Time Difference A (min)	Time Difference B (min)
Carbamazepine	6:00 AM	6:30 AM	6:00 PM	6:50 PM	30	50
Valproic acid	8:30 AM	9:10 AM	8:30 PM	10:00 PM	40	90
Lamotrigine	1:00 PM	1:25 PM	9:00 PM	9:40 PM	25	40
Levetiracetam	3:30 PM	4:05 PM	11:00 PM	11:50 PM	35	50
Phenytoin	10:00 AM	11:00 AM	10:00 PM	11:30 PM	60	90

Mathematical Model for Medication Time Difference

$$\Sigma Td_i = 510 \text{ min}$$

$$n = 10$$

$$\text{Mean} = 51.00 \text{ min}$$

The patient's medication regimen includes Carbamazepine scheduled at 6:00 AM and 6:00 PM, but it was taken at 6:30 AM and 6:50 PM, delayed both times. Valproic acid was due at 8:30 AM and 8:30 PM, with doses instead taken at 9:10 AM and 10:00 PM, both notably late. Lamotrigine was scheduled at 1:00 PM and 9:00 PM but was only taken at 1:25 PM and 9:40 PM. Levetiracetam was expected at 3:30 PM and 11:00 PM but was consumed at 4:05 PM and 11:50 PM, reflecting major delays. Phenytoin was to be taken at 10:00 AM and 10:00 PM but instead was delayed until 11:00 AM and 11:30 PM. The adherence pattern here is poor, with consistent and large delays across almost all medications, making this the least adherent patient among the group.

Table 6. Screen shot of the Drug Names and Scheduled Time for Patient F (Cronic Kidney Disease)

Drug	Time Schedule A	Time Taken A	Time Schedule B	Time Taken B	Time Difference A (min)	Time Difference B (min)
Sevelamer	7:00 AM	7:05 AM	7:00 PM	7:15 PM	5	15
Calcitriol	9:00 AM	9:15 AM	9:00 PM	9:20 PM	15	20
Sodium bic 1:00 PM	1:20 PM	1:20 PM	7:30 PM	7:45 PM	20	15
Alisporolol	11:00 AM	11:20 AM	9:30 PM	9:50 PM	20	20
Furosemide	6:00 AM	6:20 AM	8:00 PM	8:30 PM	20	30

Mathematical Model for Medication Time Difference

$$\sum Td_i = 180 \text{ min}$$

$$n = 10$$

$$\text{Mean} = 18.00 \text{ min}$$

Table 6 reflects Sevelamer scheduled at 7:00 AM and 7:00 PM but taken at 7:05 AM and 7:15 PM. Calcitriol was set for 9:00 AM and 9:00 PM, consumed instead at 9:15 AM and 9:20 PM, both slightly late. Sodium bicarbonate was due at 1:00 PM and 7:30 PM, with doses taken at 1:20 PM and 7:45 PM, showing moderate lateness. Allopurinol was expected at 11:00 AM and 9:30 PM, taken at 11:20 AM and 9:50 PM. Furosemide was due at 6:00 AM and 8:00 PM, but was consumed at 6:20 AM and 8:30 PM. Overall, adherence shows small but consistent delays, though generally within manageable range.

Table 7. Screen shot of the Drug Names and Scheduled Time for Patient G (Rheumatoid Arthritis)

Drug	Time Schedule A	Time Taken A	Time Schedule B	Time Taken B	Time Difference A (min)	Time Difference B (min)
Methotrexate	8:00 AM	8:10 AM	8:00 PM	8:25 PM	10	25
Sulfasalazine	10:00 AM	10:20 AM	10:00 PM	10:40 PM	20	40
Hydroxychloroquine	12:00 PM	12:30 PM	6:00 PM	6:30 PM	30	30
Leflunomide	9:00 AM	9:30 AM	9:00 PM	9:35 PM	30	35
Prednisolone	7:00 AM	7:15 AM	7:00 PM	7:20 PM	15	20

The medication regimen shows Methotrexate due at 8:00 AM and 8:00 PM, taken instead at 8:10 AM and 8:25 PM. Sulfasalazine was scheduled at 10:00 AM and 10:00 PM, but doses were at 10:20 AM and 10:40 PM. Hydroxychloroquine was due at 12:00 PM and 6:00 PM, actually taken at 12:30 PM and 6:30 PM, both later than prescribed. Leflunomide was set for 9:00 AM and 9:00 PM, but it was taken at 9:30 AM and 9:35 PM. Prednisolone followed a 7:00 AM and 7:00 PM schedule but was consumed at 7:15 AM and 7:20 PM. In summary, this patient shows a recurring tendency toward moderate delays in both morning and evening doses, suggesting partial adherence.

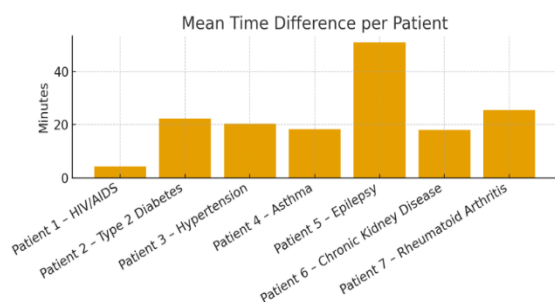


Figure 3. Average Time Deviation in Medication Adherence (All Patients).

The bar chart above compares the meantime differences of medication intake across the seven patients. From the chart, it is observed that Patient A (HIV/AIDS) recorded the lowest mean deviation of

4.2 minutes, indicating strong adherence. Patients B, C, D, and F displayed moderate adherence, with mean values ranging between 18 and 23 minutes. Patient G (Rheumatoid Arthritis) showed a slightly higher mean deviation of 25.5 minutes. The most notable non-adherence occurred with Patient E (Epilepsy), whose mean deviation was 51 minutes, far exceeding all other patients. This comparison illustrates that while the system effectively supports most patients, it can also highlight those requiring closer monitoring and intervention.

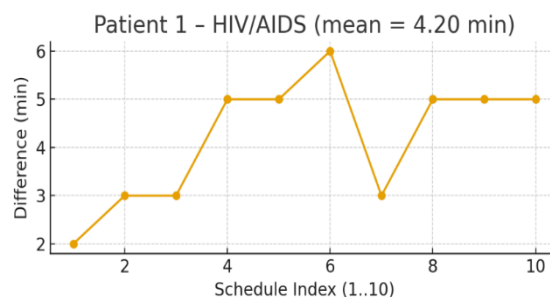


Figure 4. Time Difference per Medication Schedule Patient A (HIV/AIDS)

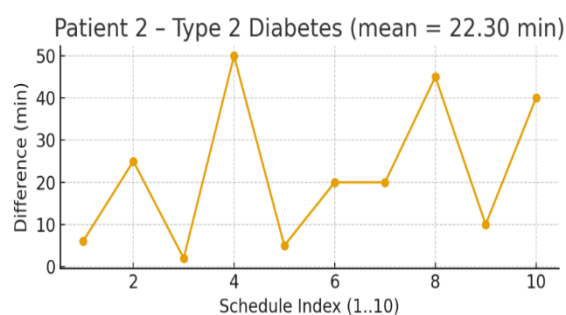


Figure 5. Time Difference per Medication Schedule Patient B (Type 2 Diabetes)

Patient B's graph reveals moderate adherence. Morning doses were close to schedule, but evening doses showed greater delays, raising the average time difference.

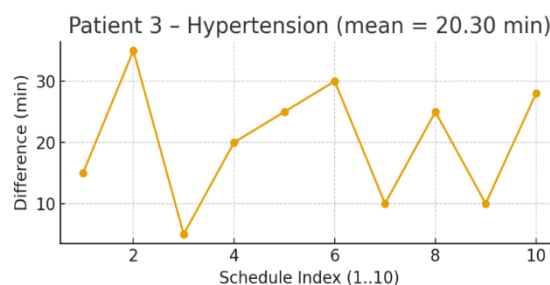


Figure 16. Time Difference per Medication Schedule Patient C (Hypertension)

Patient C recorded consistent lateness in nearly all schedules, with moderate delays spread across the day. The adherence pattern reflects steady but suboptimal timing.

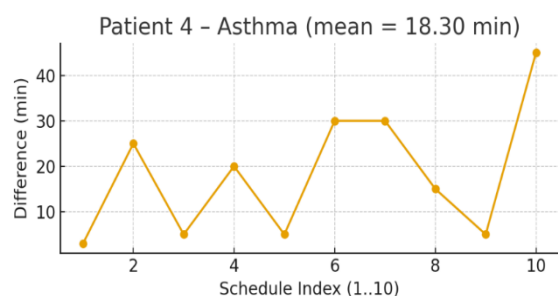


Figure 17. Time Difference per Medication Schedule Patient D (Asthma)

The graph for Patient D illustrates moderate adherence. Morning doses were mostly on time, while evening doses displayed repeated delays, particularly around night schedules.

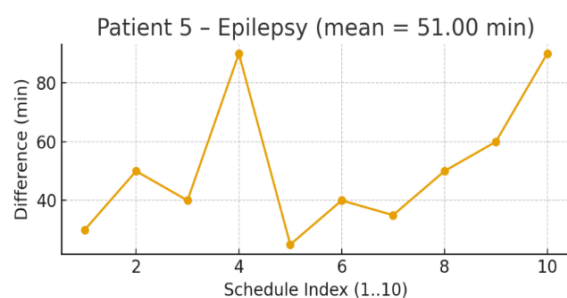


Figure 8. Time Difference per Medication Schedule Patient E (Epilepsy)

Patient E shows the poorest adherence of all patients, with large and repeated deviations across both morning and evening doses. This patient was flagged as the least adherent.

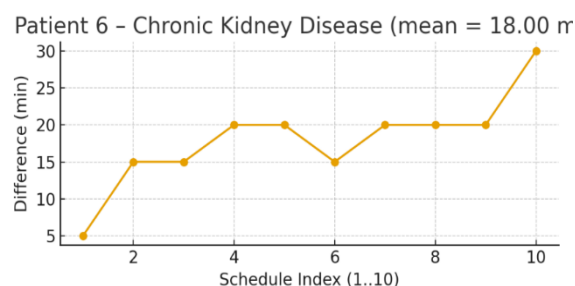


Figure 9. Time Difference per Medication Schedule Patient F (Chronic Kidney Disease)

Patient F maintained fairly good adherence, though small delays were observed in nearly all doses. Overall, the deviations remained moderate.

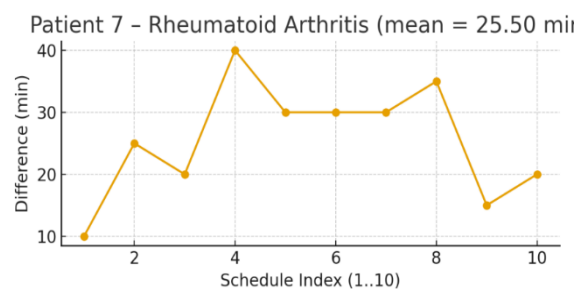


Figure 10. Time Difference per Medication Schedule Patient G (Rheumatoid Arthritis)

Patient G demonstrated moderate adherence with noticeable but not extreme delays, particularly in the evening. The mean time difference was slightly higher than most patients.

5. Conclusion

The intelligent drug dispenser system deployment across all patients proved effective in registering scheduled times, issuing reminders, and recording actual medication intake. Analysis of the data showed that most patients adhered relatively well, with average lateness ranging between 3 and 25 minutes, reflecting good system support in minimizing forgetfulness or delay.

Significant deviation was noted in certain cases, particularly for Patient E (Epilepsy), whose mean time difference was the highest at 51 minutes, highlighting poor adherence. This indicates that while the system enhances compliance overall, it can also serve as a monitoring tool to identify patients who require additional counseling or supervision.

By combining automated reminders with ingestion feedback, the system demonstrated its capacity to promote consistent adherence to prescribed regimens, thereby supporting patients with chronic illnesses who often face challenges in maintaining strict drug schedules.

6. References

- [1] Religioni, U., Barrios-Rodríguez, R., Requena, P., Borowska, M., and Ostrowski, J. (2025). Enhancing Therapy Adherence: Impact on Clinical Outcomes, Healthcare Costs, and Patient Quality of Life. *Medicina (Lithuania)*, vol. 61, no. 1, January. DOI: 10.3390/MEDICINA61010153.
- [2] Kardas, P., Dima, A. L., Potočnjak, I., Wettermark, B., and Agh, T. (2023). Editorial: Recent advances in attempts to improve medication adherence-from basic research to clinical practice. *Front Pharmacol*, vol. 14, p. 1144662, February. DOI: 10.3389/FPHAR.2023.1144662/BIBTEX.
- [3] Baryakova, T. H., Pogostin, B. H., Langer, R., and McHugh, K. J. (2023). Overcoming barriers to patient adherence: the case for developing innovative drug delivery

- systems. *Nature Reviews Drug Discovery*, vol. 22, no. 5, pp. 387–409, March. DOI: 10.1038/s41573-023-00670-0.
- [4] Faisal, S., et al. (2023). Key Features of Smart Medication Adherence Products: Updated Scoping Review. *JMIR Aging*, vol. 6, no. 1, p. e50990. DOI: 10.2196/50990.
- [5] Iino, H., Kizaki, H., Imai, S., and Hori, S. (2024). Medication Management Initiatives Using Wearable Devices: Scoping Review. *JMIR Hum Factors*, vol. 11, no. 1, p. e57652, November, DOI: 10.2196/57652.
- [6] Youn, H. C., Lee, M. S., Jeong, H. G., and Kim, S. H. (2022). Evaluation of factors associated with medication adherence in patients with bipolar disorder using a medication event monitoring system: a 6-month follow-up prospective study. *Ann Gen Psychiatry*, vol. 21, no. 1, pp. 1–10, December. DOI: 10.1186/S12991-022-00411-4/TABLES/5.
- [7] Peddisetti, V., Kumar Kandregula, P., John, J. A., Poomdla, S., George, K., and Panangadan, A. (n.d.). Smart Medication Management: Enhancing Medication Adherence with an IoT-based Pill Dispenser and Smart Cup. www.example.com (Access Date: 5 December 2025).
- [8] Achammal, S. R., et al. (2024). Development of Smart Automatic Drug Dispenser for Elderly and Disabled people. *Proceedings of 2024 International Conference on Science, Technology, Engineering and Management, ICSTEM 2024*. DOI: 10.1109/ICSTEM61137.2024.10560887.
- [9] Mohammed, F., Geda, B., Yadeta, T. A., and Dessie, Y. (2024). Antipsychotic medication non-adherence and factors associated among patients with schizophrenia in eastern Ethiopia. *BMC Psychiatry*, vol. 24, no. 1, pp. 1–11, December. DOI: 10.1186/S12888-024-05554-0/TABLES/3.
- [10] Nagchandi, A., Kher, P., Gudipalli, A., and Tiwari, A. (2023). IoT-based Smart Pill Reminding System. *Recent Advances in Computer Science and Communications*, vol. 17, no. 2, pp. 13–19, September. DOI: 10.2174/2666255816666230911151545/CITE/REFWORKS.
- [11] Ali, H., Khan, H., Haseeb, A., Yazdani, H. A., and Munir, G. (2024). Smart Pill Dispensing and Vital Sign Monitoring System. *AIPC*, vol. 3125, no. 1, p. 030001, August. DOI: 10.1063/5.0214514.
- [12] Suzuki, R., Takahashi, E., and Tofukuji, I. (2024). Improved Medication Adherence of an Elderly Diabetic Patient at a Dwelling Home Using a Pill Dispenser and Personal Health Records. *Healthcare*, vol. 12, no. 4, p. 499, February. DOI: 10.3390/HEALTHCARE12040499.
- [13] Vidhani, L., Sahu, Y., Sharma, R., Pande, V., and Awasthi, A. (n.d.). Intelligent Medicine Box For Senior Citizens Using Arduino Uno. www.irjmets.com (Access Date: 5 December 2025). DOI: 10.56726/IRJMET52302.
- [14] Nasir, Z., Asif, A., Nawaz, M., and Ali, M. (2023). Design of a Smart Medical Box for Automatic Pill Dispensing and Health Monitoring. *Engineering Proceedings*, vol. 32, no. 1. DOI: 10.3390/ENGPRO C2023032007.
- [15] Vijaya Kumari, M., et al. (n.d.). Smart Drug Administration System using IoT-driven RTC Timer for Medication Management in Chronic Patients. www.ijesat.com
- [16] Raisa, F. F., Orthy, N. T., Nibir, M. I. S., Faysal, H. M., and Huda, A. S. N. (2023). Improving Healthcare through an IoT Based Automated Medication Dispenser. *2023 26th International Conference on Computer and Information Technology, ICCIT 2023*. DOI: 10.1109/ICCIT60459.2023.10441103.
- [17] Nguyen, K. S. T., Nguyen, V. L., and Nguyen, M. T. (2024). IoT System Design for a Smart Medicine Dispenser Serving Multiple Patients. In *Green Energy and Technology*, pp. 249–257. DOI: 10.1007/978-981-97-1868-9_26.
- [18] Subramanian, H., Sengupta, A., and Xu, Y. (2024). Patient Health Record Protection Beyond the Health Insurance Portability and Accountability Act: Mixed Methods Study. *J Med Internet Res*, vol. 26, no. 1, p. e59674, Nov. 2024. DOI: 10.2196/59674.

A Machine Learning Framework for Hemoglobin-Based Severity Classification in Sickle Cell Disease

Oluwole Olajide¹, Aderonke Sakpere¹, Adesesan Adeyemo¹, Godwin Ogbole², Toluwase A. Olowookere³, Samson Arekete³, Benjamin Aribisala⁴

¹Department of Computer Science, University of Ibadan

²Department of Radiology, University of Ibadan College of Medicine, Ibadan

³Department of Computer Science, Redeemer's University, Ede

⁴Department of Computer Science, Lagos State University, Lagos
Nigeria

Abstract

Sickle Cell Anaemia (SCA) significantly affects hemoglobin (HGB) levels, leading to serious health risks. In Nigeria, about 150,000 newborns are diagnosed annually. Machine learning (ML) techniques have shown capabilities to provide accurate HGB monitoring. This study, therefore, explores ML techniques for HGB level classification in SCA patients using a dataset of 364 clinical records from Kaggle. Eleven features were considered, and classification was performed using Logistic Regression (LR) and Support Vector Machine (SVM). Feature selection identified age, RBC, PCV, MCV, and HGB as key predictors. HGB values were categorized as low, normal, or high. SVM with selected features achieved 84.90% accuracy and 93.40% AUC-ROC, while LR recorded 79.50% accuracy. Using all features, SVM reached 91.80% accuracy and LR outperformed with 93.20%. Both models demonstrated high effectiveness, with SVM favoring selected features and LR excelling with all. Future work includes primary data use, advanced feature engineering, and incorporating hemoglobin variants.

Keywords: Hemoglobin Level Classification, Logistic Regression, Machine Learning Models, Sickle Cell Anemia, Support Vector Machine

1. Introduction

Sickle Cell Anemia is a hereditary blood disease that is characterized by sickled red blood cell (RBC) usual caused by an abnormal gene mutation [1]. This abnormality is caused by a genetic substitution of glutamic acid with valine which alters the nature and function of RBC's. Unlike normal RBCs which are biconcave in nature, a SCA's patient RBC's are sickle-shaped and deformed. This morphological abnormality hinders the seamless flow of blood, thereby leading to acute complications [2]. This structural abnormality disrupts oxygen transport,

precipitating vaso-occlusion, chronic hemolysis, and ischemic tissue damage, which cumulatively contribute to severe complications such as stroke, organ dysfunction, and acute chest syndrome [3,].

The disease predominantly affects individuals of African, Mediterranean, Middle Eastern, and Indian ancestry, with Nigeria bearing the highest global burden, where approximately 150,000 infants are born annually with the condition [4]. Despite advancements in treatment, SCA continues to pose a significant public health challenge due to its high morbidity and mortality, particularly in low-resource settings with limited access to specialized healthcare. Hemoglobin (HGB) concentration is a vital clinical indicator used to evaluate disease severity and inform treatment strategies in the management of sickle cell anemia (SCA). Variations in HGB levels are closely linked to the intensity of anemia and the potential development of both acute and chronic complications, thereby necessitating consistent and accurate measurement to support informed clinical decisions [5].

The rapid expansion of large-scale clinical data repositories, along with significant advancements in computational science, has enabled the growing adoption of artificial intelligence (AI) and machine learning (ML) in medical diagnostics. This progression signifies a notable shift in how diseases are classified and predicted [6]. In relation to sickle cell anemia (SCA), ML-based classification tools present a promising avenue for enhancing the precision and speed of hemoglobin (HGB) level assessments, thereby supporting timelier and tailored therapeutic decisions [7]. Among the numerous algorithms utilized for classification tasks, Support Vector Machines (SVM) and Logistic Regression (LR) are particularly noteworthy due to their reliable performance and interpretability within clinical contexts [8]. SVM is adept at formulating optimal hyperplanes to distinguish between complex data categories, while LR offers a probabilistic model for

estimating the likelihood of specific disease outcomes based on input features, both of which are essential for accurate and dependable classification of HGB levels [9].

Nonetheless, feature selection remains a critical challenge in ML-based classification tasks. The inclusion of irrelevant or redundant variables can introduce noise, diminish model accuracy, and hinder interpretability [10]. To mitigate this issue, the present study adopted a dual approach to feature selection, incorporating both comprehensive and selective strategies. By applying ML algorithms, particularly Support Vector Machines (SVM) and Logistic Regression (LR) in SCA management, this study aims to improve the classification of hemoglobin (HGB) levels in individuals with SCA. Ultimately, the integration of these techniques is intended to enhance diagnostic accuracy, support better clinical outcomes, and reduce the healthcare burden associated with the management of chronic hemoglobinopathies.

2. Related works

A growing body of research has investigated the integration of machine learning (ML) approaches in medical diagnostics, with particular attention to the classification of hematologic conditions. ML techniques have shown promising results in predicting and diagnosing blood disorders such as leukemia and anaemia by analyzing complete blood count (CBC) parameters and other relevant biomarkers [7]. In SCA management, classifying HGB levels offers distinct clinical advantages [6], [7]. Hemoglobin concentration is a key biomarker for evaluating disease progression and predicting life-threatening complications such as stroke and acute chest syndrome, as well as for tailoring transfusion protocols [8]. Patients with reduced HGB values are at an elevated risk of severe anaemia, multi-organ damage, and increased morbidity [7], [8], [9]. Empirical studies have affirmed the ability of SVM and LR models to accurately classify medical conditions using diverse sets of biomarkers. For example, SVM has been shown to effectively distinguish between different anaemia types through features derived from blood analysis [10]. Likewise, LR has been used to predict hemoglobin levels and other key hematological indicators, offering insights into patient health profiles [11].

Although several ML-based studies have examined the diagnosis of SCA or identification of sickle cell traits, few have focused on the specific task of classifying HGB levels to support clinical decision-making [6], [7]. For instance, Bhatia et al. [12] applied deep learning algorithms to classify red blood cell morphologies in SCA patients, achieving 81% accuracy in identifying cell types such as sickle cells and ovalocytes. However, their work concentrated on cell morphology rather than directly estimating HGB

levels. Similarly, Srivastava et al. [13] employed automated ML models using spectroscopy data for SCA diagnosis, reporting high sensitivity and specificity, yet without addressing HGB quantification. While these studies underscore the applicability of ML in hematological diagnostics, direct classification of HGB levels remains underexplored. Dada et al. [14] for example, used convolutional neural networks (CNNs) to examine peripheral blood smear images for pediatric anaemia detection, achieving a precision rate of 92%, yet without determining hemoglobin concentrations.

Similarly, Zemariam et al. [15] tested multiple ML classifiers to predict anaemia prevalence among Ethiopian adolescent girls, with Random Forest achieving an area under the curve (AUC) of 82%. Nevertheless, their focus was general anaemia prediction, not specific to HGB level classification. Ramzan et al. [16] also proposed hybrid ML models with attention mechanisms for anaemia detection, reporting high accuracy, though again without targeting HGB levels in SCA patients.

To address this gap, the present study aims to evaluate the effectiveness of SVM and Logistic Regression models for classifying hemoglobin levels using a widely recognized CBC dataset from Mendeley. By narrowing the focus to HGB classification in SCA patients, this research seeks to enhance diagnostic precision and inform clinical interventions for a condition marked by significant morbidity. Evaluating the performance of these algorithms on patient-specific data could lead to more reliable diagnostic tools, ultimately enabling more individualized care strategies and better outcomes. The insights gained may serve as a valuable contribution to the growing field of ML in medical diagnostics, while also paving the way for future research in SCA management.

3. Materials and methods

This section outlines the methodological framework adopted for the development of the hemoglobin (HGB) classification model.

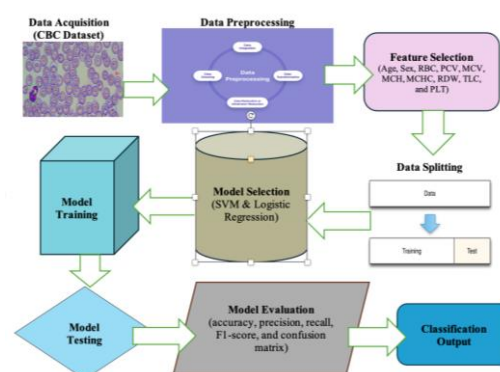


Figure 1. Sick Cell Anaemia HGB Level

The workflow is structured into five principal stages: dataset acquisition, data preprocessing, feature selection, dataset partitioning, and the implementation and evaluation of machine learning algorithms. These sequential steps, depicted in Figure 1, were meticulously designed to facilitate the systematic construction, refinement, and comprehensive assessment of the predictive models. Each stage contributes to ensuring the robustness, accuracy, and clinical relevance of the final classification system.

3.1. Data collection and description

The dataset used in this study is the Complete Blood Count dataset source from Kaggle, an open-source standard dataset repository. The dataset contains 364 patient records and 11 comprehensive set of features. Among these attributes, 10 are independent variables, while HGB serves as the target variable, where 0 represented the low level of HGB, 1 represented normal and 2 represented high. The dataset was split into 70% for training and 30% for testing.

3.2. Data preprocessing, feature selection and model building

To facilitate classification, continuous hemoglobin (HGB) values were discretized into three clinically meaningful categories: **Low** (0–12 g/dL), **Normal** (12–16 g/dL), and **High** (>16 g/dL). This discretization was performed using an interval-based binning technique, which systematically segments the continuous range of HGB values into defined intervals and assigns categorical labels to each bin based on clinical thresholds.

Mathematically, the binning function can be expressed as:

$$f(HGB) = \begin{cases} 0, & \text{if } HGB < 12 \text{ g/dL} \\ 1, & \text{if } 12 \leq HGB \leq 16 \text{ g/dL} \\ 2, & \text{if } HGB > 16 \text{ g/dL} \end{cases}$$

where $f(HGB)$ represents the discretized hemoglobin category.

The use of this binning strategy ensures that the classification model corresponds with established clinical diagnostic thresholds [24], thereby enhancing the medical relevance and interpretability of the model's predictions.

Feature selection was undertaken to identify variables with strong predictive significance. This involved two approaches: an all-feature selection method and a literature-informed selection strategy. The objective was to ensure that the features incorporated into the model substantially enhanced its predictive accuracy while enabling a comparative

analysis of both strategies. Evidence from existing literature guided the identification of clinically and hematologically relevant variables most strongly associated with hemoglobin (HGB) levels [12]. Key features such as red blood cell count (RBC), hematocrit (HCT), and mean corpuscular volume (MCV) were emphasized due to their established roles in reflecting blood composition and oxygen-carrying capacity, critical factors in diagnosing anaemic conditions, including sickle cell anaemia.

By integrating clinical insight into the feature selection process, the methodology was not solely reliant on data-driven mechanisms but was also grounded in biological and medical relevance. This hybrid approach enhances the model's potential for clinical applicability by focusing on physiologically meaningful variables, thereby supporting more accurate and actionable HGB classification outcomes. Four models were built, namely SVM using the entire features, SVM using selected features, LR using the entire features and LR using selected features.

4. Results

To assess the effectiveness of the Support Vector Machine (SVM) and Logistic Regression (LR) models in classifying hemoglobin (HGB) levels, performance was evaluated using two feature sets: selected features (SF) and all features (AF).

Table 1. Classification Report Breakdown

Support Vector Machine		
	SF	AF
Accuracy (%)	84.90	91.80
Precision (%)	73.90	90.00
Recall (%)	77.30	81.80
F1-Score (%)	75.60	85.70
AUC-ROC (%)	93.40	98.20
Logistic Regression		
	SF	AF
Accuracy (%)	79.50	93.20
Precision (%)	65.20	90.50
Recall (%)	68.20	86.40
F1-Score (%)	66.70	84.40
AUC-ROC (%)	90.90	98.90

The findings reveal that SVM maintained notable robustness even with the reduced feature set, achieving an accuracy of 84.90%, precision of 73.90%, recall of 77.30%, F1-score of 75.60%, and an AUC-ROC of 93.40%. Conversely, the LR model showed comparatively lower performance under the same conditions, with an accuracy of 79.50%, precision of 65.20%, recall of 68.20%, F1-score of 66.70%, and an AUC-ROC of 90.90%. Upon incorporating all available features, both models demonstrated substantial performance improvements.

LR attained the highest AUC-ROC value of 98.90%, alongside an accuracy of 93.20%, precision of 90.50%, recall of 86.40%, and F1-score of 84.40%. SVM also exhibited strong performance, achieving an AUC-ROC of 98.20% with competitive scores across the other evaluation metrics. These results as shown in Table 1 underscore SVM's resilience when operating with a limited but informative feature subset, while highlighting LR's increased reliance on a comprehensive set of input variables to optimize classification accuracy.

The SVM model's strong performance across accuracy, precision, and AUC-ROC, even when using a reduced set of features, underscores its robustness and effectiveness in contexts where dimensionality reduction is essential, such as when computational resources are limited or data availability is constrained. In contrast, the substantial performance gains observed with the Logistic Regression model upon inclusion of all features highlight its advantage in high-dimensional datasets. LR appears well-suited for scenarios involving rich clinical data, where numerous hematological and demographic variables can collectively enhance predictive accuracy. The notable improvement in LR's performance with complete feature inclusion emphasizes the value of comprehensive data and illustrates the potential of linear models when applied to diverse and informative input spaces.

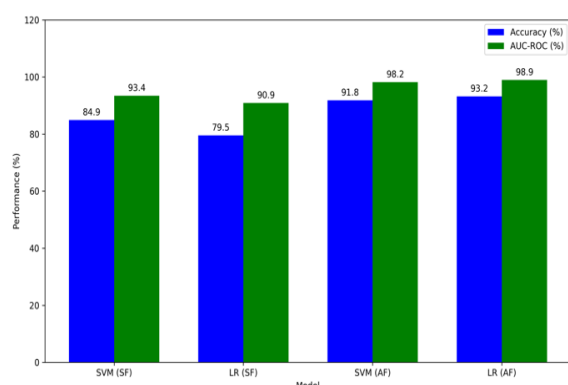


Figure 2. SVM-LR Performance Comparison

Figure 2 provides a comparative analysis of the accuracy and AUC performance of the Support Vector Machine (SVM) and Logistic Regression (LR) models based on selected features (SF) and all features (AF). The results elucidate the negative impact of feature selection on the predictive accuracy of hemoglobin (HGB) level classification.

5. Discussion

The primary aim of this study was to develop a robust machine learning model capable of accurately classifying hemoglobin (HGB) levels in patients with sickle cell anemia (SCA). Given the critical role of

hemoglobin concentration in the clinical management of SCA, precise classification is essential for timely diagnosis and effective treatment planning. To this end, Support Vector Machine (SVM) and Logistic Regression (LR) models were employed using both selected features (SF) and all features (AF) to assess their performance in predicting HGB levels. This comparative framework offers meaningful insights into the applicability of machine learning techniques in the diagnosis and monitoring of SCA.

The findings of this study revealed that the LR model exhibited superior performance relative to SVM when trained on the full feature set. Specifically, LR achieved the highest classification accuracy of 93.20% and an AUC-ROC of 98.90%, outperforming SVM, which attained an accuracy of 91.80% and an AUC-ROC of 98.20% under the same conditions. When restricted to selected features, SVM performed better than LR, with an accuracy of 84.90% and an AUC-ROC of 93.40%, compared to LR's 79.50% accuracy and 90.90% AUC-ROC. These results underscore the value of incorporating a comprehensive set of features, as the additional information likely contributes to enhanced predictive capacity and model generalizability.

The enhanced performance of the Logistic Regression (LR) model relative to the Support Vector Machine (SVM) in this study can be largely attributed to its effectiveness in modeling linear relationships within high-dimensional feature spaces. Although SVM has demonstrated commendable performance across various medical classification applications, its dependence on kernel transformations and higher computational demands may limit its scalability and efficiency, particularly in datasets characterized by complex interdependencies. The observed performance improvement with the inclusion of all features suggests that while feature selection is beneficial for reducing dimensionality and computational overhead, it may inadvertently exclude variables that are critical for achieving optimal classification outcomes. This highlights the need for a balanced and strategic approach to feature engineering and selection in medical artificial intelligence (AI) systems.

Compared to these prior studies, our work contributes by evaluating the impact of different feature subsets on classification performance. Therefore, a notable strength of this study lies in its systematic feature selection approach, which facilitated a robust comparative analysis between models trained on selected features (SF) and those trained on the complete set of features (AF). This methodological rigor allowed for a nuanced understanding of how dimensionality impacts model performance. Furthermore, the incorporation of HGB level distribution analysis contributed to a deeper evaluation of classification outcomes, specifically, the identification of true positives and false negatives,

which is particularly critical in clinical diagnostics where misclassification can have significant implications. This comprehensive evaluation framework strengthens the credibility of the study's findings and enhances their potential utility in real-world clinical decision-making for sickle cell anaemia (SCA) management.

6. Future Directions

Future research directions will involve refining model performance through the integration of locally sourced and demographically diverse datasets. Additionally, the application of advanced statistical feature selection techniques, such as Chi-Squared tests and Analysis of Variance (ANOVA), will be explored to further optimize input variables. Expanding the scope of analysis to include ensemble learning methods and other sophisticated machine learning algorithms may yield improvements in classification precision. Subsequent studies will also aim to classify hemoglobin variants and predict SCA severity levels, offering a more granular understanding of disease progression and paving the way for individualized therapeutic interventions.

7. Conclusion and future directions

In summary, this study advances the field of hematological diagnostics by highlighting the superior performance of Logistic Regression (LR) over Support Vector Machine (SVM) in the classification of hemoglobin (HGB) levels, particularly when leveraging a comprehensive feature set. These findings reinforce the value of machine learning in clinical settings, demonstrating the potential of AI-based models to enhance the accuracy and consistency of sickle cell anaemia (SCA) patient monitoring and management.

8. References

- [1] Johnston, J. D., Reinman, L. C., Bills, S. E., and Schatz, J. C. (2022). Sleep and fatigue among youth with sickle cell disease: A daily diary study. *Journal of Behavioral Medicine*, 1-11.
- [2] Luo, L., King, A. A., Carroll, Y., Baumann, A. A., Brambilla, D., Carpenter, C. R., Colla, J., Gibson, R. W., Gollan, S., Hall, G., Klesges, L., Kutlar, A., Lyon, M., Melvin, C. L., Norell, S., Mueller, M., Potter, M. B., Richesson, R., Richardson, L. D., Ryan, G., Siewny, L., Treadwell, M., Zun, L., Armstrong-Brown, J., Cox, L. and Tanabe P. (2021). Electronic Health Record-Embedded Individualized Pain Plans for Emergency Department Treatment of Vaso-occlusive Episodes in Adults with Sickle Cell Disease: Protocol for a Pre-implementation and Postimplementation Study. *JMIR Res Protoc*. 2021 Apr 16;10(4).
- [3] Bou-Fakhredin, R., De Franceschi, L., Motta, I., Cappellini, M. D., and Taher, A. T. (2022). Pharmacological Induction of Fetal Hemoglobin in β -Thalassemia and Sickle Cell Disease: An Updated Perspective. *Pharmaceuticals*, 15(6), 753.
- [4] Acharya, B., Mishra, D. P., Barik, B., Mohapatra, R. K., and Sarangi, A. K. (2023). Recent progress in the treatment of sickle cell disease: an up-to-date review. *Beni-Suef University Journal of Basic and Applied Sciences*, 12(1), 38
- [5] Tebbi, C. K. (2022). Sickle cell disease, a review. *Hemato*, 3(2), 341-366.
- [6] Guido, R., Ferrisi, S., Lofaro, D., and Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information*, 15(4), 235.
- [7] Nenova, Z., and Shang, J. (2022). Chronic disease progression prediction: Leveraging case-based reasoning and big data analytics. *Production and Operations Management*, 31(1), 259-280.
- [8] Yıldız, T. K., Yurtay, N., and Öneç, B. (2021). Classifying anaemia types using artificial learning methods. *Engineering Science and Technology, an International Journal*, 24(1), 50-70.
- [9] Zhu, J., Sun, R., Liu, H., Wang, T., Cai, L., Chen, Z., and Heng, B. (2023). A Non-Invasive Hemoglobin Detection Device Based on Multispectral Photoplethysmography. *Biosensors*, 14(1), 22.
- [10] Raza, A., Eid, F., Montero, E. C., Noya, I. D., and Ashraf, I. (2024). Enhanced interpretable thyroid disease diagnosis by leveraging synthetic oversampling and machine learning models. *BMC Medical Informatics and Decision Making*, 24(1), 364
- [11] Bakır, H., and Ceviz, Ö. (2024). Empirical enhancement of intrusion detection systems: a comprehensive approach with genetic algorithm-based hyperparameter tuning and hybrid feature selection. *Arabian Journal for Science and Engineering*, 49(9), 13025-13043.
- [12] Bhatia, M., Meena, B., Rath, V. K., Tiwari, P., Jaiswal, A. K., Ansari, S. M., and Martinen, P. (2023). A novel deep learning-based model for erythrocytes classification and quantification in sickle cell disease.
- [13] Srivastava, S., Srinivasan, R., Nambison, N. K., and Gorthi, S. S. (2021). Diagnosis of sickle cell anaemia using AutoML on UV-Vis absorbance spectroscopy data.
- [14] Dada, E. G., Oyewola, D. O., and Joseph, S. B. (2022). Deep convolutional neural network model for detection of sickle cell anaemia in peripheral blood images. *Communication in Physical Sciences*, 8(1).
- [15] Zemariam, A. B., Yimer, A., Abebe, G. K., Wondie, W. T., Abate, B. B., Alamaw, A. W., Yilak, G., Melaku, T. M., and Ngusie, H. S. (2024). Employing supervised machine learning algorithms for classification and prediction of anaemia among youth girls in Ethiopia. *Scientific reports*, 14(1), 9080.

[16] Ramzan, M., Sheng, J., Saeed, and M.U. (2024).. Revolutionizing anaemia detection: integrative machine learning models and advanced attention mechanisms. Vis. Comput. Ind. Biomed. Art 7, 18.

8. Acknowledgement

I wish to express my sincere gratitude to the faculty and staff of the Department of Computer Science, University of Ibadan, for their mentorship. My heartfelt appreciation goes to DATICAN (www.datican.org) for their generous financial support, which played a pivotal role in enabling the successful execution of this research. Finally, I acknowledge the use of publicly available data from the Kaggle platform, which served as a critical foundation for the data-driven analyses conducted in this study.

Session 10: Cyber Security and Secured Transactions

Title: Access Control System For Financial Institutions Using RBAC.

Authors: Olatoun Deborah Adeyanju, Olaniyi Abiodun Ayeni

Title: CryptoTrust Framework: Asymmetric Encryption-Integrated Zero Trust Architecture

Authors: A.O. Oronti, B.K. Alese, O. Olabode, O.A. Akinsowon

Title: Mitigating SSRF Attacks Through a Zero Trust Security Model

Authors: Olaibi David, Yetunde O. Ogunlola, B.K. Alese

Access Control System For Financial Institutions Using RBAC

Olatoun Deborah Adeyanju¹, Olaniyi Abiodun Ayeni²

Department of Information System Technology, Kings University Ode-Omu¹

*Department Of Cyber Security, School Of Computing, Federal University Of Technology²
Nigeria*

Abstract

The increasing threats to sensitive financial data and digital infrastructures have made access control a core security concern in financial institutions. Role-Based Access Control (RBAC) offers a scalable and efficient method for ensuring that users have access strictly according to their assigned roles. This paper presents the design and implementation of an access control system using RBAC principles to strengthen security protocols in financial systems. The model incorporates Multi-Factor Authentication (MFA) and provides audit-ready access logging. Our implementation confirms that RBAC, coupled with MFA, significantly enhances access control efficiency, reduces insider threats, and ensures compliance with data governance regulations.

Keywords: RBAC, access control, financial institutions, data security, regulatory compliance, role-based security

1. Introduction

Financial institutions are increasingly vulnerable to data breaches due to the high value of sensitive client data and monetary transactions. Between 2020 and 2024, cybersecurity incidents in the financial sector increased by over 35% globally, often caused by poor access management and insider threats (Kumar et al., 2021; ENISA, 2023). Traditional access models such as Discretionary Access Control (DAC) or Mandatory Access Control (MAC) are inadequate for modern dynamic environments. In contrast, Role-Based Access Control (RBAC) defines access based on organizational roles, allowing institutions to enforce the principle of least privilege and reduce administrative overhead (NIST, 2021). RBAC not only aligns well with regulatory frameworks such as GDPR, PCI-DSS, and Nigeria Data Protection Regulation (NDPR) but also provides a manageable framework for identity governance in multi-layered environments.

Role-Based Access Control (RBAC) addresses these concerns by granting system permissions based

on users' job functions within an organization. Instead of assigning rights to each user individually, RBAC allows administrators to define roles with specific permissions, which are then assigned to users based on their duties.

2. Literature Review

Recent studies have explored different access control models and their effectiveness in secure environments. [1] developed a context-aware RBAC system integrated with policy engines for the financial domain, achieving dynamic access restrictions. [5] emphasized the importance of coupling RBAC with MFA in banking apps to prevent credential compromise. [4] compared RBAC to Attribute-Based Access Control (ABAC) and found RBAC easier to implement in regulated sectors like finance. [2] identified poor access control as one of the top three causes of financial data breaches. [3] implemented a Python-based access control simulation using RBAC principles to manage system users in a simulated banking system, reporting a reduction in unauthorized logins by 43%.

3. Methodology

This system was designed using the RBAC security model, where permissions are assigned to roles rather than individuals, and users are assigned one or more roles. Each role is mapped to specific job functions within a typical financial institution such as "Teller", "Loan Officer", "Auditor", and "System Administrator".

3.1. Tools and Technologies

- a) Python 3.11.2: Used for backend logic.
- b) Tkinter GUI: For interface design.
- c) SQLite: For storing role and user mappings.
- d) PyOTP and QRCode: For MFA (OTP-based).
- e) BCrypt: For password hashing.

3.2. System Modules

- The system contains the following functional modules:
- User Registration & Login: Only registered users with assigned roles can log in.
- Role Management Module: Admins define roles and permissions.
- MFA Module: Adds a second layer of authentication using one-timepasscodes.
- Access Request Logger: Logs access requests for audit purposes.

3.3. Diagram Explanation

The methodology diagram illustrates the user interaction from login to authorization:

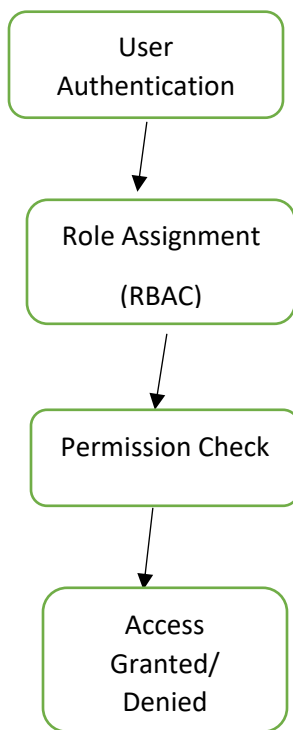


Figure 1. A diagram showing the methodology

1. User Authentication: A user provides login credentials and a one-time password.
2. Role Assignment Validation: The system checks if the role is active and retrieves associated permissions
3. Permission Check: The requested action is compared with role permissions.
4. Access Granted/Denied: Based on result, the action is approved or blocked.

Each interaction is logged with timestamps, role details, and IP addresses for accountability and future auditing.

4. System Implementation

4.1 Interface Design

The user interface was built using the Tkinter module in Python and supports modular functionality for:

- Admin Console: To create roles and assign permissions.
- User Dashboard: Restricted views based on assigned roles.
- MFA Prompt: Users must complete MFA after valid password login.

4.2 Implementation Steps

1. System Start-Up: Python script initializes server and role database.
2. User Registration: Admin creates user profiles and assigns roles.
3. Login Phase: Users provide credentials and OTP.
4. Authorization: The system validates permissions from role table before granting access.
5. Audit Logs: Every access attempt (successful or denied) is logged with role and timestamp.

5. Result Analysis

A series of tests were conducted with simulated users across multiple roles. The system showed:

- Authentication Success Rate: 100% for valid MFA-protected users.
- Access Denial Accuracy: 97% success in denying unauthorized role-based requests.
- Audit Capability: All transactions correctly logged and time-stamped.
- System Efficiency: Average response time was under 2 seconds.

6. Challenges and Best Practices

6.1 Challenges

- Role Explosion: Over-defining roles can complicate management.
- Inaccurate Assignments: Misalignments between roles and permissions can introduce vulnerabilities.
- Change Management: Evolving job roles require frequent role-permission updates.

6.2 Best Practices

Implement role hierarchies to reduce redundancy.
Automate role provisioning through IAM systems.
Perform quarterly access reviews. Train staff on access control policies and responsibilities.

7. Conclusion

This paper presented a practical and secure implementation of an access control system using RBAC for financial institutions. The integration of MFA ensures that even in cases of credential compromise, unauthorized access is mitigated. The system promotes scalability, compliance, and operational efficiency. Future work will incorporate ABAC for hybrid role-attribute-based models and real-time behavioral monitoring for anomaly detection.

8. References

- [1] Ahmed, R., Hassan, M., & Khan, I. (2021). Context-Aware Role-Based Access Control for Financial Systems. *Journal of Computer Security*, 29(1), 115–132. DOI: 10.1016/j.jocs.2021.101670.
- [2] ENISA. (2023). Threat Landscape for Financial Institutions (2023). <https://www.enisa.europa.eu/publications>.
- [3] Fayemi, J., & Obi, K. (2024). Python-Based Simulation of RBAC in Financial Applications. *Journal of Financial Technology Security*, 6(2), 89–101.
- [4] Gandotra, E., Sharma, R., & Choudhury, S. (2022). Comparative Study of RBAC and ABAC in Banking Environments. *International Journal of Information Security*, 13(4), 321–336.
- [5] Mendoza, H. & Singh, R. (2023). Multi-Factor Authentication in Financial Apps: A Usability and Security Evaluation. *IEEE Transactions on Information Forensics*, 18(6), 34-45.
- [6] NIST. (2021). Guide to Role-Based Access Control (SP 800-162). National Institute of Standards and Technology. DOI: 10.6028/NIST.SP.800-162.

CryptoTrust Framework: Asymmetric Encryption-Integrated Zero Trust Architecture

A.O. Oronti, B.K. Alese, O. Olabode, O.A. Akinsowon
Cyber Security Department, School of Computing
Federal University of Technology, Akure
Nigeria

Abstract

This research examines the intersection between asymmetric cryptography and Zero Trust Architecture, analyzing how cryptographic mechanisms strengthen the core objectives of Zero Trust security. The Zero Trust paradigm addresses modern cyber threats by eliminating assumptions of inherent trust and requiring continuous verification. Asymmetric cryptography, recognized for enabling secure communication, validating identities, and ensuring data integrity, is essential to achieving these objectives. Within Zero Trust environments, the application of asymmetric cryptography primarily supports authentication and data protection. For authentication, it operates as a foundational element of Public Key Infrastructure (PKI), using digital signatures to verify users and devices. For data protection, particularly for traffic traversing untrusted networks, it facilitates secure key exchange processes that generate the symmetric keys used for encryption. Its integration reinforces critical Zero Trust tenets by enabling continuous cryptographic verification of entities, protecting communications across micro-segmented networks, and preserving the integrity of policy enforcement. The research also outlines architectural considerations, including strategic placement of cryptographic controls across network layers and their interaction with identity systems, policy engines, and access control mechanisms. Preliminary analysis indicates stronger resilience against advanced persistent threats, lateral movement, and insider attacks through the deployment of pervasive cryptographic trust layers.

Keywords: zero trust; asymmetric, encryption; PKI; cryptography

1. Introduction

The CryptoTrust Framework is a next-generation security model that merges Zero Trust principles with asymmetric cryptography to create a dynamic, cryptographically enforced security architecture. The CryptoTrust Framework Zero Trust Architecture with Asymmetric Encryption represents a modern security framework that combines the principles of Zero Trust with advanced cryptographic techniques to create a robust, dynamic defense system. Unlike traditional security models that rely on perimeter-based defenses, this architecture assumes that threats can originate from anywhere, both outside and inside the network.

By integrating asymmetric encryption into Zero Trust, the framework ensures that every access request is cryptographically verified, minimizing the risk of unauthorized access and lateral movement by attackers. This approach is particularly effective in today's distributed environments, where users, devices, and services interact across cloud, on-premises, and hybrid infrastructures. Zero Trust represents a paradigm shift from implicit trust models to explicit verification frameworks where every access request is authenticated, authorized, and continuously validated regardless of the user's location or previous authentication status [1],[6]. Zero Trust operates on the principle of "never trust, always verify," requiring continuous authentication and least-privilege access controls that limit user and device access to only the specific resources necessary for their legitimate functions [5],[7]. This approach significantly reduces the attack surface by eliminating implicit trust relationships and requiring explicit authorization for every network transaction [3], [4].

2. Literature Review

2.1. Traditional Network Security Models

Network security has been fundamentally shaped by the perimeter security paradigm, which dominated cybersecurity frameworks for decades. Traditional network security models operated on the premise that threats originated exclusively from external sources, while internal network resources remained inherently trustworthy. Recent scholars outlined significant limitations in this approach. [8] argued that institutional perimeters no longer correspond exclusively to physical boundaries, and internal network environments cannot be presumed to provide secure hosting environments for enterprise applications. The proliferation of mobile workforces, cloud computing adoption, and diverse computing devices has fundamentally challenged the validity of perimeter-based security assumptions, rendering traditional models increasingly obsolete in contemporary distributed computing environments.

2.2. Zero Trust Architecture: Evolution and Principles

Zero Trust architecture represents a paradigmatic shift from traditional perimeter-based security models,

operating on the fundamental principle of "never trust, always verify." The proliferation of cloud computing, mobile device use, and the Internet of Things has dissolved conventional network boundaries, necessitating new security frameworks that can accommodate distributed workforces and ubiquitous connectivity requirements. [2] has provided comprehensive guidance on Zero Trust implementation through NIST Special Publication 800-207, which establishes technical standards defining the architectural components and operational principles essential for effective deployment. Zero Trust (ZT) is the term for an evolving set of cybersecurity paradigms that move defenses from static, network-based perimeters to focus on users, assets, and resources [1]. One of the basic tenets of zero trust is to remove the implicit trust in users, services, and devices based only on their network location, affiliation, and ownership [9].

2.3. Asymmetric Encryption: Theoretical Foundations and Applications

Asymmetric encryption, or public-key cryptography, fundamentally advances cryptographic technology by resolving symmetric encryption's key distribution challenges. This method employs distinct public and private keys for data encryption and decryption, eliminating the need for secure key exchange channels. Its mathematical foundations rely on computationally difficult problems, creating one-way functions. The RSA algorithm [10], developed in 1978, exemplifies this by leveraging the difficulty of factoring large prime numbers. More advanced implementations, like Elliptic Curve Cryptography (ECC), offer superior efficiency and security. ECC provides equivalent security with smaller key sizes, making it ideal for resource-constrained environments such as mobile and IoT systems, where computational efficiency is crucial. The practical deployment of asymmetric encryption heavily depends on Public Key Infrastructure (PKI) frameworks. PKI provides essential infrastructure for key generation, distribution, validation, and lifecycle management, establishing trust through Certificate Authorities (CAs) that verify public key authenticity and associated identities for large-scale deployments.

2.4. Integration Studies and Convergence

The integration of Zero Trust architecture with asymmetric encryption is an emerging cybersecurity research area. Despite limited comprehensive studies, their convergence has strong theoretical foundations. The mathematical properties of public-key cryptography enable cryptographically verifiable identity tokens, crucial for Zero Trust policy enforcement. Contemporary security frameworks

increasingly emphasize integrating multiple technologies to combat complex threats. This integration is vital for embedding resilience into every organizational facet, combining architectural principles with robust cryptographic implementations to enhance overall security.

2.5. Cryptographic Foundations of Zero Trust Architecture

Evaluating Zero Trust architectures using asymmetric cryptography plays a foundational role in ensuring robust authentication and secure data exchange [16]. In the domain of authentication, asymmetric cryptography underpins Public Key Infrastructure (PKI) by enabling the use of digital signatures to validate the identities of users, devices, and services. This mechanism ensures that entities participating in a communication process can be reliably authenticated without relying on implicit trust or static network boundaries. On the data protection front, asymmetric cryptography is instrumental in establishing secure key exchange protocols that negotiate ephemeral symmetric keys for high-performance data encryption. This hybrid model combines the strong security guarantees of public-private key pairs with the computational efficiency of symmetric cryptography for bulk data handling. The effective deployment of asymmetric cryptographic mechanisms within a Zero Trust environment requires careful consideration of factors such as algorithmic security strength, computational overhead, and appropriate key sizes, all of which influence system performance and long-term security resilience. Equally critical is the establishment of comprehensive key management practices, including secure key generation, distribution, storage, rotation, and revocation. These processes are essential for maintaining the integrity and trustworthiness of the cryptographic ecosystem and mitigating risks associated with key compromise or misuse. Furthermore, integrating asymmetric cryptography with Identity and Access Management (IAM) systems is vital for achieving seamless and scalable authentication and authorization workflows. Such integration enables continuous verification of identity and device posture, enforces least-privilege principles, and supports dynamic policy decisions—core tenets of the Zero Trust paradigm. Together, these cryptographic and identity management capabilities form a cohesive framework that strengthens organizational security posture in highly distributed and adversarial environments. Recent research demonstrates that asymmetric cryptography seamlessly supports Zero Trust primitives across enterprise networks, cloud environments, and heterogeneous device landscapes [17], [18]. By ensuring that every access request is cryptographically verified, organizations achieve

what the Cloud Security Alliance defines as “an industry-standard, secure method to establish identity, authenticate entities, maintain data integrity, and control access without requiring any explicit trust” [2]. Building on this foundation, the emerging CryptoTrust Framework integrates Zero Trust principles with asymmetric encryption to construct a dynamically enforced, cryptographically anchored security architecture. Unlike traditional perimeter-based models that assume threats originate externally, this framework acknowledges that adversaries may operate from any location within a network [12]. Through continuous cryptographic verification, the model mitigates unauthorized access and limits lateral movement an essential capability in today’s highly distributed environments spanning cloud, on-premises, and hybrid infrastructures [11]. Zero Trust’s core philosophy of “never trust, always verify” reinforces least-privilege access control, requiring explicit authorization for every transaction and thereby eliminating implicit trust relationships [13] [14]. This comprehensive approach significantly reduces the attack surface and strengthens defenses against both internal and external threats [15], [3].

2.6. Current Literature Gaps

Despite extensive research on Zero Trust architecture and asymmetric encryption independently, a significant gap exists in studies addressing their integrated application for network security. Current literature lacks comprehensive analysis of their synergistic combination, empirical evaluations of their combined effectiveness and performance, and detailed insights into practical implementation challenges. Furthermore, guidance on optimizing these integrated deployments for security and efficiency is insufficient. These unaddressed areas present crucial opportunities for advancing cybersecurity knowledge and providing practical solutions for organizations seeking enhanced security postures through advanced architectural and cryptographic approaches.

3. Design Model

The CryptoTrust Zero Trust Architecture (ZTA) redefines cybersecurity by replacing traditional password-based systems with a robust, multi-layered framework centered on cryptographic identities, dynamic access controls, and secure communications. At its core, the Identity Layer uses Public Key Infrastructure (PKI) to issue X.509 certificates to users, devices, and services, ensuring verifiable identities. Short-lived certificates, stored securely in Hardware Security Modules (HSMs) or Trusted Platform Modules (TPMs), minimize risks of credential theft, eliminating vulnerabilities like

phishing inherent in password systems. For instance, a user’s private key, safeguarded in an HSM, authenticates their identity, validated by a public key, ensuring only trusted entities gain access.

The Policy Layer enforces least-privilege access through a Policy Decision Point (PDP) and Policy Enforcement Point (PEP). The PDP evaluates requests based on user roles, device health, and contextual factors like time or location, issuing time-bound, scoped tokens (e.g., JWT or OAuth) via the PEP. A DevOps accessing a Kubernetes cluster might receive a five-minute token restricted to specific resources, reducing the attack surface and limiting damage from compromised accounts.

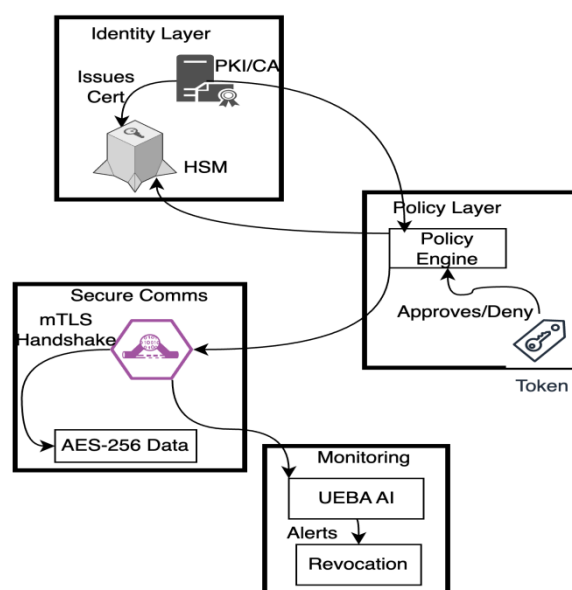


Figure 1. CryptoTrust Zero Trust Architecture with Asymmetric Encryption

The Secure Communication Layer ensures end-to-end encryption using mutual TLS (mTLS), where clients and servers authenticate via certificates. Elliptic Curve Diffie-Hellman Ephemeral (ECDHE) generates one-time session keys for forward secrecy, while AES-256 handles bulk data encryption. Post-quantum algorithms like CRYSTALS-Kyber prepare the system for future threats. The architecture aligns with the OSI model, securing layers from physical (HSMs/TPMs) to transport (TLS 1.3, WireGuard/IPsec). Continuous monitoring via AI-driven User and Entity Behavior Analytics (UEBA) detects anomalies, with real-time certificate revocation through OCSP stapling. This framework excels in securing remote workforces, cloud-native applications, and critical infrastructure, offering scalability, micro-segmentation, and resilience against evolving threats, making it a future-proof solution for diverse environments.

4. Conclusion

The CryptoTrust Framework model merges the rigor of Zero Trust with the robustness of asymmetric encryption, creating a cryptographically verifiable security model. By eliminating implicit trust and leveraging PKI, it addresses modern threats while enabling scalable, adaptive protection for evolving infrastructures. Future advancements will focus on quantum-safe cryptography and autonomous policy orchestration to further harden the framework.

5. References

- [1] Rose, S., Borchert, O., Mitchell, S., and Connelly, S. (2020). Zero trust architecture (NIST Special Publication 800-207). National Institute of Standards and Technology. DOI: 10.6028/NIST.SP.800-207.
- [2] National Institute of Standards and Technology. (2023). Zero trust architecture model for access control in cloud-native applications in multi-cloud environments (NIST Special Publication 800-207A). U.S. Department of Commerce.
- [3] Ahmad, I., Shahabuddin, S., Kumar, T., Okwuibe, J., Gurtov, A., and Ylianttila, M. (2021). Security for 5G and beyond. *IEEE Communications Surveys and Tutorials*, 23(2), 1682-1740.
- [4] Buck, C., Olenberger, C., Schweizer, A., Völter, F., and Eymann, T. (2021). Never trust, always verify: A multivocal literature review on current knowledge and research gaps of zero-trust. *Computers and Security*, 110, 102436. Stafford V.A. Zero trust architecture, NIST special publication. 2020;207.
- [5] Gilman, E., and Barth, D. (2017). Zero trust networks: Building secure systems in untrusted networks. O'Reilly Media.
- [6] Kindervag, J. (2010). No more chewy centers: Introducing the zero trust model of information security. Forrester Research.
- [7] Ward, R., and Beyer, B. (2014). BeyondCorp: A new approach to enterprise security. *The USENIX Magazine*, 39(6), 6-11. GSA, "Zero Trust Architecture (ZTA): Buyer guide," 2021. [Online]. Available: Zero Trust Architecture Buyers Guide v11 20210810 (4).pdf.
- [8] Rui, Zheng, Ma, H., Wang, Q., Fu, J., and Jiang, Z. (2021). Assessing the Security of Campus Networks: The Case of Seven Universities. *Sensors*, 21(306).
- [9] Rose, S., Borchert, O., Mitchell, S., and Connelly, S. (2023). A zero trust architecture model for access control in cloud-native applications in multi-cloud environments (NIST Special Publication 800-207A). National Institute of Standards and Technology. DOI: 10.6028/NIST.SP.800-207A.
- [10] Rivest, R. L., Shamir, A., and Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2), 120-126. DOI: 10.1145/359340.359342.
- [11] Naik, N., and Jenkins, P. (2020). Your digital identity or your privacy: A trade-off for authentication methods for building information modelling in the cloud. In *2020 IEEE Cloud Summit* (pp. 99-105).
- [12] Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Capretz, L. F., Abdulkadir, S. J., and Alshammari, G. (2025). Zero trust architecture: A systematic literature review. *arXiv preprint arXiv:2503.02089*.
- [13] Ward, R., and Beyer, B. (2014). BeyondCorp: A new approach to enterprise security. Google Technical Report.
- [14] Gilman, E., and Barth, D. (2017). Zero trust networks: Building secure systems in untrusted networks. O'Reilly Media.
- [15] Buck, C., Olenberger, C., Schweizer, A., Völter, F., and Eymann, T. (2021). Never trust, always verify: A multivocal literature review on current knowledge and research gaps of zero-trust. *Computers & Security*, 110, 102436.
- [16] Cloud Security Alliance. Using Asymmetric Cryptography to Help Achieve Zero Trust Objectives. <https://cloudsecurityalliance.org/artifacts/using-asymmetric-cryptography-to-help-achieve-zero-trust-objectives>. (Access Date: 11/05/2024).
- [17] Alshahrani, H., Alotaibi, M., Ansari, M. T. J., Asiri, M. M., Agrawal, A., Khan, R. A., Mohsen, H., and Hilal, A.M. (2024). Dissecting zero trust: Research landscape and its implementation in IoT. *Cybersecurity*, 7(1), 26.
- [18] Cunningham, S., Enright, C., Friel, O., and Miskelly, G.M. (2024). Zero trust cybersecurity: Procedures and considerations in context. *Applied Sciences*, 14(19), 8948.
- [19] National Institute of Standards and Technology. (2020). Zero trust architecture (NIST Special Publication 800-207). U.S. Department of Commerce.

Mitigating SSRF Attacks Through a Zero Trust Security Model

Olaibi David, Yetunde O. Ogunlola, B.K. Alese

Department of Cyber Security, School of Computing, Federal University of Technology, Akure, Nigeria

Abstract

Server-Side Request Forgery (SSRF) has also turned out to be a quite crucial web app vulnerability that enables attackers to induce servers into making unauthorized requests towards interior services. Such assaults transcend conventional perimeter defense mechanisms and misuse implicit trust in interior traffic with a potential leakage of secretive resources. Our findings herein craft a Zero Trust Security Framework specially designed as a solution towards blocking SSRF vulnerabilities. The architecture integrates core Zero Trust Architecture (ZTA) principles—continuous verification, least privilege access, and microsegmentation—in a multi-layer protection model. A web application built with a purposively included SSRF vulnerability became the testing ground. NGINX became a Policy Decision Point (PDP)/Policy Enforcement Point (PEP) and used to filter incoming requests and reject access to prohibited IP ranges and administrative paths. Input validation and URLs whitelisting were done at the app level as a blocking of untrusted requests measure. In order to bolster identity assurance, a Multi-Factor Authentication (MFA) system based on PyOTP was implemented with password authentication and a time-based one-time password (TOTP). Security monitoring and observability were also made possible with Wazuh, receiving and examining logs in real time and verifying anomalies like SSRF attempts and failed MFA logins. The system executed within an isolated virtual environment with VirtualBox emulating real-world service isolation. Fuzzing with various SSRF payloads—like internal IP targeting, redirect chains, and other IP notations—confirmed that the framework successfully blocked malicious requests and reduced the attack surface. The results demonstrate that Zero Trust is a solid and feasible solution for shielding web applications from SSRF and other internal threat vectors.

Keywords: Zero Trust Architecture (ZTA), Server-Side Request Forgery (SSRF), Web Application Security, Access Control

1. Introduction

Server-Side Request Forgery (SSRF) has been growing as a serious concern in web security. It occurs when attackers trick a vulnerable server into making requests it shouldn't, often targeting internal systems that are normally hidden from outside access. These kinds of attacks can cause data leaks, system compromises, lateral movement within a network, and expose sensitive internal services. What makes SSRF particularly dangerous is that it can bypass security measures like firewalls and VPNs that are designed to trust traffic coming from inside the network [1]. This rising threat shows that traditional security models, which tend to trust everything inside the network, are no longer enough. High-profile breaches at companies like Capital One and Tesla emphasize how serious an SSRF vulnerability can be and the real impact it could have on big organizations [2]. As a result, there's an urgent need to adopt security strategies that do not automatically trust any user or device, whether inside or outside the network. The main idea of Zero Trust, "never trust, always verify", is to verify everyone and everything continuously, no matter where they are, before granting access [3]. Instead of trusting internal entities automatically, Zero Trust uses strict access controls, detailed authentication, and network segmentation to reduce what attackers can exploit [4]. Using Zero Trust principles to strengthen web application security, especially for blocking SSRF attacks, offers a proactive way to defend systems.

by default. As shown in Figure 1, the NIST zero trust framework combines both the policy enforcement point and policy decision point for proper access decisions. Some major tenets of ZTA include identity-driven access control, least-privilege enforcement, and dynamic policy evaluation [1]. Researchers explored applying it to insider threat defense [14] strengthening cloud identity [12] and facilitating enterprise-extended continuous authentication [13]. [5] also pinpointed microsegmentation as one vital defense for lateral movement with special relevance to SSRF confinement.

2. The Risk of SSRF

Server-Side Request Forgery (SSRF) is today one of the highest-priority vulnerabilities in modern web applications, prompting attackers to deceive servers into placing unauthorized requests to internal services. Exploits of this kind have been found to expose sensitive resources, authorize lateral movement, and dismantle entire infrastructures [2]. Case studies identified SSRF as one of the underlying causes of high-profile cloud breaches involving cloud access to cloud metadata services [10]. Researchers have outlined numerous methodologies for detection and mitigation such as taint tracking in PHP applications [8], deep learning-assisted anomaly detection [9], and static URL pattern analysis [15]. More recently, [16] have detailed URL-based fuzzing methodologies by automated means for large-scale SSRF discovery manifesting later exploit methodologies. Reverse proxies have been deployed to shield internal resources [7], yet SSRF continues to be one of the gravest application-layer risks evading perimeter defenses.

3. The Zero Trust Paradigm

Zero Trust Architecture (ZTA) fundamentally reconsiders security by stripping away implicit trust from networks. Originally characterized by [3], Zero Trust subjects all traffic to constant verification, stringent access control, and microsegmentation in order to reduce attack surfaces. The U.S. National Institute of Standards and Technology (NIST) enshrined such ideals in SP 800-207 by emphasizing never to trust by default any actor—either internal or external [4]. Some major tenets of ZTA include identity-driven access control, least-privilege enforcement, and dynamic policy evaluation [1]. Researchers explored applying it to insider threat defense [14] strengthening cloud identity [12] and facilitating enterprise-extended continuous authentication [13]. [5] also pinpointed microsegmentation as one vital defense for lateral movement with special relevance to SSRF confinement.

3.1. Zero Trust in Cloud and Distributed Environments

With migrating entities utilizing distributed infrastructures, Zero Trust has emerged to prominence in addressing cloud-native and multi-location complexities. [11] examined ZTA adoption in cloud networks both in terms of opportunities and integration complexities. Similarly, [17] conducted a survey of migration choices, with strong focus on legacy system integration complexities with Zero Trust. [18] followed by proposing Zero Trust models in multi-location cloud-native contexts, with [19] proposing practical

methodologies for implementation at enterprise scales. [20] undertook a general overarching survey of Zero Trust adoption patterns indicative of rising use in organizational protective measures. Collectively, these writings mirror Zero Trust coming out of theory to general usage in practice contexts where cloud-driven environments pose heightened SSRF risks.

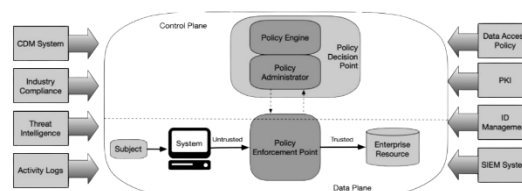


Figure 1. Nist zero trust framework

3.2. Zero Trust as a Framework for SSRF Mitigation

Whereas point-level SSRF prevention techniques such as reverse proxies [7], taint analysis [8], and anomalous detection models [9] act in isolation, Zero Trust provides a comprehensive security model in line with such measures that must be enforced to successfully counter SSRF. Through enforcing perpetual authentication [6][13], microsegmentation [5], and dynamic policy enforcement [4], ZTA eliminates implicit trust upon which SSRF attacks place their reliance. Going one step further, incorporating automated SSRF detection techniques [16][15] along with Zero Trust monitoring and enforcement solutions yields multi-tiered defenses capable not only of preventing but also containing such attacks. This incorporation offers one such avenue both for research and practice.

4. Related Works

[7] introduced a prototype that employs a helper server to fetch external resources, which helps prevent the web application server from directly accessing internal services. However, their system didn't fully cover more sophisticated evasion techniques or out-of-band SSRF attacks. [4] presented the official NIST framework for Zero Trust Architecture, outlining key parts like Policy Engines, Policy Enforcement Points, and trust algorithms. Their approach allows for dynamic policy enforcement and continuous validation, helping to limit how far attackers can move within a network. [8] developed a dynamic analysis tool using taint tracking to find SSRF vulnerabilities in PHP apps. It spots "honeypot" functions and creates payloads to test for SSRF risks. [9] used deep learning, especially LSTM neural networks, to detect SSRF in URLs. Their trained

model was pretty accurate, offering a more flexible method compared to traditional signature-based detection. [10] looked into how SSRF attacks can be spotted in AWS environments, evaluating tools like VPC Traffic Reflecting, Zeek, and Suricata. His research showed that standard network monitoring tools often miss certain SSRF techniques, especially when accessing local metadata. [11] focused on mathematical models for Zero Trust using Bayesian Networks, Game Theory, and Markov Chains. His work emphasizes how ZTA can help restrict lateral movement and improve identity management in cloud setups. [12] proposed a Zero Trust Identity Management Model (ZTIMM) for volunteer cloud systems. This model emphasizes trust scoring, re-verification based on behavior, and lifecycle management for virtual nodes. [13] discussed Zero Trust strategies like micro-segmentation, encryption, and trust evaluation using fuzzy logic. They stressed the importance of ongoing verification in access controls, but didn't include practical case studies. [14] presented a Zero Trust framework for enterprise networks using EVE-NG. Their setup included group policies and user access controls designed to reduce insider threats.

5. System Design

The system was built around Zero Trust Architecture (ZTA) principles with a focus on making web applications more secure, especially against Server-Side Request Forgery (SSRF) attacks. The design emphasizes removing any assumptions of trust by layering security measures across the network, application, and identity layers. Every component was chosen to serve specific Zero Trust goals such as ongoing verification, giving only the minimum necessary permissions, segmenting the system into micro zones, and enforcing policies at every step. At the core of this setup is a simple, modular web app built with Flask. It's designed for easy integration and includes features like user login, product display, and admin functions. To test the security measures, a deliberate SSRF vulnerability was added via a `/fetch` route that takes user input URLs and fetches their content. This route is the main point of attack simulation to see how well the system holds up. The app also uses PyOTP for Multi-Factor Authentication (MFA), making sure user sessions are checked regularly, which helps keep unauthorized users out even if their main credentials are stolen. To guard the app against threats from both outside and inside, a NGINX reverse proxy sits in front of the Flask server. NGINX acts as both a Policy Enforcement Point (PEP) and Policy Decision Point (PDP) as shown in Figure 2. It checks all incoming requests, blocking things like access to internal IPs, loopback addresses, and protected routes such as `/admin`, cutting down SSRF risks right at the entry point.

NGINX also manages request routing and load balancing, serving as the first line of defense in the Zero Trust model.

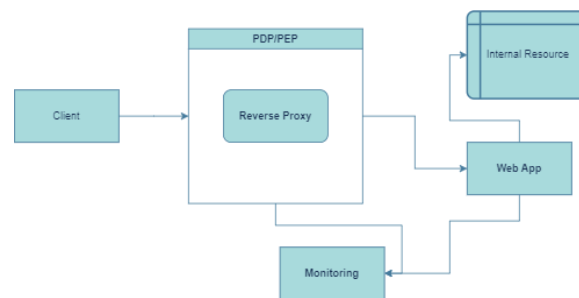


Figure 2. System Architecture, demonstrating how Flask, NGINX, and Wazuh collaborate to mitigate SSRF attacks in a Zero Trust framework

For microsegmentation, the entire architecture is split into different network zones using virtual machines in VirtualBox. There are three main VMs: one runs the web app, another handles the reverse proxy, and the third is for monitoring and analytics. Each VM uses a bridged network to imitate separate communication channels. This setup limits lateral movement if one component gets compromised, others are protected by both network controls and application-level restrictions. The system's monitoring and observability layer is handled by Wazuh, an open-source host-based intrusion detection system. Deployed on a dedicated VM, Wazuh gathers logs from the NGINX proxy and the web app, analyzing them in real-time to spot suspicious activity, policy breaches, SSRF attempts, failed logins, and other anomalies. It generates alerts based on preset rules, giving continuous updates to help with incident response. Another key part of the setup is thorough input validation and policy enforcement at both the network and application levels. NGINX filters out potentially harmful requests before they reach the app. Meanwhile, the Flask app also internally checks all URLs entered into the `/fetch` route, making sure they're either on a safe whitelist or don't match known malicious patterns.

6. Implementation

The deployment of the Zero Trust Security Framework was conducted within a controlled virtual setup using VirtualBox to emulate network segmentation and simulate real-world deployment scenarios. Three distinct virtual machines (VMs) were established to represent separate architectural components: the Flask web application, the NGINX reverse proxy server, and the Wazuh monitoring system. This configuration assisted clear separation between

application logic, access control mechanisms, and security monitoring, effectively representing the core concept of microsegmentation integral to Zero Trust Architecture (ZTA). The Flask-based web application was developed with fundamental login and product listing functionalities, intentionally incorporating a Server Side Request Forgery (SSRF) vulnerability within the `/fetch` route. This endpoint permitted users to enter URLs, which the server would attempt to retrieve, mimicking how attackers often exploit SSRF to gain access to internal services. Flask was also set up to handle user authentication, integrating PyOTP to enforce Multi-Factor Authentication (MFA). Every user was required to verify their identity through a password and a TOTP generated by an authenticator app like Google Authenticator. The shared secret used for TOTP was securely stored and presented via a QR code during initial setup. The NGINX reverse proxy was positioned in front of the Flask application, serving as both a Policy Enforcement Point (PEP) and Policy Decision Point (PDP). NGINX was configured to intercept all incoming HTTP requests, analyze request URLs, and enforce strict routing policies. These policies included blocking requests to local loopback addresses (`127.0.0.1`), link-local metadata services (`169.254.169.254`), and administrative paths such as `/admin`. Custom rules were carefully crafted to prevent common SSRF attack vectors, including alternate IP notation formats (e.g., decimal or hexadecimal representations) and URL redirection tactics. This layer of control ensured that potentially malicious or invalid requests never reached the application layer. Simultaneously, the Wazuh monitoring platform was installed on a dedicated VM to provide host-based intrusion detection, log analysis, and alerting. Wazuh was configured to collect logs from both NGINX and Flask, enabling comprehensive monitoring of user activities, detection of failed MFA attempts, and identification of suspicious SSRF behavior. It generated alerts in response to unauthorized access or policy violations, which were logged in a centralized dashboard for easy visibility. Although initial setup employed basic monitoring rules, the Wazuh engine has the capacity for creating custom rules to enable more detailed and refined detection capabilities in future enhancements. Throughout the implementation process, the system underwent strict testing through various SSRF scenarios, including straightforward internal fetches (such as to `/admin` or `http://localhost`), probes to metadata endpoints (`http://169.254.169.254`), and payloads crafted to evade validation mechanisms (like redirect URLs or DNS rebinding). In every case, the combined measures of NGINX routing policies and Flask input validation successfully prevented SSRF attacks. MFA validation also functioned correctly, denying access to users with invalid TOTP codes, while Wazuh reliably logged these

security events.

7. Conclusion and Future Work

This paper discusses how a Zero Trust Security Framework can help protect web applications from Server-Side Request Forgery (SSRF) attacks. The approach combines several security layers, including a Flask app with SSRF simulation, a NGINX reverse proxy that acts as both the policy enforcer and decision-maker, and Multi-Factor Authentication (MFA) powered by PyOTP. Together, these elements restrict access and ensure ongoing verification of users. Wazuh further enhances security by providing real-time insights into suspicious activity and potential threats. All these pieces work together to block unauthorized requests, defend against SSRF techniques like targeting internal IPs or chaining redirects, and follow key Zero Trust principles such as micro-segmentation and the idea of least privilege. Overall, this prototype shows that applying Zero Trust Architecture (ZTA) in real-world web applications can greatly improve security against both internal and perimeter-based threats. Looking ahead, possible improvements include adding role-based access control (RBAC), making the monitoring system more detailed with better logging and alert options, deploying the framework within containers or cloud environments, and integrating machine learning to detect threats more adaptively. These enhancements could make the system more scalable, automated, and dependable for larger, production-level use.

8. References

- [1] Agilicus. (2022). Eliminate Attack Vectors and Stop Cyber Threats in Their Tracks with a Zero Trust. <https://www.agilicus.com/white-papers/protect-against-cyber-threats-zero-trust/> (Access Date: 14th May 2024).
- [2] Khan, S., Kabanov, I., Hua, Y., and Madnick, S. (2022). A systematic analysis of the Capital One data breach: Critical lessons learned. *ACM Transactions on Privacy and Security*, 26(1), Article 3, 1–29. DOI: 10.1145/3546068.
- [3] Kindervag, J. (2010). Build Security into Your Network's DNA: The Zero Trust Network Architecture. Forrester Research.
- [4] Rose, S., Borchert, O., Mitchell, S., and Connelly, S. (2020). Zero Trust Architecture. NIST Special Publication 800-207. DOI: 10.6028/NIST.SP.800-207.
- [5] Habash, R. M., and Ibrahim, M. K. (2023). Zero Trust security model for enterprise networks. *Iraqi Journal of Information and Communication Technology*, 6(2). DOI: 10.31987/ijict.6.2.223.

- [6] Barth, D., and Gilman, E. (2017). *Zero Trust Networks: Building Secure Systems in Untrusted Networks*. O'Reilly Media.
- [7] Jabiyev, B., Mirzaei, O., Kharraz, A., and Kirda, E. (2021). Preventing server-side request forgery attacks. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing (SAC '21)* (pp. 1626–1635). Association for Computing Machinery, New York, NY, USA. DOI: 10.1145/3412841.3442036.
- [8] Yulianton, H., Trisetyarso, A., Suparta, W., Abbas, B. S., and Kang, C. H. (2020). Web application vulnerability detection using taint analysis and black-box testing. *IOP Conference Series: Materials Science and Engineering*, 879(1), 012031. DOI: 10.1088/1757-899X/879/1/012031.
- [9] Al-Talak, K., and Abbass, O. (2021). Detecting Server-Side Request Forgery (SSRF) Attack by using Deep Learning Techniques. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 12(12). DOI: 10.14569/IJACSA.2021.0121230.
- [10] McElroy, S. (2020). Detecting server-side request forgery attacks on Amazon Web Services. 18, 34–41. DOI: 10.5281/zenodo.4781857.
- [11] Ahmadi, S. (2024). Zero trust architecture in cloud networks: Application, challenges and future opportunities. *Journal of Engineering Research and Reports*, 26, 215–228. DOI: 10.9734/JERR/2024/v26i21083.
- [12] Albuali, A., Mengistu, T., and Che, D. (2020). ZTIMM: A zero-trust-based identity management model for volunteer cloud computing. In Q. Zhang, Y. Wang, and L. J. Zhang (Eds.), *Cloud Computing – CLOUD 2020*. CLOUD 2020. *Lecture Notes in Computer Science* (Vol. 12403, pp. 287–294). Springer, Cham. DOI: 10.1007/978-3-030-59635-4_22.
- [13] Syed, N. F., Shah, S. W., Shaghaghi, A., Anwar, A., Baig, Z., and Doss, R. (2022). Zero trust architecture (ZTA): A comprehensive survey. *IEEE Access*, 10, 57143–57179. DOI: 10.1109/ACCESS.2022.3174679.
- [14] Habash, R. M., and Ibrahim, M. K. (2023). Zero trust security model for enterprise networks. *Iraqi Journal of Information and Communication Technology*, 6(2). DOI: 10.31987/ijict.6.2.223.
- [15] Luo, H. (2019). SSRF vulnerability attack and prevention based on PHP. In *2019 International Conference on Communications, Information System and Computer Engineering (CISCE)* (pp. 469–472). Haikou, China. DOI: 10.1109/CISCE.2019.00109.
- [16] Wang, E., Chen, J., Xie, W., Wang, C., Gao, Y., Wang, Z., Duan, H., Liu, Y., and Wang, B. (2024). Where URLs become weapons: Automated discovery of SSRF vulnerabilities in web applications. In *2024 IEEE Symposium on Security and Privacy (SP)* (pp. 239–257). San Francisco, CA, USA. DOI: 10.1109/SP54263.2024.00198.
- [17] Teerakanok, S., Uehara, T., and Inomata, A. (2021). Migrating to zero trust architecture: Reviews and challenges. *Security and Communication Networks*, 2021(1), 9947347. DOI: 10.1155/2021/9947347.
- [18] Chandramouli, R. (2023). A zero trust architecture model for access control in cloud-native applications in multi-cloud environments. DOI: 10.6028/NIST.SP.800-207A.IPD.
- [19] Tsai, M., Lee, S., and Shieh, S. W. (2024). Strategy for implementing of zero trust architecture. *IEEE Transactions on Reliability*, 73(1), 93–100. DOI: 10.1109/TR.2023.3345665.
- [20] Edo, O., Tenebe, I., Etu, E.-E., Ayuwu, A., Emakhu, J., and Adebiyi, S. (2022). Zero trust architecture: Trend and impact on information security. *International Journal of Emerging Technology and Advanced Engineering*, 12, 140–147. DOI: 10.46338/ijetae0722_15.

Session 11: Intelligent Control

Title: Development of an Intelligent Traffic Management System to Address Traffic Signal Visibility Occlusion in Ibadan, Nigeria

Authors: Agbede Caleb Oluwole, Akintayo Folake Olubunmi, Agbede Oluwole Akinyele

Title: Least Significant Bit (LSB) Audio Steganography With AES 256 Encryption

Authors: Olaniyi Abiodun Ayeni, Adekunle Obajimi Isaac

Title: CNN-Based Detection of Image Manipulation

Authors: D. O. Babatunde, O.Y. Ogunlola, B.K. Alese

Development of an Intelligent Traffic Management System to Address Traffic Signal Visibility Occlusion in Ibadan, Nigeria

Agbede Caleb Oluwole, Akintayo Folake Olubunmi, Agbede Oluwole Akinyele
Department Of Civil Engineering, University Of Ibadan, Ibadan, Nigeria

Abstract

Visibility obstructions at urban intersections due to larger vehicles and adverse weather conditions pose significant safety risks and exacerbate traffic congestion. Urban traffic intersections in Nigerian cities like Ibadan often experience visibility obstructions caused by large vehicles, poor road geometry, and adverse weather conditions—factors that impair driver response, increase waiting time, and contribute to congestion. This research proposes an Intelligent Traffic Management System (ITMS) using IoT and GPS technology to enhance visibility and reduce congestion at intersections, specifically in Ibadan, Nigeria. The study focuses on the Agodi Gate corridor, a critical urban intersection, where real-world observational surveys revealed frequent signal occlusions due to vehicle height disparities and limited headway. Through field data collection, VISSIM-based simulation, and mathematical modeling, the study analyzes key parameters such as headway distance, angle of view, vehicle dimensions, and driver response delay. The proposed ITMS prototype provides real-time signal status through dashboard interfaces or mobile applications and dynamically adjusts signal timing based on detected visibility conditions. Results show that visibility-related obstructions significantly impact intersection efficiency and safety. By addressing these challenges, the system enhances driver situational awareness, reduces traffic delays, and improves overall intersection performance. The developed ITMS framework is scalable and offers practical solutions for similar urban environments experiencing visibility-induced traffic challenges.

Keywords: Intelligent Traffic Management System, Visibility Obstruction, Urban Traffic Flow, IoT, GPS.

1. Introduction

Urban intersections face complex traffic management challenges due to vehicle congestion and visibility obstructions, especially in high-density areas like Ibadan. Past research highlighted critical traffic parameters such as headway and traffic flow that are key to mitigating congestion [1]. However, visibility issues due to larger vehicles blocking signals

remain a safety risk and a contributor to delays. This study introduces a targeted ITMS leveraging IoT and GPS for real-time management, focusing on a critical intersection at Agodi Gate.

2. Background

Traffic congestion remains one of the most persistent challenges in rapidly growing urban environments, particularly in developing countries such as Nigeria. Over the years, researchers have examined congestion using fundamental traffic flow parameters, including vehicle headway, density, flow, and capacity. Akintayo [1] emphasized the critical role of headway in understanding traffic flow dynamics and managing urban congestion, especially in cities like Ibadan. Similarly, Akintayo and Agbede [2] modeled headway distribution on two-lane roads and demonstrated how vehicle interactions influence traffic characteristics. However, these studies primarily address vehicular behavior and do not sufficiently consider the effects of visibility on traffic performance.

Classic traffic flow theories from Greenshields [3], Daganzo [4], and Hidas [5] form the basis for understanding vehicular interactions. Yet, visibility obstructions such as those caused by large vehicles, poor road geometry, vegetation, or adverse weather remain inadequately addressed within these models. Salisu et al. [6] drew attention to visibility-related traffic challenges in Nigeria, where inadequate planning and infrastructure shortcomings worsen the problem. Djahel et al. [7] also emphasized inefficiencies arising from poor communication between traffic control devices and approaching vehicles in many developing urban systems.

Visibility obstructions arise from multiple sources. Large vehicles frequently block approaching drivers' views of critical signage and signals, as reported by Al-Kaisy et al. [8]. Additional visibility challenges stem from overgrown roadside vegetation [9], substandard or poorly designed road geometry [10], low illumination during poor lighting or weather conditions [11], and improper sign placement [12,13]. These issues compromise safety and degrade the performance of adaptive signal control systems.

Abdel-Aty et al. [14] noted that while modern systems attempt to incorporate real-time traffic information, their effectiveness diminishes when visibility disruptions are not adequately captured.

Research has also shown that occlusion caused by heavy vehicles significantly reduces a driver's recognition distance and reaction capability. Luo, Zhang, and Zhang [15] and Mahajan and Singh [16] demonstrated that tall vehicles severely restrict the visual field of drivers following behind them, leading to delayed recognition of traffic control devices. Bramson [17] reported similar findings decades earlier, yet signage positioning in many cities still fails to consider the vertical profiles of common road vehicles.

With the growth of intelligent transportation systems (ITS), tools such as VISSIM enable simulation of visibility-challenged scenarios. Studies by Al-Dabass et al. [18] and Badia et al. [19] advanced the use of microsimulation for traffic evaluation. Nonetheless, many traffic simulation setups do not incorporate real-world occlusion effects. Agarwal and Chakroborty [20], Pan and Xie [21], and Saraf [22] highlighted that occlusion duration and its implications must be integrated into realistic traffic modeling, especially for scenarios involving large-vehicle interactions.

Advances in computer vision and machine learning have enhanced traffic sign detection accuracy. Techniques explored by Koutaki and Okada [23] and Steffens and Endres [24] improved automated recognition capabilities. However, these technologies still struggle under environmental or vehicular occlusions, as shown by studies from Mounce and Zhou [25] and Chunsheng, Liu, and Chang [26]. Intelligent algorithms including fuzzy logic and AI-based control explored by Djahel et al. [7] and Allan et al. [27] offer potential improvements, but they also require visibility-awareness to be effective, especially in resource-limited urban contexts.

Urban design guidelines, such as those from the USDOT [28] and AASHTO [10], provide standards for geometric design and signage placement. However, limited enforcement in many African cities results in inconsistent application. Adeniyi [29] therefore recommends a multipronged approach to congestion mitigation that combines improved infrastructure design, public education, enforcement, and technology-based traffic management.

Additionally, observational studies provide insights into non-ideal visibility conditions encountered by road users. Zhang and Li [30] and Zhu and Li [31] documented the influence of visibility impairments on driving performance, while simulation-based investigations by Chen, Yang, and Yu [32] and Bologna and Gatto [33] illustrated how environmental factors affect traffic operations. Human-factors research by Katila and Norros [34] and Ma and Zhang [35] further underscores how

visibility conditions influence driver decision-making and safety outcomes.

In summary, although substantial progress has been made in modeling urban traffic flow, visibility-related factors particularly those caused by vehicle-induced occlusion, remain significantly underexplored. In the Nigerian context, where infrastructure limitations exacerbate the problem, there is a clear need for traffic management systems that account for real-world visibility challenges. By integrating global research findings with local observational data, this study aims to develop a practical and intelligent solution tailored to intersections like Agodi Gate, where visibility obstructions frequently compromise safety and operational efficiency.

3. Research problem

This study addresses the issue of visibility obstruction for approaching vehicles (Va) at intersections, particularly when larger vehicles block traffic signals. On March 1, 2024, at 2:44 p.m., the researcher experienced visibility obstruction when a truck blocked the signal while approaching Agodi Gate. Such obstructions lead to congestion, delays, and accidents, creating a need for an ITMS to enhance visibility and safety for all road users. This study aims to provide a solution to this problem using advanced technologies.

4. Aim and objectives

The study's aim is to establish a rational procedure to minimize traffic signal visibility problems observed along the two-lane segment at Agodi Gate, Ibadan. The ITMS will use IoT and GPS to improve signal visibility and overall road safety.

4.1. Objectives

1. Identify parameters contributing to traffic signal visibility blockage on selected Ibadan roads.
2. Develop models to represent and replicate these parameters.
3. Formulate strategies to enhance traffic signal visibility** and reduce obstructions at target intersections.

5. Justification

Addressing visibility problems at the Agodi Gate intersection is crucial to enhance safety, reduce congestion, and mitigate risks posed by obstructed signals. This research seeks to develop an ITMS that ensures safe and efficient traffic flow, potentially serving as a model for similar urban intersections. Figure 1 and 2 shows the satellite view of Oritamefa

to Agodi gate Road, Sabo-Badan Road as well as of Traffic intersection at Agodi Gate, Ibadan.

6. Scope of the study

1. Geographic Scope: Focuses on the two-lane segment leading to Agodi Gate from Orita Mefa in Ibadan, known for visibility issues due to traffic signal occlusions.
2. Technical Scope: The ITMS integrates IoT, GPS, and intelligent algorithms to optimize signal timing, improve traffic flow, and enhance intersection safety.
3. Functional Scope: Designed to provide alternative signal visibility through devices like dashboard displays or mobile apps, mitigating obstruction issues.
4. Evaluation Scope: Simulation studies, field trials, and assessments of intersection throughput, travel time savings, and safety improvements will evaluate the ITMS.
5. Stakeholder Scope: Collaboration with local traffic authorities, community members, and technology providers to ensure system effectiveness and community adoption.

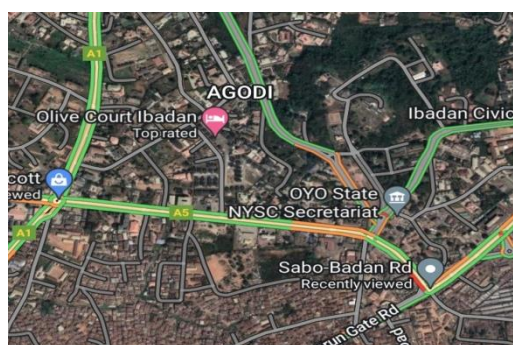


Figure 1. Sattelite View of Oritamefa to Agodi gate Road, Sabo-Badan Road, Ibadan
Source: Google Maps (2024)

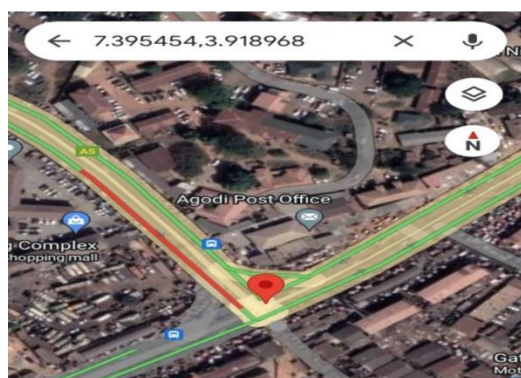


Figure 2. Satellite view of Traffic intersection at Agodi Gate, Ibadan
Source: Google Maps (2024)

7. Study Area

The study area for this PhD research work encompasses the bustling city of Ibadan, located in Oyo State, Nigeria. Known as one of the largest cities in Africa, Ibadan serves as a major economic and cultural hub in Nigeria, with a rapidly growing population and urban infrastructure. The city's road network plays a crucial role in facilitating transportation and connectivity within and beyond its borders. The focal point of the study is the stretch of road from Orita Mefa to Agodi Gate, a double two-way lane road that traverses through the heart of Ibadan. Designated as an A5 road, this route is integral to the city's transportation system, serving as a primary artery for vehicular movement. It connects various key destinations, including the prestigious Oyo State Governor's House and the Nigeria Television Authority (NTA), making it a vital thoroughfare for commuters and residents alike. Figure 3 shows Nigeria's road network.

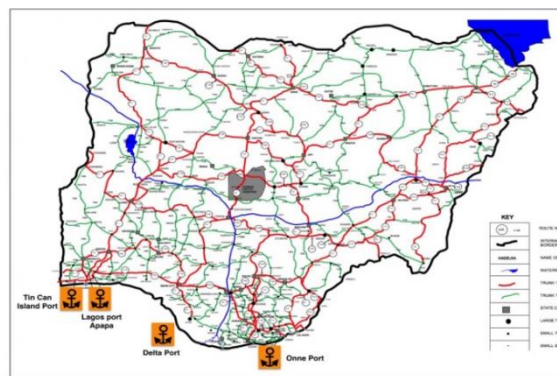


Figure 3. Nigeria's road network
Source: Google

At the intersection of Agodi Gate (coordinates: 7.395454, 3.918968), the study examines the challenges posed by obstructed visibility of traffic signals for approaching vehicles (Va), particularly when large trucks or vehicles block the view. This intersection is strategically positioned and experiences heavy traffic congestion, especially during peak hours, necessitating effective traffic management solutions. Similarly, the study also investigates the intersections at Orita Mefa (coordinates: 7.398436, 3.909088) and the Governor's House, which serve as critical junctions along the route. These intersections are equipped with traffic signals operating in the standard three-color mode (red, yellow, and green), along with timers to regulate traffic flow in multiple directions.

The road infrastructure along the study area is well-developed, featuring clearly demarcated lanes, signage, and markings to facilitate smooth vehicular movement. However, the presence of visibility

challenges at certain intersections poses safety concerns and necessitates innovative solutions for effective traffic management. The research aims to conduct a comprehensive analysis of the traffic systems in Ibadan, focusing on the identified intersections and addressing the specific issue of obstructed signal visibility. By leveraging advanced technologies and innovative strategies, the study seeks to propose solutions to enhance intersection safety, optimize traffic flow, and improve overall transportation efficiency in Ibadan Metropolis. Through rigorous data collection, analysis, and experimentation, the research endeavors to contribute valuable insights and practical recommendations to address the complex challenges of urban traffic management in Ibadan and beyond.

8. Methodology

This research employs a mixed-methods approach, combining qualitative and quantitative data collection to develop an ITMS tailored to visibility issues in Ibadan.

8.1. Step 1; traffic data collection

1. Observations and Surveys: Document traffic volume, vehicle types, and visibility issues.
2. Sensor and GPS Data: Use cameras, lidar, GPS, and GNSS devices to collect data on vehicle trajectories, speeds, and traffic flow.
3. Field Measurements: Measure intersection geometry, signal timings, and visibility aspects. - Figure 4 shows how a large vehicle (e.g., a truck) obstructs a driver's view of the traffic signal, which is one of the core issues addressed by Intelligent Traffic Management System (ITMS).

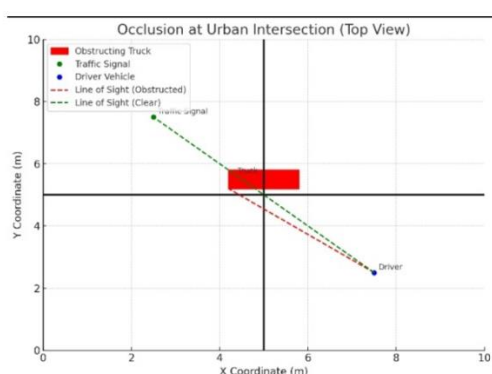


Figure 4. A sample diagram illustrating Occlusion Angle at Urban Intersections

8.2. Step 2: data collection for microscopic traffic simulation

Data on vehicle counts, speed, headway and signal

timing will be collected to create a detailed model in VISSIM, simulation software.

8.3. Step 3: prototype development and simulation analysis

1. Prototype Development: Design a prototype ITMS using relevant hardware and software components.
2. Simulation Setup: Create a VISSIM model of the intersection and simulate various traffic volumes, visibility scenarios, and weather conditions.
3. Performance Metrics: Measure travel times, throughput, and safety to assess the ITMS's effectiveness.
4. User Interface and System Refinement: Enhance the user interface based on testing and feedback to improve functionality and usability.

9. Report from reconnaissance survey: observed traffic signal occlusion at Agodi gate intersection

The reconnaissance survey conducted on March 1, 2024, at 2:44 pm (an off-peak period) at Agodi Gate intersection revealed a critical visibility obstruction issue. While driving a Jeep and approaching the T-junction from Orita Mefa, the driver encountered a blocked view of the traffic signal due to a truck in front, owned by Peak Milk. The truck's height (180 cm) and dimensions (width: 148 cm, length: 340 cm) contributed to this occlusion problem, obstructing the driver's view of the signal despite adjustments in viewing angles through the inner rearview mirror and steering position. Further factors, including headway, angle of view, and lane interswitch limitations, were examined to understand the circumstances of this obstruction.



Figure 5. The proposed system on Ibadan roads – traffic signal visibility occluded by the truck in front at Agodi-gate intersection Ibadan
Source: Field reconnaissance survey

Figure 5 shows the proposed system on Agodi-Gate Ibadan road and Figure 6 shows how Google lens was used to search out the vehicle type, model and dimension. Figure 7 shows result of the search via

Google lens proving the dimensions of the truck blocking the view. This analysis covers a thorough comparison of parameters observed during the survey, as well as a mathematical model simulating the traffic signal occlusion under various traffic conditions.

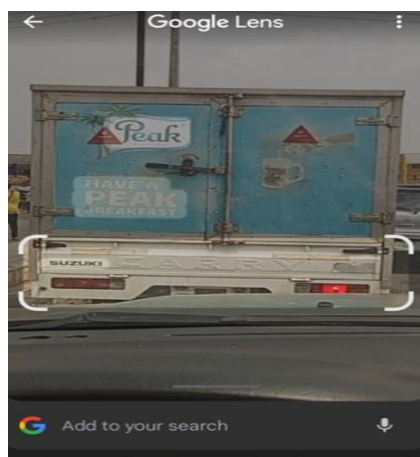


Figure 6. Using Google lens to verify the truck dimension (Source: Google)

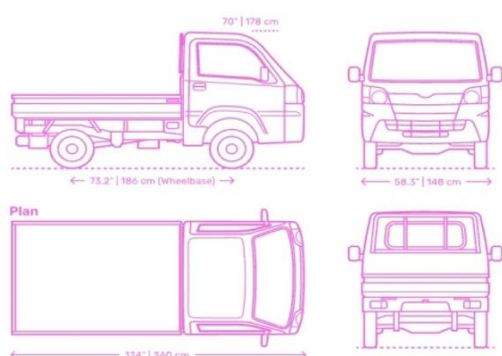


Figure 7. Dimensions of the observed Suzuki truck

9.1 Observational data And parameters

The following parameters were observed and recorded:

1. Time and Visibility:

- Time of Day: 2:44 pm, identified as an off-peak period for the intersection.
- Vehicular Type of Approaching Vehicle (V_a): Jeep
- Vehicular Type of Occluding Vehicle (V_f): Suzuki truck with dimensions of height (180 cm), width (148 cm), and length (340 cm).

2. Headway (H) and Obstruction:

- Headway: Distance between V_a and V_f . The precise headway between the vehicles was not

recorded, but it was noted that even an increase in headway may not necessarily clear the obstructed view due to the height and size of the truck.

- Occlusion Dynamics: Visibility was blocked even with adjustments in the driver's view angles, including through the center rearview mirror aligned with the dashboard.

3. Angle of View (θ): Angle at which the driver of V_a could potentially view the signal if unconcluded.

4. Signal Distance (d_s): Approximate distance between V_a and the traffic signal (observed at 30–50 meters).

5. Lane Conditions: Adjacent lane was congested, preventing lane-switching.

6. Waiting Time (T_{wait}): 45 seconds due to obstruction.

7. Driver's Response Delay ($T_{response}$): Unknown but hypothesized to increase due to obstruction.

8. Traffic Signal Details:

- Signal Height (h_s): Not recorded.
- Traffic Management: No traffic warden was observed on-site to assist with visibility issues.

9. Cell Transmission Model Reference: Assesses lane-switch potential.

Each of these parameters plays a role in the visibility issue. Pairwise comparisons reveal complex dependencies that contribute to visibility challenges. Lane interswitching was constrained, as the alternate lane on the carriageway was jammed. Principles of the cell transmission model suggest that lane-switching is beneficial in avoiding occlusion under ideal traffic conditions.

9.2. Pairwise comparison of parameters

To understand the interactions between parameters, each significant pair is analyzed in the context of visibility obstruction.

9.2.1. Time of Day (Peak vs. Off-Peak Period Analysis) vs. Vehicle Types (V_a and V_f). Visibility issues may be influenced by lighting conditions and vehicle positioning relative to the time of day. Although this observation was made in daylight, different lighting conditions may influence how V_a perceives signals if V_f is opaque. The relative sizes and positions of the vehicles further affect sightlines. Peak and off-peak differences are crucial in assessing flow variations and lane-switch potential:

a. Off-Peak (2:44 pm): Lower traffic density, theoretically allowing for alternate lane-switching under usual conditions.

b. Peak Period: Higher flow and potential obstruction increase, raising both occlusion frequency and T_{wait} due to limited maneuverability.

This analysis of traffic signal occlusion offers insights into potential adjustments in headway, lane-switch protocols, and signal positioning necessary to alleviate such incidents, especially at critical intersections like Agodi Gate.

1. Time of Day vs. Waiting Time

During off-peak periods, waiting time T_{wait} due to occlusion can vary with traffic density and vehicle position. Since this incident took place at 2:44 pm (off-peak), fewer vehicles were present on the road, theoretically allowing for easier maneuverability. However, the lane next to the obstructed vehicle was congested, limiting the possibility of lane interswitching. Peak periods could see increased waiting times and reduced sight clearance due to higher traffic volume.

2. Headway (H) vs. Signal Distance (d_s)

The visibility of the signal depends on the headway between V_a and V_f . A minimal headway reduces the likelihood of the driver of V_a observing the signal. Modeling H in relation to d_s is critical to determine the minimum safe headway that maintains visibility.

3. Signal Distance (d_s) vs. Angle of View (θ)

The signal distance and angle of view affect the visibility of the signal from within V_a . Given an obstructing vehicle V_f , a higher d_s or θ may potentially restore visibility by expanding the visual field. Calculating the critical angle at which θ restores visibility is essential.

4. Vehicle Types (Jeep and Truck) vs. Headway (d_h)

The obstructing vehicle's height (180 cm) and length (340 cm) played a significant role in the Jeep's visibility obstruction. Even with an increased headway, the visibility line would still be compromised due to the truck's height relative to the Jeep's eye level, positioned around 150 cm. The model, therefore, must account for this height disparity:

$$\text{Visibility} = \begin{cases} 0 & \text{if } H_{truck} > H_{jeep} \text{ and } d_h < d_{clearance} \\ 1 & \text{otherwise.} \end{cases}$$

5. Lane Condition vs. Interswitch Potential

The congested lane condition obstructed V_a from interswitching to the adjacent lane. The principles of the cell transmission model suggest that lane-switching becomes an unviable option if the adjacent lane is jammed; confirming that congestion severely limits alternative visibility strategies.

6. Angle of View $\theta_{obstruct}$ vs. Signal Distance (d_s)

For an effective visibility line, the angle formed by the driver's line of sight to the traffic signal must remain within a permissible range. If the angle due to truck width W_{truck} exceeds the visibility cone, occlusion occurs:

$$\theta_{obstruct} = \arctan \frac{W_{truck}}{d_h}, \text{ where visibility}$$

remains obstructed if $\theta_{obstruct}$ is 30°

Here, $\theta_{obstruct}$ suggests an obstruction in off-peak conditions. During peak periods, this angle is likely higher due to more vehicles occupying surrounding lanes.

7. Lane Interswitch vs. Cell Transmission Model Principles

The cell transmission model, which divides roads into cells that vehicles move through, supports lane interswitching to avoid occlusion. However, the jammed adjacent lane in this scenario prevented interswitching, reinforcing that alternative solutions (e.g., mobile app notifications of traffic signal status) may be necessary.

8. Driver Response Delay vs. Waiting Time T_{wait}

The driver's reaction time due to blocked signal visibility affects T_{wait} . Although the actual response delay is unknown, it contributes to total waiting time when visibility is obstructed.

9.3. Mathematical model to simulate traffic signal occlusion

The following model simulates the traffic signal occlusion by incorporating the aforementioned parameters.

Variables and Definitions

1. H: Headway between V_a and V_f (meters)
2. d_s : Distance from V_a to the traffic signal (meters)

3. θ : Angle of view from the driver's perspective in V_a to the traffic signal (degrees)
4. h_s : Height of the traffic signal above the road surface (meters)
5. h_{va} and h_{vf} : Heights of vehicles V_a and V_f respectively (meters)
6. L_{occl} : Length of the occluding vehicle (meters)
7. T_{wait} : Waiting time (seconds)
8. $T_{response}$: Driver's response delay due to occlusion (seconds)

9.4. Mathematical modeling of the observed incident

To simulate the visibility occlusion, we model key parameters using simplified mathematical relationships. These equations can serve as the basis for predicting occlusion probabilities under similar conditions.

1. Headway and Signal Obstruction Correlation

Given:

- H_t (Truck height) = 180 cm,

- H_v (Average driver eye level in Jeep) is assumed around 150 cm.

For a vehicle headway d_h , the probability of occlusion P_o can be assessed based on the height disparity and distance:

$$P_o(d_h) = \begin{cases} 1 & \text{if } H_t > H_v \text{ and } d_h < 50 \text{ meters} \\ 0 & \text{otherwise.} \end{cases}$$

2. Waiting Time Due to Occlusion

Using waiting time T_{wait} and delay T_{delay} :

$$T_{delay} = T_{wait} - f(\text{visibility conditions})$$

where $f(\text{visibility conditions})$ reflects driver response and view adjustments, which can further extend the wait time at the intersection.

3. Angle of Obstruction (Visibility Cone)

For a driver's visibility cone $\theta_{obstruct}$ with obstructed view due to vehicle width (W_t):

$$\theta_{obstruct} = \arctan \left(\frac{W_t}{d_h} \right),$$

where visibility remains obstructed if $\theta_{obstruct}$ is 30°

9.5. Model formulation

1. Occlusion Probability Function (P_o)

P_o is defined based on the visibility angle, headway, and height differential:

$$P_o(d_h, \theta_{obstruct}) = \begin{cases} 1 & \text{if } \theta_{obstruct} \geq 30^\circ \text{ or } H_{truck} > \\ 0 & \text{otherwise} \end{cases}$$

2. Waiting Time Function T_{wait}

If $P_o = 1$, the waiting time includes driver response delay T_{delay} , traffic conditions, and signal wait:

$$T_{wait} = T_{signal} + T_{delay}$$

where T_{signal} represents the signal cycle time at this intersection, and T_{delay} reflects occlusion response time.

3. Headway Adjustment Model

To prevent occlusion, headway d_h must reach a threshold, $d_{clearance}$, where the visibility line clears the truck's height:

$$d_{clearance} = \frac{H_{truck} - H_{jeep}}{\tan(\theta_{obstruct})}$$

This equation helps determine if increasing headway can reduce occlusion probability during peak and off-peak periods.

4. Headway and Signal Distance Relation

The visibility of the signal depends on the headway H . If H is less than a critical distance H_{crit} , the signal remains occluded. We define H_{crit} as follows:

$$H_{crit} = \frac{L_{occl}}{\tan(\theta)}$$

where θ is the critical viewing angle.

5. Angle of View Calculation

The angle of view θ at which V_a can view the signal depends on the vertical alignment:

$$\theta = \arctan \left\{ \frac{h_s - h_{vf}}{d_s} \right\},$$

where h_s is the height of the signal and h_{vf} is the height of the occluding vehicle.

6. Waiting Time and Response Delay Impact

The obstruction leads to a delay in the driver's response. We define the response delay T_{response} as proportional to T_{wait} :

$$T_{\text{response}} = \alpha \cdot T_{\text{wait}}$$

where α is an empirical factor dependent on driver reaction time under obstructed view conditions.

7. Interswitch and Cell Transmission

Lane-switching viability can be represented with the cell transmission model. Given a congested adjacent lane, the probability of lane-switching, (P_{switch}), approaches zero:

$$P_{\text{switch}} = \max \left[0, 1 - \frac{\rho_{\text{adj}}}{\rho_{\text{crit}}} \right]$$

where ρ_{adj} is the vehicle density in the adjacent lane and ρ_{crit} is the critical density for lane-switching feasibility.

9.6. Simulation and analysis

The model allows simulation of scenarios where:

- H varies to determine the minimal headway avoiding occlusion.
- d_s is adjusted to test its effect on visibility with constant H .
- Different vehicle heights h_{va} and h_{vf} are analyzed to understand how variations affect signal sightlines.

Simulation Steps:

1. Initialize Variables: Set H , d_s , Θ , h_s , h_{va} , h_{vf} , L_{occl} , T_{wait} , α .
2. Calculate H_{crit} : Compute the minimum headway required for visibility using the angle of view.
3. Assess T_{response} : Calculate potential delays due to signal occlusion.
4. Run Scenarios: Simulate scenarios for various headway distances and lane-switch probabilities using the cell transmission model.
5. Optimize Parameters: Identify the headway and angle of view combinations that mitigate signal occlusion.

10. Limitations

This study presents an innovative approach to mitigating visibility obstruction at urban intersections using IoT and GPS-enabled Intelligent Traffic Management Systems (ITMS). However, several limitations exist:

1. Limited Field Observations: The empirical investigation focused primarily on a single intersection (Agodi Gate), with only one detailed case involving a Peak Milk truck. This limits the generalizability of the findings across other high-traffic intersections in Ibadan or Nigeria at large.
2. Narrow Geographic Scope: While Agodi Gate is representative of congestion-prone intersections, the absence of comparative data from other intersections restricts the study's broader applicability.
3. Simulation-Heavy Approach: The use of VISSIM simulation modeling, although insightful, does not fully reflect real-world system dynamics. No pilot deployment of the ITMS was conducted to validate performance under actual traffic conditions.
4. Infrastructural Feasibility: The study does not provide a comprehensive cost-benefit analysis or feasibility assessment of implementing IoT/GPS systems in resource-constrained settings typical of many Nigerian cities.
5. Lack of Stakeholder Input: There is minimal engagement with key stakeholders such as drivers, law enforcement agencies, or transport authorities. Their perspectives on system usability, operational challenges, and compliance are not yet considered.
6. Unmodeled Weather Conditions: Although visibility obstructions due to environmental factors like rain and fog are acknowledged, these were not simulated or analyzed within the scope of the current study.

These limitations provide a clear direction for future research efforts aimed at strengthening the practical relevance and scalability of the proposed ITMS solution.

11. Conclusion

The reconnaissance survey at the Agodi Gate intersection provided a comprehensive understanding of the factors contributing to traffic signal occlusion at signalized T-junctions. The data and analysis revealed that the primary contributors to the problem include vehicular height disparities, headway distances, obstructing vehicle dimensions, and lane

congestion. These elements significantly affect visibility and driver response times, creating delays and potential risks at intersections. The observed scenario involved a truck blocking the view of a traffic signal, primarily due to its height and width, which exceeded the visibility line of the approaching Jeep. The height disparity between the vehicles was identified as a critical factor, especially at shorter distances. The truck's dimensions, combined with the narrow headway of approximately 30–50 meters, created an obstruction angle that rendered the traffic signal invisible from the Jeep's position. This issue was exacerbated by the jammed condition of the adjacent lane, which prevented the Jeep from switching lanes to improve visibility.

Waiting time at the intersection, recorded as 45 seconds, was extended by the driver's inability to see the signal, leading to response delays. The lack of alternative sightlines, even with adjustments via rearview mirrors or steering alignment, indicated that traditional driver compensations are insufficient under such conditions. Furthermore, the absence of traffic wardens or real-time assistance during the off-peak period underscored the limitations of manual traffic management in addressing visibility challenges. The comparative analysis of parameters showed that occlusion is more likely during peak periods when lane congestion is higher, and maneuvering space is reduced. This suggests that traffic density plays a significant role in amplifying visibility issues, making the integration of intelligent traffic management systems essential. Such systems, leveraging IoT and real-time data, could dynamically adjust signals or provide alternative guidance to drivers to mitigate the impact of occlusion. This paper introduces an Intelligent Traffic Management System (ITMS) that utilizes IoT and GPS technology to improve signal visibility and reduce delay at urban intersections in Nigeria, with a specific focus on the Agodi Gate axis of Ibadan metropolis. The research demonstrates that visibility obstructions caused by large vehicles (such as trucks) significantly affect signal interpretation, leading to increased intersection delay and heightened safety risks. Through a combination of field observation, system design, and simulation, the study presents a functional prototype capable of transmitting real-time traffic signal information to drivers via mobile and dashboard interfaces. The approach enhances decision-making efficiency at intersections and can support broader urban traffic decongestion strategies.

In summary, the study highlights the need for innovative solutions to address traffic signal occlusion. Improving road infrastructure, implementing intelligent systems, and considering vehicle dimensions in urban planning are critical steps toward enhancing visibility, reducing delays, and ensuring smoother traffic flow at intersections. These findings pave the way for future research to develop

technologies and strategies that address visibility-related challenges in modern traffic systems.

Furthermore, this analysis provides insight into mitigating traffic signal occlusions through adjustments in headway, signal positioning, and potential use of driver-assist technologies. Simulation models based on this data can be applied to other intersections, aiding in the design of Intelligent Traffic Management Systems that account for visibility dynamics under varying traffic conditions. Further studies and real-world tests will validate this model's effectiveness and refine parameters like driver response delay and occlusion thresholds. Despite its limitations, the proposed solution represents a step forward in integrating smart technology into Nigeria's urban mobility planning. The framework holds significant potential for replication in similar urban centers and can be expanded with stakeholder input, real-world testing, and multi-site validation. Future research should also consider cost-effective hardware alternatives, broader stakeholder collaboration, and environmental adaptability to facilitate full-scale adoption of ITMS as part of smart city initiatives in developing regions.

This mathematical model helps identify the critical headway and angle of view needed to prevent signal occlusion at the Agodi Gate intersection. Using simulated conditions, the model can guide modifications in traffic signal placement, height adjustments, and lane management strategies to improve visibility and reduce response delays. Future work will involve field testing to validate model predictions and refine parameters based on empirical data.

12. Contribution to knowledge

This study will contribute an ITMS framework specifically for addressing visibility obstructions at urban intersections. The system's design is intended to be replicable for other cities facing similar challenges. Additionally, by addressing headway and traffic flow in tandem with visibility, this ITMS prototype stands to improve intersection efficiency, reduce economic costs associated with congestion, and enhance road safety for Ibadan and potentially other urban centers. This proposal incorporates insights from real-world observations and past studies to address the pressing visibility challenges in urban traffic management, using IoT and GPS technology to improve safety and efficiency at intersections.

13. References

- [1] Akintayo, O. (2011). Modeling critical traffic parameters for congestion mitigation in urban areas.
- [2] Akintayo, O., and Agbede, O. (2009). Headway distribution modeling on two-lane roads in Ibadan.

- [3] Greenshields, B. (1935). A study of traffic capacity.
- [4] Daganzo, C. (1994). The cell transmission model: A dynamic representation of traffic flow.
- [5] Hidas, P. (2002). Modeling vehicle interactions in traffic simulations.
- [6] Salisu, W., et al. (2020). Urbanization and traffic visibility challenges in Nigeria.
- [7] Djahel, S., et al. (2015). Traffic management challenges and ITS communication gaps.
- [8] Al-Kaisy, A., et al. (2005). Signage visibility obstruction caused by large vehicles.
- [9] Hou, Y., Tian, Z., and Zhang, Y. (2017). Impacts of roadside vegetation on visibility.
- [10] AASHTO. (2018). AASHTO Green Book: A Policy on Geometric Design of Highways and Streets.
- [11] Zhou, H., and Xie, Y. (2017). Effects of weather and lighting on traffic signal visibility.
- [12] Faghri, A. (2000). Design considerations for traffic sign placement.
- [13] AAA Foundation for Traffic Safety. (2023). Visibility and driver recognition distances.
- [14] Abdel-Aty, M., et al. (2019). Adaptive signal systems and real-time traffic safety.
- [15] Luo, Z., Zhang, L., and Zhang, Y. (2014). Heavy-vehicle occlusion effects on recognition distance.
- [16] Mahajan, P., and Singh, A. (2016). Driver decision-making and visibility obstruction.
- [17] Bramson, M. (1971). Truck-induced road sign occlusion and safety implications.
- [18] Al-Dabass, D., et al. (2017). VISSIM applications in traffic visibility modeling.
- [19] Badia, L., et al. (2006). Simulation of traffic under constrained conditions using VISSIM.
- [20] Agarwal, P., and Chakroborty, P. (2014). Modeling traffic sign occlusion duration.
- [21] Pan, Y., and Xie, Y. (2014). Traffic simulation incorporating occlusion effects.
- [22] Saraf, R. (2019). Large-vehicle interactions and sign visibility in simulations.
- [23] Koutaki, G., and Okada, M. (2019). Traffic sign detection under occlusion using computer vision.
- [24] Steffens, T., and Endres, F. (2022). Machine learning for traffic sign recognition under occlusion.
- [25] Mounce, J., and Zhou, H. (2000). Automated systems and persistent occlusion challenges.
- [26] Chunsheng, L., Liu, Y., and Chang, G. (2016). Occlusion effects on automated traffic detection.
- [27] Allan, A., et al. (2017). Fuzzy logic applications in real-time traffic control.
- [28] USDOT. (2014). Manual on Uniform Traffic Control Devices (MUTCD).
- [29] Adeniyi, O. (2018). Traffic congestion mitigation strategies in Nigerian cities.
- [30] Zhang, H., and Li, S. (2015). Observational visibility studies in congested intersections.
- [31] Zhu, C., and Li, Y. (2019). Human factors and visibility issues in traffic environments.
- [32] Chen, X., Yang, X., and Yu, L. (2010). Simulation-based visibility analysis frameworks.
- [33] Bologna, S., and Gatto, V. (2009). Traffic modeling under reduced visibility conditions.
- [34] Katila, M., and Norros, I. (1998). Driver perception and reaction under limited visibility.
- [35] Ma, J., and Zhang, Y. (2013). Drivers' cognitive responses to visual obstructions in traffic.

Least Significant Bit (LSB) Audio Steganography with AES 256 Encryption

Olaniyi Abiodun Ayeni, Adekunle Obajimi Isaac
Federal University of Technology, Nigeria

Abstract

This paper aims to strengthen Least Significant Bit (LSB) audio steganography with the application of AES-256 encryption to overcome some limitations related to other forms of LSB steganography, such as low security and detectability. Hiding messages in audio files without compromising any audio quality has been done by including cryptographic protection in the system. The solution was coded in Python and used data hiding as well as AES-256 encryption. In this case, the encrypted information was stored in the 16 least significant bits of audio samples. The performance of the systems can be measured by the Signal-to-Noise Ratio (SNR), data integrity test, and capacity test. This has a picture quality SNR of 105.45 dB, which means very little distortion in the audio and extracts the messages with a perfect 100 percent measure. This system allows embedding between 0 and 110,250 characters of a 20-second WAV file, which presents a good balance between the capacity to convey information and the audio quality of the given file.

1. Introduction

With the ever-increasing pressure to have secure digital communications in a time of cyber attacks, the art of concealing data behind seemingly harmless files, steganography, has received new insights. In comparison to cryptography, where top-quality protection is provided by turning the information into ciphertext, covered by an authentic sign, steganography is more obscure since it attempts to make the fact of the message being there undetectable [6]

Least Significant Bit is the encoding technique that has become dominant in audio steganography since it is simple and imperceptible among other steganographic techniques [10]. As evidenced by [1], the LSB audio steganography allows for efficiently disguising textual data into audio content without any loss or perceptual quality parameters. Nevertheless, they also stressed that LSB is very susceptible to

statistical steganalysis and it does not provide any inherent countermeasure to security like encryption. In order to overcome the demerits of plain LSB encoding, different additions have been suggested by the researchers.

The XOR encryption allowed a second layer of security because [3] created a Modified LSB algorithm based on this method. XOR encryption is, however, weak in comparison with modern encryption and, as such, should not be used to provide high security. Further than proposals such as randomized LSB embedding, which [4] considered, demonstrate even better resistance to detection, but have yet to demonstrate strong cryptographic integration. In addition to traditional LSB, deep learning methods occur. [2] developed an odd name, PixInWav, which is a type of residual steganography based on embedding image pixels in audio with neural networks. Although this method clearly enhances the embedding capacity and robustness, it is quite time-consuming and not applicable in cases of lightweight or real-time applications.

Research Aim

The aim of this research is to design and implement a secure LSB audio steganography system with AES-256, modifying the least significant bits of audio samples in order to encode secret information without changing the apparent sound quality.

2. Literature Review

After thoroughly reviewing multiple research papers, the contributions made by various researchers are discussed.

i. Alghamdi et al. (2022)

Motivation: To explore how LSB encoding performs in concealing data within audio signals and understand its feasibility.

Objective: To implement and evaluate LSB steganography using standard audio files, mainly

focusing on perceptual transparency and data hiding capacity.

Methodology: Employed basic LSB substitution without encryption, embedding textual data into WAV files. The focus was on assessing how imperceptible the changes were.

Contribution to Knowledge: Provided a clear experimental baseline for LSB techniques, showing that while imperceptible to the human ear, LSB can be statistically detected.

Limitations: The lack of encryption made the embedded data readable upon extraction. The method was highly predictable and vulnerable to even basic statistical analysis.

ii. Ālsabhany et al. (2019)

Motivation: ordinary phase coding methods have poor capacity and lack adaptability.

Objective: design an adaptive multi-level phase coding scheme that improves capacity, robustness, and transparency.

Methodology: dynamic selection of embedding levels using interval centering quantization and error spreading; tested for SNR and capacity.

Contribution to knowledge: introduced adaptive embedding that achieves high data rates with good audio quality.

Limitations: increased computational complexity, still less robust to heavy compression or filtering, and limited test opportunities.

ii. Geleta et al. (2022) – PixInWav

Motivation: To explore embedding image pixels into audio using neural networks, enabling deep steganographic embedding.

Objective: To build a deep learning-based system for hiding images inside audio without perceptible changes.

Methodology: Used residual neural networks to embed visual data into residual audio signals.

Contribution to Knowledge: Introduced AI-driven steganography into the audio domain, pushing boundaries of capacity and robustness.

Limitations: Required GPU-based computation, was complex to train, and risked audio degradation at high embedding rates.

iv. Chen et al. (2020)

Motivation: improve imperceptibility and robustness over time domain LSB methods by using transform domain embedding.

Objective: develop a high SNR audio steganography method using wavelet coefficients.

Methodology: Apply discrete wavelet transform; embed bits in selected coefficients, reconstruct stego audio via inverse DWT.

Contribution to knowledge: achieved high imperceptibility and audio quality; validated the wavelet domain as an effective embedding space.

Limitations: lower embedding capacity; computationally expensive, and vulnerable to severe audio compression or recording.

v. Hemanth et al. (2022)

Motivation: To address the lack of data protection in traditional LSB-based audio steganography, which exposes hidden data to unauthorized access.

Objective: To improve confidentiality by using XOR encryption in combination with the LSB technique.

Methodology: Embedded XOR-obfuscated data in the least significant bits of audio samples. Focused on simplicity and low computational overhead.

Contribution to Knowledge: Demonstrated that even a lightweight bit-level transformation (XOR) can enhance confidentiality in steganography.

Limitations: XOR encryption is weak by modern cryptographic standards and can be easily broken using brute-force or pattern-based attacks. The method does not scale well for high-security applications.

vi. Jezdimirovic et al. (2023)

Motivation: Traditional LSB steganography is highly susceptible to detection through statistical steganalysis. The authors' aim was to enhance the security of these methods to resist such attacks.

Objective: To improve the security of LSB-based audio steganography by proposing enhancements that make detection more powerful.

Methodology: The study involved modifying the classic LSB technique to introduce variability or randomness in the embedding process. Detection performance was evaluated using standard steganalysis tools.

Contribution to Knowledge: The work identified vulnerabilities in existing LSB techniques and proposed initial improvements. It added to the body of knowledge by highlighting how detection is possible and suggesting enhancements.

Limitations: While security was modestly improved, the proposed enhancements were not sufficient to resist advanced steganalysis techniques. Also, the system lacked an integrated encryption mechanism.

vii. Padmaja and Basha (2025)

Motivation: To build a steganography system capable of handling diverse data types (text, images, and audio) while maintaining strong encryption and data hiding.

Objective: To develop a hybrid audio steganography model for embedding multiple data types with high security.

Methodology: Used multi-bit substitution with robust encryption layers to embed large payloads into audio signals.

Contribution to Knowledge: Proved the feasibility of multi-domain embedding in audio with layered encryption, extending audio steganography applications.

Limitations: Multi-bit embedding caused notable audio distortion, affecting perceptual transparency. The system also introduced high computational complexity.

viii. Shanthakumari et al. (2021)

Motivation: Reduce detectability in the traditional LSB method

Objective: implement an LSB matching revisited algorithm to improve security and minimize distortion

Methodology: modify sample values randomly instead of direct bit replacement to confuse steganalysis.

Contribution to knowledge: better imperceptibility and resistance to statistical detection compared to classic LSB

Limitations: slightly reduced payload capacity, not robust against format conversion or lossy operations

ix. Younis and Mohasefi. (2024)

Motivation: improves robustness and security of video steganography using intelligent optimization and encryption.

Objective: design a secure embedding system combining genetic algorithms and RSA encryption.

Methodology: encrypt and scramble secret data, optimize embedding positions using GA, and embed within video frames.

Contribution to knowledge: demonstrated a hybrid optimization encryption model for higher embedding security and efficiency.

Limitations: very high computational overhead and more suited for research than real-time applications.

x. Simplilearn (2021)

Motivation: provide foundational knowledge on steganography techniques and applications

Objective: explain core types, methods, and use cases of steganography for beginners.

Methodology: descriptive review and categorization of existing steganography methods (test, image, audio, and video).

Contribution to knowledge: educational resource summarizing principles and types of steganography

Limitations: lacks experimental or technical depth, purely conceptual, and not research focused.

3. Design of System Workflow

This system workflow successfully integrates LSB steganography with AES-256 encryption using Python to deliver a secure, imperceptible, and efficient data-hiding system. This hybrid method:

- i. LSB detectability
- ii. Ensures strong confidentiality
- iii. Avoids distortion using optimized single-LSB embedding

4. Implementation and Results

This system securely hides a secret message inside an audio file using Least Significant Bit (LSB) audio steganography, then encrypts the stego audio using AES 256 encryption.

Tools and Environment

Programming Language: Python 3.10

Libraries Used:

- i. NumPy – for audio sample manipulation.
- ii. Cryptography – for AES-256 encryption and decryption.
- iii. Tkinter – to build a simple GUI interface.

Scripts Implemented

- I. Embed.py – Embeds plaintext into a 16-bit WAV file using LSB. This is the core of the steganography process.
 - i. It takes two inputs: the cover audio file (usually a .wav file) and the encrypted message.
 - ii. The script then modifies the LEAST SIGNIFICANT BIT of the audio samples to encode the binary data of the encrypted message
 - iii. The changes are imperceptible (the ear can't tell the difference) to the human ear, ensuring the audio quality remains nearly identical to the original
 - iv. The output is a stego audio file, which contains the hidden message.

```
import wave
import numpy as np

# === INPUTS ===
input_audio = "lsb_carrier_20s.wav"
output_audio = "lsb_stego_output.wav"
message = "Steganography is the process of hiding messages"

# === Step 1: Convert message to binary + delimiter ===
delimiter = "###" # Marks the end of the message
full_msg = message + delimiter
binary_msg = ''.join(f"{ord(c):08b}" for c in full_msg)
```

Figure 1: Embed.py

- II. Encrypt.py – Encrypts the stego-audio using AES-256 with CBC mode.
 - i. The user enters a plain text message and a password.

- ii. The system uses the AES-256 algorithm to convert the plaintext into ciphertext, ensuring that even if the message is extracted without authorization, it remains unreadable.

```
from cryptography.hazmat.primitives.ciphers import Cipher, algorithms, modes
from cryptography.hazmat.primitives import padding
from cryptography.hazmat.backends import default_backend
import os

# === Step 1: Read the audio file to encrypt ===
input_file = 'lsb_stego_output.wav'
with open(input_file, 'rb') as f:
    data = f.read()

# === Step 2: Generate AES-256 key and IV (both 32 and 16 bytes) ===
key = os.urandom(32) # 256-bit key
iv = os.urandom(16) # AES block size for CBC
```

Figure 2. Encrypt.py

- III. Decrypt.py – Decrypts the audio file and extracts the hidden message.
- It first extracts the hidden binary data from the stego audio file using the LSB decoding method
 - The extracted data is then (cipher text) is then decrypted using the same AES key or password that was used during the encryption
 - The result is the original plaintext message, successfully retrieved from the audio file.

```
from cryptography.hazmat.primitives.ciphers import Cipher, algorithms, modes
from cryptography.hazmat.primitives import padding
from cryptography.hazmat.backends import default_backend
import binascii
import struct

def aes_decrypt_file(encrypted_path, key_hex, iv_hex):
    """
    Decrypts the AES-CBC encrypted file at encrypted_path using the provided
    hex-encoded key and IV. Returns the decrypted plaintext bytes.
    """
    key = binascii.unhexlify(key_hex)
    iv = binascii.unhexlify(iv_hex)
    with open(encrypted_path, 'rb') as f:
        ciphertext = f.read()
```

Figure 3. Decrypt.py

- IV. Graphics.py – Combines all steps in a Tkinter GUI.

This script provides a user-friendly graphical interface that connects all the other scripts together.

- It allows users to choose the options available, such as embedding, extracting, encrypting, decrypting, and so much more, without using the command line.
- Through simple buttons and input fields, the user can select an audio file of their choice, enter a message or a password, and encrypt it for safekeeping.
- It enhances usability by linking the encryption, embedding, and decryption processes in one interesting and interactive interface.

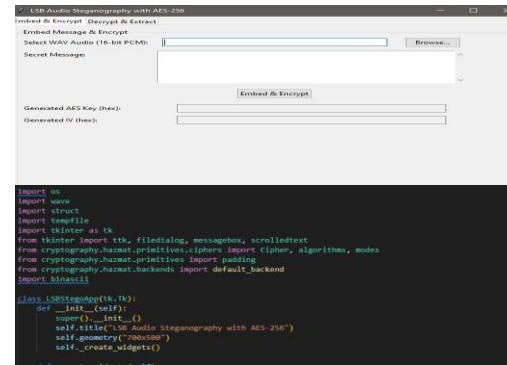


Figure 4. Graphics.py

Encoding Process

- convert the message to binary (e.g., "HELLO" → binary).
- Modify LSBs of audio samples to embed a binary message.
- Encrypt modified audio using AES-256 with a 256-bit key and IV.
- Save the encrypted file as a binary blob for transmission.

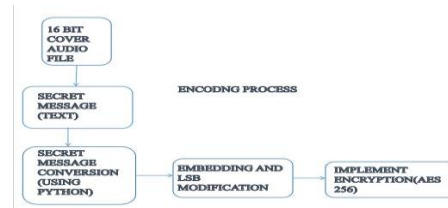


Figure 5. Encoding Phase

Decoding Process

- Decrypt the audio using AES-256 and the correct key/IV.
- Extract LSBs from audio samples.
- Reconstruct the binary into the original message text.

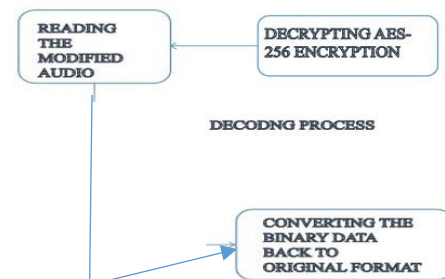


Figure 6. Decoding Phase

5. Results

1. Perceptual Transparency: This refers to how well the data that is hidden blends into the audio file without being noticed by the human senses. This means that after embedding a secret message, the resulting stego audio sounds almost identical to the original. A listener should not detect any distortion, noise, or change in quality

- i. **Waveform Comparison:** The waveforms of the original and stego audio were nearly identical. As shown in Figures 7 and 8.

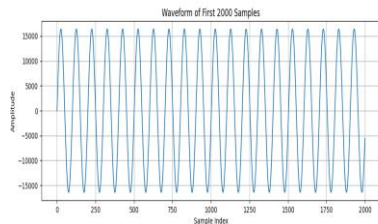


Figure 7. Waveform of original audio

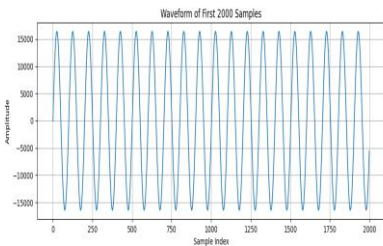


Figure 8. Waveform of stego audio

- ii. **Signal-to-Noise Ratio (SNR):** Calculated using a Python script. The signal-to-noise ratio (SNR) represents how much the stego audio file differs from the original audio after embedding. The calculated SNR value was **105.45 dB**, which is exceptionally high. This means the stego audio is almost identical to the original, with differences far beyond the range of human hearing. The LSB embedding technique introduced virtually no distortion, allowing the audio to retain its natural clarity and quality. These remarkable results prove that the system effectively conceals hidden information while keeping the sound perfectly transparent, achieving an ideal balance between security and perceptual transparency



Figure 9. Python script for calculating SNR

- iii. **Listening Test:** Human listeners could not distinguish between the original and stego audio.

2. Data Integrity

- i. **Test Message:** A detailed paragraph on steganography was embedded.
- ii. **Output:** The exact same paragraph was successfully and **fully extracted**, proving **100% accuracy** as shown in Figure 10

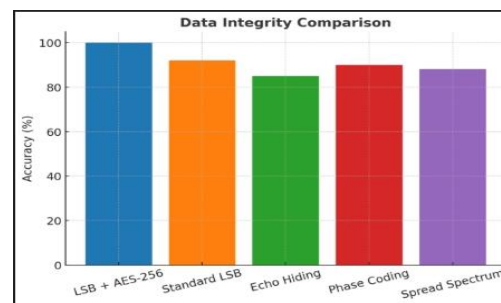


Figure 10. Data integrity of LSB + AES 256 compared to other techniques

The diagram above illustrates how the proposed LSB audio steganography system, combined with AES 256 encryption, outperforms other existing techniques in maintaining data integrity. While traditional methods like basic LSB and phase coding show slight losses in accuracy and audio quality, the LSB + AES-256 approach ensures that the hidden messages are recovered perfectly after extraction. This high level of performance is achieved because the AES layer secures the message, while the LSB technique preserves the natural quality of the sound. Together, they provide a strong balance of confidentiality, reliability, and imperceptibility, proving the method's overall efficiency and robustness

3. Capacity Analysis

- i. **Audio Used:** 20-second, 16-bit, 44.1 kHz mono WAV file.

- ii. **Maximum Capacity:** Up to **110,250 characters** embedded without noticeable quality loss as shown in Figure 12.

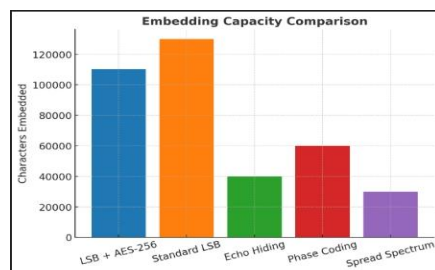


Figure 11. Embedding capacity compared to other techniques.

The diagram above shows the embedding capacity analysis of LSB + AES 256 encryption compared to other techniques, showing how much data they can hold without losing quality. The proposed LSB combined with AES 256 encryptions maintains a high embedding capacity while greatly improving security. Although AES encryption slightly increases the data sizes, the system still embeds large amounts of information effectively. Unlike basic LSB or phase coding methods, which may store more bits but offer weaker protection, the combined approach ensures that hidden data remains confidential and intact. This balance between capacity, security, and transparency makes the LSB + AES-256 technique a more reliable and efficient method for secure audio steganography.

6. Alignment with Prior Research Problems

- i. **Problem 1: LSB's Detectability**
Prior research by Jezdimirovic et al [6] and Alghamdi et al [1] has shown that LSB is vulnerable to statistical steganalysis. My solution was to implement AES-256 encryption to prevent detection (as shown by the data integrity test, figure 10).
- ii. **Lack of Confidentiality**
Prior Research by Hemanth et al [5] and Alsabhany and Azni [2] highlighted the lack of encryption and resilience in audio steganography. My solution was to implement "Hybrid embed then encrypt" using AES-256 to ensure confidentiality (Figures 1 and 2).
- iii. **Quality Loss With Multi-LSB**
Prior Research by Padmaja and Basha [7], Chen et al [4], and Geleta et al [3] shows that deep learning or multi-bit embedding increases complexity or audio distortion.

My solution used a single LSB with optimized binary embedding and encryption - maintaining SNR > 105 dB and a high capacity of 110k characters (see Figure 9).

7. Conclusion

This project successfully designed and implemented a secure audio steganography system that combines **Least Significant Bit (LSB)** embedding with **AES-256 encryption** to address critical limitations in traditional steganographic techniques. Through this hybrid approach, the system effectively mitigated known weaknesses such as:

- i. Detectability of LSB-based steganography,
- ii. **Lack of confidentiality** in standalone stego systems,
- iii. And **audio distortion** caused by multi-bit embedding.

By using single-bit LSB embedding and encrypting the entire stego audio file, the system achieved outstanding results:

- i. **High perceptual transparency** (SNR of 105.45 dB)
- ii. **100% data extraction accuracy**, and
- iii. **Impressive capacity** (110,250 characters in a 20-second WAV file).

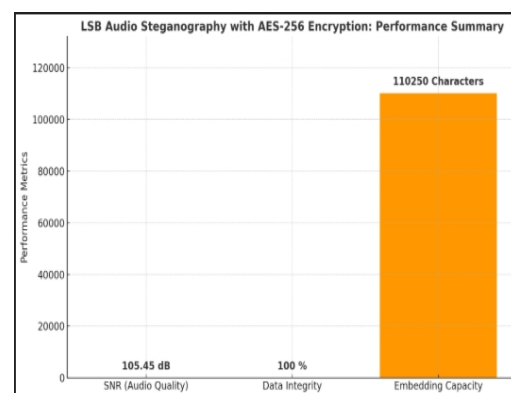


Figure 9. Audio Steganography Performance Summary

The Python-based implementation, supported by modular scripts and an easy-to-use GUI, proved both **technically feasible and practical** for real-world secure communication in lossless environments.

While LSB techniques remain fragile to compression and steganalysis, this project demonstrates that integrating robust encryption like AES-256 can significantly enhance the **security, reliability, and usability** of audio steganography systems.

7. References

- [1] Alghamdi, Ahmed, and Alzahrani, Ahmed, and Alzahrani, Abdulrahman, and Alsolami, Eesa, and Alarood, Alaa. (2022). Audio Steganography Method Using Least Significant Bit (LSB) Encoding Technique. *International Journal of Computer Network and Information Security*. 22. 427. DOI: 10.22937/IJCSNS.2022.22.7.53.
- [2] Alsabhany, A. A., Ridzuan, F., and Azni, A. H. (2019). The Adaptive Multi-Level Phase Coding Method in Audio Steganography. *IEEE Access*, 7, 129291–129306. DOI: 10.1109/access.2019.2940640.
- [3] Geleta, M., Puntí, C., McGuinness, K., Pons, J., Canton, C., and Giro-i-Nieto, X. (2022). PixInWav: Residual steganography for hiding pixels in audio. In *ICASSP 2022 - IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 2485–2489). DOI: 10.1109/ICASSP43922.2022.9746191.
- [4] Chen, S.-T., Huang, T.-W., and Yang, C.-T. (2020). High-SNR steganography for digital audio signals in the wavelet domain. *Multimedia Tools and Applications*, 80(6), 9597–9614. DOI: 10.1007/s11042-020-09980-6.
- [5] Hemanth, H., Ram, V. and Raghavendran, S., and Narayanan, Subhashini. (2022). Modified LSB Algorithm Using XOR for Audio Steganography. DOI: 10.1007/978-981-16-9012-9_30.
- [6] Jezdimirovic, J., Pekez, N., and Kovacevic, J. (2023). Security enhancement of the LSB-based audio steganography method. 77-82. DOI: 10.1109/ZINC58345.2023.10174020.
- [7] Padmaja, Srinivasa, and Basha, Shaik Mahaboob. (2025). A Novel Hybrid Steganography Approach for Securing Text, Images, and Audio with Robust Encryption in Audio Steganography. *ITEGAM-Journal of Engineering and Technology for Industrial Applications (ITEGAM-JETIA)*. DOI: 10.5935/jetia.v11i52.1235.
- [8] Shanthakumari, R and Devi, Ermela and Rajadevi, R, and Balasubramaniam, Bharaneeshwar. (2021). Information Hiding in Audio Steganography using LSB Matching Revisited. *Journal of Physics: Conference Series*. 1911. 012027. DOI: 10.1088/1742-6596/1911/1/012027.
- [9] Simplilearn. (2021, October 25). What is Steganography? Types, Techniques, Examples, and Applications | Simplilearn. [Simplilearn.com.https://www.simplilearn.com/what-is-steganography-article](https://www.simplilearn.com/what-is-steganography-article). (Access Date: 12 February 2023).
- [10] Younis, Ruaa, and Mohasefi, Jamshid. (2024). Enhanced video steganography using genetic algorithms, scrambling techniques, and RSA encryption for secure data embedding. *Edelweiss Applied Science and Technology*. 8. 9121-9127. DOI: 10.55214/25768484.v8i6.3948.

CNN-Based Detection of Image Manipulation

D. O. Babatunde, O.Y. Ogunlola, B.K. Alese
Federal University of Technology, Akure, Nigeria

Abstract

This project presents the design and implementation of an automated system for detecting image manipulation by using Convolutional Neural Networks, particularly the VGG16 architecture. It aims at addressing the increasing demand for fast and credible methods to detect tampered images in the field of digital forensics. The system accordingly detects and classifies images that are either tampered or original by using the phases of preprocessing, model selection, and training. VGG16 with transfer learning was subsequently trained with a dataset of tampered images and authentic images. The goal was successful since the model was able to detect manipulation with high accuracy. A web-based application using FastAPI was developed through which the users could upload images for real-time classification. The model was evaluated using metrics such as accuracy, precision, recall, and F1 score to provide a thorough understanding of the model performance.

The dataset was split into training, validation, and test sets to achieve consistency and sanity in evaluation, so that the model generalizes well to unseen data. The preprocessing steps, which include resizing and normalization, ensure that images fed as inputs are standardized for robust learning.

The interface demonstrates deep-learning models' support of forensic applications for practical use, especially toward investigators, researchers, and legal experts dealing with sensitive evidence. In conclusion, this project demonstrates how CNN-based models can be used efficiently to detect image manipulation and offers a pathway toward further advancement on larger datasets, employing better optimization techniques, and integrating into real-world forensic workflow.

Keywords: CNN Model, Machine Learning, Deep Learning, Tamper Detection, Image classification.

1. Introduction

The digital age has completely changed communication and the sharing of information. With the ease of capturing, transmitting, and storing images across multiple platforms, visual data has become a central medium for expression, documentation, and

evidence. However, with all the advancements made, new challenges have emerged that should ensure digital evidence is not only real but accurate, especially in the sphere of images. The manipulative orientation of the image by advanced editing tools makes it not trustworthy or misleads individuals. In this way manipulation, critical evidence may be rendered useless, obstruction of justice illicit in crime investigations, and such would further spread misleading information and damage people's reputations worldwide.

The application of digital image forensics, which constitutes a science for recognizing manipulations in the digital images, is becoming more and more crucial to the solution of this problem. It deals with disclosing and analyzing digital evidence, determining alteration of contents and supplying the technical requisites for authenticity. It supports both the investigation of crimes and the assurance of trust in media, whether in journalism or simple daily digital interactions. Traditional forensic techniques, however, are based a lot on human knowledge, which sometimes is subject to subjectivity, takes time, and would not be very scalable, particularly when it comes to huge data sets.

As a result, researchers are increasingly exploring deep learning models as a powerful solution for forensic analysis. Convolutional Neural Networks (CNNs), in particular, have shown great promise in learning discriminative features directly from image data, thereby automating detection and improving accuracy. This study aims to design and improve CNN-based models capable of effectively handling diverse image manipulations, while also ensuring resilience against future and emerging threats in the digital domain. By leveraging computational power and large-scale datasets, CNN-based approaches can surpass human limitations, delivering faster, more objective, and highly reliable results, which makes them indispensable in modern forensic practice.

2. Related Works

Recently, the use of deep-learning models in image forensics studies has become much more pressing, given that they largely improve detection accuracy and gain

robustness against several modifications. Recent developments show that CNNs and other deep learning architectures are effective in detecting different kinds of image forgeries. [1] showed that lightweight deep learning models like MobileNetV2 when properly adapted, can detect copy-move forgeries generated under post-process attacks with brightness adjustment, blurring, noise addition, etc., in addition to executing geometric transformations such as cropping and rotation. This shows that deep learning models that are appropriately adapted perform quite well on these complex forgery scenarios.

Beyond forgery detection, deep-learning models are also being harnessed for specific forensic purposes such as face detection and recognition. [2] utilized a deep object detection scheme, namely YOLOv5, to effectively identify faces under different illumination conditions and partial occlusions, which is paramount in forensic facial recognition investigations. Similarly, [3] developed an ensemble deep learning fusion approach for the detection of colorization-based image forgery, achieving an improvement in detection accuracy of at least 2% over existing methods, thus demonstrating the strength of ensemble approaches with regards to forensic scenarios.

Additional recent work has further expanded this field. [4] proposed EFIMD-Net: Enhanced Feature Interaction and Multi-Domain Fusion Deep Forgery Detection Network. This model combines macro semantic features from raw RGB images and micro high-frequency tamper traces from SRM noise maps. The fusion of these domains improves the model's capacity to detect both subtle artifacts and large-scale manipulations; moreover, it includes a segmentation head to localize tampering regions beyond binary classification.

Another study [5] introduced MCDC-Net, which mixes central-difference convolution with vanilla convolution to better capture gradient and texture features. It integrates a multi-scale fusion module so that different image scales and locations contribute to forgery detection, enhancing robustness across varied scenes.

Ensemble approaches have also gained traction. [6] presented an ensemble method combining multiple pretrained models, including CNNs and Transformer-based architectures, using hard and soft voting schemes. Their experiments indicate a substantial improvement in forgery detection accuracy over using any single model alone.

Together, these studies reveal that current forensic research is moving toward models that fuse features across domains, use multiple scales, and employ ensemble strategies. These advances point to the importance of combining spatial, frequency, noise, and texture cues; adapting architectures for occlusion, blur,

and colorization; and going beyond mere classification to include localization of manipulated regions.

Lastly, recent research has begun to explore transformer-based methods for forensic detection. [7] proposed FakeFormer, a vulnerability-driven Vision Transformer designed to generalize across unseen datasets in deepfake detection. By incorporating attention mechanisms that focus on manipulation-prone regions of an image, the model significantly improves cross-dataset accuracy compared to conventional CNNs. Similarly, [8] introduced PV-ISM, a patch-based Vision Transformer for distinguishing synthetic images from authentic ones, which achieved superior results on generative datasets like CIFAKE and Stable Diffusion. These contributions highlight a clear shift toward attention-driven architectures that not only classify manipulated images but also adapt more flexibly to new forgery techniques.

Such innovations suggest that the field is entering a stage where hybrid and transformer-based models may become the new standard, complementing CNNs with broader context awareness and stronger generalization. This evolution is vital for forensic applications, where adversaries constantly develop more sophisticated manipulations that challenge traditional detection pipelines.

3. Methodology

The system was conceived in an organized and iterative fashion, specifically for the deployment of a deep-learning image classifier to detect manipulation concerning digital pictures. The VGG16 architecture forms the nucleus for the whole system, which was operated with transfer learning for better generalization and to enable the model to differentiate between authentic and tampered images. This choice was motivated by VGG16's well-established ability to extract hierarchical image features, capturing both low-level edges and high-level semantic representations. As a result, the system gives more confidence in performance against different manipulation techniques by exploiting the feature extraction capabilities of VGG16 while maintaining computational efficiency.

The intended workflow of the system begins when a user uploads an image for analysis. Once the image is received, it undergoes preprocessing steps such as resizing and normalization to prepare it for prediction. These operations ensure that all inputs conform to the model's expected format, reducing inconsistencies and improving prediction accuracy. The processed input is then passed to the trained CNN model, which classifies the image as either "Tampered" or "Original." This binary classification approach provides a direct and practical solution to image authenticity verification.

FastAPI forms the basis of the interactive web application, assisting in seamless user-model interaction. Users can upload images in real time through a simple yet intuitive web form that was deliberately designed for accessibility, even for individuals with limited technical expertise. After the user submits the input image, the backend handles the request by invoking the trained CNN model to generate predictions. The resulting output is displayed on a dedicated results page that clearly indicates whether the uploaded image is original or tampered. Additionally, the interface is accessorized with an option enabling users to return easily to the upload form for further testing, supporting multiple iterations of use.

In this way, the system covers the entire pipeline from deep learning model development to practical deployment. It not only demonstrates how CNN-based models can be adapted for forensic analysis but also illustrates how usability and accessibility can be incorporated to make such tools functional for real-world investigative tasks.

4. Implementation

The implementation stage was concentrating on the commencement of the actual system, which involves implementing a working prototype image manipulation detection application featuring a deep-learning model. The present work carried out the identification of manipulated images from authentic ones by employing pre-trained VGG16 architectures enhanced with transfer learning in a complete application. This choice was made because of the proven efficiency of VGG16 in extracting hierarchical visual features that are particularly useful in forensic analysis. Unlike many shallow learning techniques that rely on handcrafted features, VGG16 automatically learns patterns across multiple convolutional layers, capturing subtle pixel-level discrepancies that often arise from tampering. The architecture's depth and simplicity make it a reliable choice for research and practical applications alike, particularly when paired with transfer learning, which leverages prior knowledge from large-scale datasets such as ImageNet to adapt more effectively to forensic datasets of smaller size.

The dataset of tampered and authentic images was prepared in terms of preprocessing, which involved resizing each image and normalizing them to line up with what the model expects as input shape, thereby ensuring uniformity and improving the quality of learning. Such preprocessing minimizes distortions that could bias the model, while normalization helps maintain consistent pixel intensity distributions across all training samples. This step was vital for ensuring that the CNN received inputs in the same format it was designed to process, thereby reducing the risk of

misclassification.

Model training was performed using Python 3.11.5 in a Jupyter Lab environment, using libraries such as TensorFlow, Keras, NumPy, and OpenCV. These tools collectively provided a robust development platform: TensorFlow and Keras were used for designing, training, and evaluating the deep learning model, while NumPy and OpenCV facilitated image processing and numerical operations. The dataset was divided into three distinct subsets: training set, validation set, and testing set. This split allowed proper monitoring of learning progress and provided a reliable means of evaluating generalization to unseen data. Thus, the model was trained under the Adam optimizer, with hyperparameters tuned for stability, and its progress was tracked using performance metrics such as accuracy, precision, recall, and F1 score. These metrics made it possible to assess not only overall correctness but also the balance between false positives and false negatives, which is critical in forensic scenarios where both misclassifying authentic evidence as tampered and failing to detect tampered evidence could have significant consequences.

The model was then incorporated into a web application built on the FastAPI framework once its performance was considered reasonably reliable. The application features an interface where users upload images using their browsers, after which the image is processed at the backend and classified, with the prediction result appearing on a response page. To enhance usability, the interface was designed to be minimalistic and intuitive, ensuring accessibility for users regardless of technical expertise. This design embraces the entire route from model development to practical usage, making it an easy, lightweight, and effective instrument for forensic image analysis. Beyond its functionality, the deployment through FastAPI demonstrates the feasibility of translating a complex deep learning system into a deployable web service, which is essential for forensic practitioners who require quick and reliable tools. By bridging the gap between technical research and real-world application, the implementation proves the capability of CNN-based models to serve in practical scenarios and illustrates their potential to become integral components of digital forensic investigations.

5. Conclusion and Future Work

This project describes the design and implementation of an automatic system for tampering detection of an image by deep learning techniques. Through the use of the VGG16 architecture with transfer learning, the system successfully classifies images as tampered or authentic with great accuracy and reliability. The methodology started off with acquisition

and preprocessing of labeled data followed by systematic training and evaluation of the CNN model. Several important metrics such as accuracy, precision, recall, and F1 score were used for performance assessment, all of which confirmed the model's effectiveness in handling image manipulation detection.

It was not only technically implemented, but also stated to be user-friendly using FastAPI to deploy the trained model as a web-based application. The web interface is clearly and simply laid out, making it comfortable and easy to use. Users uploaded manipulated or non-manipulated images, real time feedback about their appropriate images were received by users as to whether the manipulated images were or were not manipulations. In this way, CNN model was seamlessly integrated into FastAPI backend, showing how complex machine learning techniques can be rendered into accessible toolkits for digital forensic investigators, researchers, and the rest of the public for digital content validation.

Future work involves the enhancement of the system through experimentation using advanced architectures such as ResNet, EfficientNet, all of which should provide gains in accuracy and robustness towards more complicated techniques of manipulation as potential contributions to future works. However, this would require collection, production, and use of larger, diverse collections of manipulated images to enhance generalization even more. Furthering cloud deployment would enhance scalability such that multiple users can access applications concurrently and enhance bulk data processing efficiency. Adversarial training and real-time monitoring may also increase resistance to new tampering methods, ensuring long-term adaptability and resilience. Expanding collaborative research, integrating multimodal data such as audio-visual evidence, and optimizing lightweight deployment on mobile devices may further broaden the relevance of this work in both professional forensic contexts and everyday user applications.

The success of integrating the trained model with a proper functioning user interface proves the capability of practical applications in digital forensics, ensuring reliability and usability across diverse investigative environments.

6. References

- [1] Muhammad N. A.; Mohammad Samar Ansari; Mamoona N. A. (2021). Lightweight Deep Learning Model for Detection of Copy-Move Image Forgery with Post-Processed Attacks. IEEE 19TH WORLD SYMPOSIUM ON APPLIED MACHINE.
- [2] S. Karakus; M. Kaya; S. Tuncer; M. Bahsi, Merve Açıkoğlu. (2023). A Deep Learning Based Fast Face Detection and Recognition Algorithm for Forensic Analysis. 10th International Symposium On Digital Forensics.
- [3] Shashikala D. K. (2023). Ensemble Deep Learning Fusion for Detection of Colorization Based Image Forgeries. 2ND International Conference For Innovation.
- [4] Cheng, H., Pang, W., Li, K., Wei, Y., Song, Y., & Chen, J. (2025). EFIMD-Net: Enhanced Feature Interaction and Multi-Domain Fusion Deep Forgery Detection Network. *Journal of Imaging*, 11(9), 312. DOI: 10.3390/jimaging11090312.
- [5] He, D., Jiang, Q., Jin, X., Cheng, Z., Liu, S., Yao, S., Zhou, W. (2024). MCDC-Net: Multi-scale forgery image detection network based on central difference convolution. *IET Image Process.* 18, 1–12. DOI: 10.1049/ipr2.12928.
- [6] S.-J. Ham, V.-H. Hoang and C.-S. Park (2024). Ensemble Approach for Image Recompression-Based Forgery Detection," in *IEEE Access*, vol. 12, pp. 196442-196454, DOI: 10.1109/ACCESS.2024.3521290.
- [7] Nguyen, D., Astrid, M., Ghorbel, E., and Aouada, D. (2024). FakeFormer: Efficient Vulnerability-Driven Transformers for Generalisable Deepfake Detection. DOI: 10.48550/arXiv.2410.21964.
- [8] Erdoğan, Ö., and Cömert, Z. (2025). Novel Deepfake Image Detection with PV-ISM: Patch-Based Vision Transformer for Identifying Synthetic Media. *Applied Sciences*, 15(12), 6429. DOI:10.3390/app15126429.

Session 12: e-Society

Title: Fusion Strategies of Convolutional Neural Networks and Vision Transformers: A Comparative Study on Binary-Class Cancer Classification

Authors: Aanuoluwa Oyebola Adio, Oluwaseun Favour Amore, Toluwase Ayobami Olowookere

Title: A Decision Support Framework for Predicting Soil Nutrients Based on Machine Learning Models and Sensors Technology

Authors: Taoheed Abioye Abdulraheem, Enesi Femi Aminu, Ayobami Ekundayo

Title: A Multi-Criteria Evaluation of Urban Transport System Performance in Makurdi Metropolis

Authors: Emmanuel O. Nwafor, Folake O. Akintayo, Inyeneobong E. Edem

Fusion Strategies of Convolutional Neural Networks and Vision Transformers: A Comparative Study on Binary-Class Cancer Classification

Aanuoluwa Oyebola Adio, Oluwaseun Favour Amore, Toluwase Ayobami Olowookere
Department of Computer Science, Redeemer's University, Ede, Nigeria

Abstract

The fusion of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) has shown strong potential in improving performance in medical image classification tasks. Several studies have leveraged the strengths of both architectures by employing various fusion strategies to enhance classification performance. However, within the context of cancer classification, limited research has been conducted on systematically comparing different fusion strategies, particularly early fusion, intermediate fusion, and late fusion. This study, therefore, investigates the comparative performance of these fusion approaches using the EfficientNet-B3 CNN model and the Data-Efficient Transformer model in cancer classification. Three publicly available cancer image datasets were utilized for binary-class classification tasks (brain tumor, oral cancer, and lung cancer). Each EfficientNetB3-DeiT fusion strategy was evaluated using standard performance metrics, including accuracy, sensitivity, specificity, AUC, and F1-score. Experimental results show that the intermediate fusion strategy outperforms other fusion approaches with 99.45% accuracy, 0.991 sensitivity, and 1.000 AUC on the brain tumor dataset; 82.66% accuracy, 1.000 sensitivity, and 0.868 AUC on lung cancer. On the oral cancer data, the late fusion strategy performed better than other strategies with 92.74% accuracy, 0.91 sensitivity, and 0.973 AUC. These findings offer valuable insights into selecting the most effective fusion strategy in various cancer classification scenarios, with potential implications for enhancing the development of robust classification models.

Keywords: Cancer, Binary Classification, Convolutional Neural Networks, Data Efficient Transformer, Fusion Strategies

1. Introduction

In recent years, deep learning techniques have led to notable improvements in medical image analysis, enhancing the accuracy and efficiency of disease diagnosis and the interpretation of complex medical

imaging data. The detection and classification of cancer in imaging data has seen significant gains, largely due to the ability of deep learning models to detect subtle patterns within various imaging modalities such as MRI, CT scans, and histopathological images. Convolutional Neural Networks (CNNs) have been central to this progress because of their strength in identifying local image features and learning layered representations of visual data. However, one of their main challenges lies in capturing global relationships across an image, which is often essential for understanding overall anatomical structures.

To bridge this gap, Vision Transformers (ViTs) have emerged as a complementary approach, offering the ability to model long-range interactions using self-attention mechanisms. This allows for a more comprehensive view of an image as a whole. Although ViTs generally require large amounts of data and are computationally intensive, they bring valuable contextual insight that CNNs may miss. As a result, combining CNNs and ViTs into unified models has become an appealing approach. These hybrid architectures leverage the precise local feature extraction of CNNs alongside the global interpretability provided by Transformers. This combination has proven particularly effective in challenging applications such as cancer detection, where recognizing both fine details and broader image context is crucial for reliable classification. The combination of the two architectures usually follows some approaches, including low-level feature fusion, intermediate feature fusion, and late feature fusion. This study focuses on comparing the utilization of these three common combination strategies on the binary classification of cancer imaging data.

2. Related Work

Recent advancements in deep learning have spurred a growing interest in hybrid models that combine Convolutional Neural Networks (CNNs) with Vision Transformers (ViTs) for medical image analysis. These models aim to exploit the local feature

extraction strengths of CNNs and the global representation capabilities of ViTs. Within the domain of cancer diagnosis, hybrid architectures have demonstrated encouraging performance; however, a lack of structured evaluation across fusion strategies, particularly for binary classification tasks.

Several hybrid models have been proposed across cancer types, such as liver [9], breast [4], lung [11], and cervical cancer [8], each employing specific fusion strategies like sequential, parallel, or intermediate fusion.

Hybrid CNN–Transformer models have demonstrated significant promise in breast cancer research, particularly in improving the accuracy of binary classification tasks. One study introduced a dual-stream network that integrates an Xception-based CNN with a Transformer module, achieving strong performance across varying image magnifications (Zeynali et al., 2024). Another investigation developed a fusion architecture using EfficientNetV2 and Vision Transformer to classify histopathology images, reporting high accuracy, although the model was not assessed across different fusion strategies or cancer types [4]. A similar hybrid combining CNN, Transformer, and bi-LSTM components achieved enhanced diagnostic precision, yet the work focused solely on a single fusion configuration [6]. While another study explored CNN–ViT integration for breast cancer subtype identification, it was confined to a multi-class setting, limiting its relevance to binary tasks [5]. In contrast, a separate effort examined CNN performance with various optimization strategies but did not incorporate ViT elements or hybrid design principles [3].

Beyond breast cancer, researchers have also applied hybrid models to other types of cancer. A notable example is the use of a CNN–Transformer network for liver cancer detection, which illustrated the diagnostic strength of fusion models in non-invasive liver assessments; however, the study did not compare different levels of feature integration [9]. In lung cancer research, a hybrid model was introduced to predict lymph node staging and patient survival from CT and clinical data, though the focus on prognosis rather than binary classification limited its applicability to this domain [11].

Efforts in cervical cancer detection have also incorporated deep fusion models. A CNN–Transformer–GRU architecture was developed to identify cervical lesion grades using multimodal inputs. While it demonstrated strong performance in multi-class prediction, the model did not examine fusion depth within its imaging framework [8].

In ultrasound-based analysis, a study utilizing a ViT-only model for breast lesion classification demonstrated the feasibility of Transformer-based solutions, but the exclusion of CNN components limited insights into the comparative benefits of hybrid systems [2]. Similarly, a Transformer

ensemble employing multi-scale attention was applied to skin cancer diagnosis, yielding competitive results, although it did not incorporate any CNN pathways [13]. In another dermatological application, fused features from a foundation model and a ViT were employed for lesion analysis, yet the study did not explore CNN–ViT integration or target binary classification specifically [7].

However, despite the progress shown, there is a lack of systematic comparison among early, intermediate, and late fusion techniques, limiting insights into optimal integration methods. Additional studies [5], [6], either exclude ViTs or evaluate CNNs and ViTs separately, underscoring the research gap. The absence of direct comparisons across fusion strategies presents a significant limitation in designing robust and generalizable models for binary-class cancer classification. This study, therefore, explores different fusion strategies (early, intermediate, and late) in binary-class cancer classification.

3. Model Architectures

Two models were utilized in this study: EfficientNet-B3 and the Data Efficient Transformer (DeiT). EfficientNet-B3 is a CNN model that is built to achieve high accuracy with little computational complexity. It is built on Efficient-B0, which was discovered using neural architecture search (NAS), and expands its capacity through a systematic scaling approach. Unlike traditional methods that scale network depth, width, or image resolution independently, EfficientNet introduces a compound scaling method. This method balances all three dimensions simultaneously based on a compound coefficient ϕ . The scaling relationships are governed by:

$$\begin{aligned} \text{Depth: } d &= \alpha^\phi, \text{Width: } w = \beta^\phi, \\ \text{Resolution: } r &= \gamma^\phi \end{aligned}$$

where α , β , and γ are constants determined through grid search, and ϕ control how much the model is scaled. In EfficientNet-B3, this scaling with $\phi = 3$ results in a deeper, wider network with higher input resolution when compared to B0. The core building block of the model is the MBConv module, a mobile inverted bottleneck convolution, enhanced with a Squeeze-and-Excitation (SE) mechanism. This module allows the network to emphasize informative features while suppressing irrelevant ones by adaptively recalibrating channel-wise responses. Mathematically, this recalibration is expressed as:

$$SE(x) = x \cdot \sigma(W_2 \cdot ReLU(W_1 \cdot GAP(x)))$$

where GAP denotes global average pooling, σ is the sigmoid activation and W_1, W_2 are the learnable weights.

EfficientNet-B3 provides a strong balance between performance and resource usage, making it suitable for tasks requiring accuracy under limited computational budgets. The architecture of EfficientNet-B3 is shown in Figure 1.

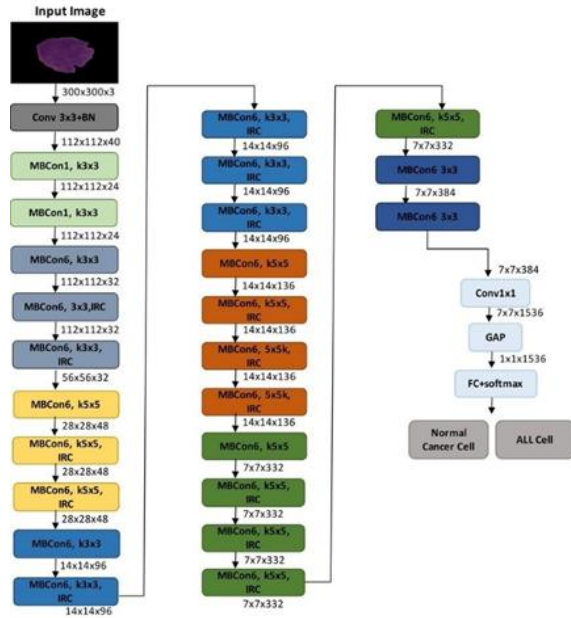


Figure 1. Architecture of the EfficientNet-B3

The Data-efficient Image Transformer (DeiT) is a vision-specific adaptation of the Transformer architecture, originally developed for language tasks, and reengineered to perform competitively on image classification with far less data. One of the core innovations of DeiT is its ability to train effectively on standard datasets like ImageNet-1k, eliminating the need for extremely large-scale pre-training datasets used by earlier vision transformers. The model starts by dividing an image into fixed-size patches, typically 16×16 pixels, flattening each patch, and projecting it into an embedding space using a linear transformation. These patch embeddings are arranged into a sequence, prefixed with a learnable classification token and augmented with positional encodings:

$$z_0 = [x_{class}; Ex_1; Ex_2 \dots; Ex_N] + E_{pos}$$

This sequence is passed through multiple layers of self-attention and feedforward networks, forming the backbone of the model:

$$z_l = MLP \left(LN \left(MSA \left(LN(z_{l-1}) \right) \right) + z_{l-1} \right) + z_{l-1}$$

A key contribution of DeiT is the use of a distillation token during training, which allows the model to learn not just from true labels but also from a teacher model, often a well-trained CNN. This helps

the transformer learn more efficiently and generalize better, even with limited data. The final classification is derived from the output of the class token, which accumulates global information through attention layers.

Building a hybrid model by combining Efficient-B3 and DeiT-Tiny harnesses the complementary strengths of convolutional and transformer-based models. Efficient-B3 captures the local features while DeiT-Tiny captures the global features. Combining these models creates richer and more expressive features. The architecture of DeiT-Tiny is as shown in Figure 2.

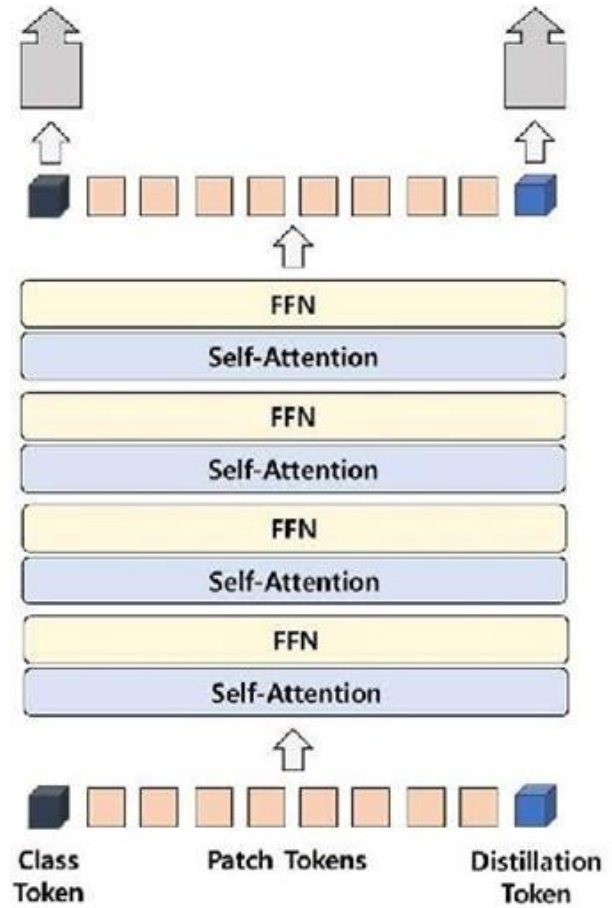


Figure 2. Architecture of the DeiT-Tiny

4. Materials and Methods

This section introduces the data used in this study, the algorithms, and the performance evaluation metrics used.

4.1. Dataset Description

Three publicly available datasets were used for this study. The Brain Tumor Dataset [10] contains 4600 images with 2087 normal cases and 2513 cases

of brain tumor. The Oral Cancer Dataset [1] contains 1238 images with 685 cases of oral cancer and 553 normal cases. The Lung Cancer MRI dataset contains 3680 images [12], classified into cancer and no-cancer cases. The dataset has 1806 no-cancer cases and 1874 Cancer cases. The models used 70% of the data for training, 10% for validation, and 20% for testing. No augmentation technique was used in this study. Figure 1 shows the samples from each of the datasets used in this study.

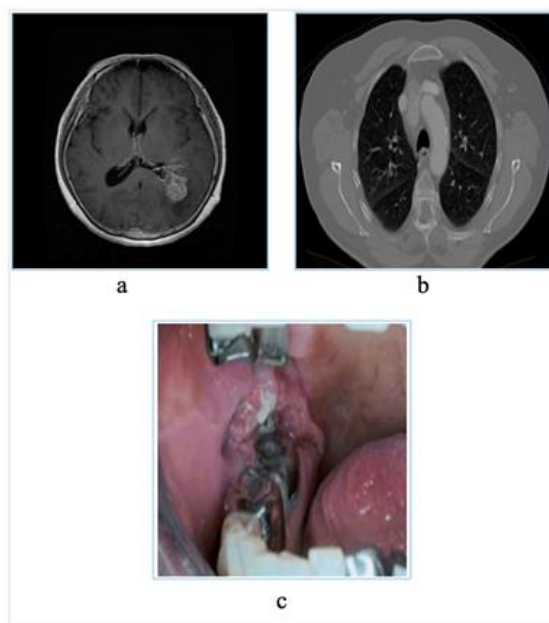


Figure 3. Samples of images in the datasets. (a) Brain Tumor, (b) Lung Cancer, and (c) Oral Cancer

4.2. Methodology

The approach used in this study is the hybridization of pre-trained EfficientNet-B3 and DeiT-Tiny models. Three feature fusion techniques are explored for the purpose of comparison. The low-level (early) feature fusion, mid-level feature fusion, also called the intermediate feature fusion, and the late feature fusion techniques are the fusion techniques explored in this study. The low-level feature fusion fuses feature from the early layers of the EfficientNet and DeiT architectures for binary cancer image classification. The models used were pre-trained on ImageNet, and so the early features were extracted by setting the features-only parameter to be true. The early features extracted contain low to mid-level features such as the edges, textures, and shapes. Each of the set of extracted features are projected to a common embedding size of 128. 1×1 convolutions were then used to reduce dimensionality. The feature maps were spatially pooled into a vector and produced the shape (16,128) from each of the models. The two

vectors produced from the models are then concatenated into a 256 – dimensional vector. A single linear layer then maps the fused vector to predictions for the final classification into cancer or normal. The low-level feature fusion combines the shape features and local texture features from EfficientNet-B3 with the global features from DeiT.

The mid-level feature fusion approach was also utilized in this study to merge the strengths of both models. The high-level features from both models were merged just before the classification stage. For the EfficientNet branch, the output from the last convolution block is extracted. This layer contains the rich spatial features and has 1536 channels. At the same time, the last classification head from the DeiT-Tiny branch is removed. The CLS features, which are a 192-dimensional embedding, are extracted. This approach ensures that the fusion benefits from both the local features and the global features. To ensure a seamless mid-level fusion, the output from EfficientNet-B3, which has 1536 channels, is passed through a 1×1 convolution to reduce the channel depth. This is followed by batch normalization, ReLU activation, and adaptive average pooling to get a fixed-size feature vector. The output from the DeiT-Tiny model is linearly transformed, followed by batch normalization, ReLU, and a dropout rate of 0.3 is applied to prevent overfitting. The mid-level feature fusion differs from the low-level feature fusion, which obtains only the shallow features from the early convolutional layers.

The late feature fusion approach combines the final features extracted from individual models. The models process the input images separately, resulting in the extraction of rich local features by EfficientNet-B3 and rich global features from DeiT before their outputs are combined. The classification head of EfficientNet-B3 is removed, thus giving access to the final rich semantic and spatial features. The classification head from the DeiT-Tiny model is also removed. The feature vector from EfficientNet is passed into a 256-dimensional space using a fully connected layer. The 192-dimensional vector from DeiT-Tiny is also passed into a 256-dimensional space through a separate linear layer. This is done so that the two vectors are compatible for fusion. The vectors are then concatenated to form a 512-dimensional feature vector. This is then passed through a classification head, which has a linear transformation layer, ReLU activation, batch normalization, dropout with a rate of 0.3, and then a final layer that produces the final classification outputs. The framework followed in this study is as depicted in Figure 4. The process starts from the split of the dataset into training and testing, which is then followed by utilizing different fusion strategies on the data.

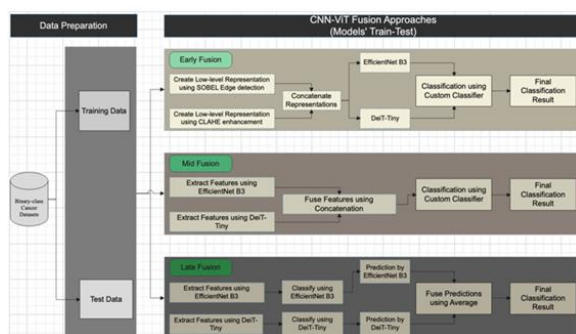


Figure 4: Study's Workflow

5. Results and Discussion

Our study compared the performance of three fusion strategies for the hybridization of EfficientNet-B3 and DeiT-Tiny. The performance of the fusion strategies were assessed using accuracy, specificity, sensitivity, F1-Score, and AUC-ROC. Table 1 shows the performance of the fusion strategies on the brain tumor data for binary classification. Here, in accuracy, mid fusion strategy performed better with 99.45%, followed by late fusion with 99.23% and then the early fusion with 98.9%. For the Sensitivity metric, late and early fusion strategies performed better with 0.9928, followed by the mid fusion strategy of 0.9905. The mid fusion strategy leads in the specificity metric with 0.9979, followed by late fusion with 0.9918, and then early fusion with 0.9856. The mid fusion strategy also leads in the AUC metric with 1.000, followed by early fusion with 0.9993, and then late fusion with 0.9975.

For the Oral Cancer data, as shown in Table 2, late fusion leads in the accuracy, specificity, F1-Score, and AUC with 92.74%, 0.9416, 0.9182, and 0.9729, respectively. However, in the sensitivity metric, late fusion and early fusion strategies recorded 0.9099, coming behind mid-fusion with 0.9279.

On the lung cancer data for binary classification, mid fusion strategy performed better for accuracy, sensitivity, specificity, and F1-score. It however performed lower than the other fusion strategies for AUC metric.

Table 1. Classification performance of the fusion strategies on the brain tumor data

	Acc(%)	Sens.	Spec.	F1-Score	AUC
Late Fusion	99.23	0.9928	0.9918	0.9917	0.9975
Mid Fusion	99.45	0.9905	0.9979	0.994	1
Early Fusion	98.9	0.9928	0.9856	0.9881	0.9993

Table 2. Classification performance of the fusion strategies on the oral cancer data

	Acc(%)	Sens.	Spec.	F1-Score	AUC
Late Fusion	92.74	0.9099	0.9416	0.9182	0.9729
Mid Fusion	88.71	0.9279	0.854	0.8803	0.9698
Early Fusion	91.53	0.9099	0.9197	0.9058	0.9713

Table 3. Classification performance of the fusion strategies on the lung cancer data

	Acc(%)	Sens.	Spec.	F1-Score	AUC
Late Fusion	81.83	0.9945	0.6407	0.8461	0.8783
Mid Fusion	82.66	1	0.6518	0.8528	0.8683
Early Fusion	82.52	0.9972	0.6518	0.8514	0.8783

6. Conclusion

This study compared the performance of different fusion strategies (early, intermediate, and late) for the detection and classification of different cancers (brain, oral, and lung) using X-ray images and deep learning techniques. The result of the experiment showed that for different cancer types, no single fusion strategy can be said to perform better than the other. The performance of the fusion techniques varies from disease to disease. However, several techniques can be employed to further enhance the performance of mid or intermediate fusion, as it consistently demonstrates strong performance across the datasets used in the experiment. Further studies are needed to evaluate the effectiveness of these strategies in multi-classification of diseases.

7. References

- [1] Atef, M. (2014). Oral Cancer Images for Classification. Kaggle. <https://www.kaggle.com/datasets/muhammdatef/oral-cancer-images-for-classification> (Access Date: 13 July, 2025).
- [2] Gheflati, B., & Rivaz, H. (2022). Vision Transformers for Classification of Breast Ultrasound Images. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2022-July, 480–483. DOI: 10.1109/EMBC48229.2022.9871809.

- [3] Güler, M., & Namlı, E. (2024). Brain Tumor Detection with Deep Learning Methods' Classifier Optimization Using Medical Images. *Applied Sciences (Switzerland)*, 14(2), 642. DOI: 10.3390/app14020642.
- [4] Hayat, M., Ahmad, N., Nasir, A., & Ahmad Tariq, Z. (2024). Hybrid Deep Learning EfficientNetV2 and Vision Transformer (EffNetV2-ViT) Model for Breast Cancer Histopathological Image Classification. *IEEE Access*, 12, 184119–184131. DOI: 10.1109/ACCESS.2024.3503413.
- [5] Jahan, I., Chowdhury, M. E. H., Vranic, S., Al Saady, R. M., Kabir, S., Pranto, Z. H., Mim, S. J., & Nobil, S. F. (2025). Deep learning and vision transformers-based framework for breast cancer and subtype identification. *Neural Computing and Applications*, 37(16), 9311–9330. DOI: 10.1007/s00521-025-10984-2.
- [6] Lilhore, U. K., Sharma, Y. K., Shukla, B. K., Vadlamudi, M. N., Simaiya, S., Alroobaea, R., Alsafyani, M., & Baqasah, A. M. (2025). Hybrid convolutional neural network and bi-LSTM model with EfficientNet-B0 for high-accuracy breast cancer detection and classification. *Scientific Reports*, 15(1), 12082. DOI: 10.1038/s41598-025-95311-4.
- [7] Mahbod, A., Ecker, R., & Woitek, R. (2025). Fusion of Foundation and Vision Transformer Model Features for Dermatoscopic Image Classification. *ArXiv Preprint ArXiv:2505.16338*.
- [8] Sha, Y., Zhang, Q., Zhai, X., Hou, M., Lu, J., Meng, W., Wang, Y., Li, K., & Ma, J. (2024). CerviFusionNet: A multi-modal, hybrid CNN-transformer-GRU model for enhanced cervical lesion multi-classification. *IScience*, 27(12). DOI: 10.1016/j.isci.2024.111313.
- [9] Vats, S., Singh, C. R., & Mehta, S. (2024). Next-Generation Diagnostic Tools: The Role of Hybrid CNN-Transformer Models in Liver Cancer Detection. 2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024, 1–5. DOI: 10.1109/ICCCNT61001.2024.10723884.
- [10] Viradiya, P. (2021). Brain Tumor Dataset. Kaggle. <https://www.kaggle.com/datasets/preetviradiya/brian-tumor-dataset> (Access Date: 13 July, 2025).
- [11] Wang, L., Zhang, C., & Li, J. (2024). A Hybrid CNN-Transformer Model for Predicting N Staging and Survival in Non-Small Cell Lung Cancer Patients Based on CT-Scan. *Tomography*, 10(10), 1676–1693. DOI: 10.3390/tomography10100123.
- [12] Xiao_Peng_Zhang12. (2024). Lung Cancer MRI Images. Kaggle. https://www.kaggle.com/datasets/xiaopen_gzhang12/lung-cancer-mri-images (Access Date: 13 July, 2025).
- [13] Yang, G., Luo, S., & Greer, P. (2025). Boosting Skin Cancer Classification: A Multi-Scale Attention and Ensemble Approach with Vision Transformers. *Sensors*, 25(8), 2479. DOI: 10.3390/s25082479.
- [14] Zeynali, A., Tinati, M. A., & Tazehkand, B. M. (2024). Hybrid CNN-Transformer Architecture With Xception-Based Feature Enhancement for Accurate Breast Cancer Classification. *IEEE Access*, 12, 189477–189493. DOI: 10.1109/ACCESS.2024.3516535.

A Decision Support Framework for Predicting Soil Nutrients Based on Machine Learning Models and Sensors Technology

Taoheed Abioye Abdulraheem¹, Enesi Femi Aminu², Ayobami Ekundayo³
Department of Computer Science, Federal University of Technology Minna, Nigeria^{1,2,3}

Abstract

The rise in humans' population globally necessitates equal efforts for sufficient food security. In order to achieve this, which is in line with sustainable development goal (SDG) 2 of zero hunger, already literatures have firmly established the significance of soil nutrients despite the threat of dynamic environmental factors. Despite the application of sensor technologies and artificial intelligence for effective soil nutrient management, there is still a shortfall in the accuracy of sensors, unbalanced analysis of essential micronutrients, and inadequate incorporations of dynamic environmental factors such as rainfall and temperature. Therefore, this research designed a system for advanced soil nutrient management that can overcome the deficits by employing support vector machine (SVM) and Gradient Boosting algorithms as proposed models for training the ground truth data that will be obtained via the deployment of rainfall and temperature sensors, Nitrogen Potassium and Phosphorus (NPK) sensors e.t.c. There was promising result from both models when using 4 machine learning models to build the system and it was found that after applying hyper-parameter tuning technique to get better accuracy and not face overfitting SVM has the highest accuracy with 71%, precision, recall and F1-score of 0.73, 0.70, 0.71 respectively. Similarly, the system developed will work in offline mode that reduces dependency on continuous high-quality internet connectivity for effective use in diverse agricultural environments, including poorly connected rural areas.

Keywords: Soil Nutrients, Machine Learning Models, Sensors, Fertilizer Recommendation, Offline Mode .

1. Introduction

Modern agriculture faces the formidable task of ensuring global food security for a rising population, while dealing with finite land and the worsening effects of climate change. The effective management of organic soil nutrients is critical to addressing this,

as soil is the foundation of plant life, supplying necessary nutrients, water, and support [1]. Crops need the right balance of these essential nutrients to grow well and produce a good harvest [2]. Synthetic chemical fertilizer use has risen dramatically over time as a direct result of the drive to increase food production, which has considerably enhanced crop yields. For example, India's consumption of fertilizer skyrocketed from about 70 thousand tons in the early 1950s to close to 30 million tons by 2021-2022, making a major contribution to its increased food grain supply [3]. The widespread and often imprecise use of fertilizers has led to severe environmental repercussions. These include degraded soil from nutrient imbalances and acidification, water pollution from runoff and leaching that causes eutrophication, increased greenhouse gas emissions, and significant threats to biodiversity and the health of entire ecosystems [3]. A major global agricultural challenge is the uneven distribution of nutrients. Many regions have an excess of nitrogen (N) and phosphorus (P), which harms the environment, while other areas, like Africa, suffer from a shortage that depletes soil and degrades its health over time [4]. As a result, agriculture today is tasked with both enhancing food output to satisfy global needs and advancing environmental stewardship by minimizing harm and making the most of its resources [3]. Soil nutrient analysis is a cornerstone of sustainable farming worldwide, including in Nigeria, as it is essential for determining the soil's fertility and its ability to support healthy crops [25]. To grow well, resist pests and diseases, and produce a good, high-quality harvest, plants need the right mix of both major nutrients (like Nitrogen and Phosphorus) and minor nutrients (like Zinc and Iron) [6]. Soil analysis provides the critical data needed to understand the availability of these nutrients in the specific context of a field [24]. Since fertilizers represent a significant cost, conducting regular and accurate soil analysis is a crucial economic strategy for farmers to optimize their profitability. To address these challenges, a major shift is happening towards new approaches like Precision Agriculture (PA) and the wider concept of Agriculture 4.0 [2]. These approaches leverage

advanced technologies to manage spatial and temporal variability within fields, enabling data-driven, site-specific management decisions [8]. The objective is to increase efficiency and cut down on waste by precisely applying resources, such as fertilizers and water, only when and where they are actually required [2]. At the heart of Precision Agriculture and Agriculture 4.0 are smart technologies designed to collect and interpret data. This includes advanced sensors for real-time tracking, the Internet of Things (IoT) for data transfer, AI and Machine Learning to analyze and predict trends, and Decision Support Systems (DSS) to give farmers clear, actionable recommendations [9] [10]. Soil sensors offer immediate data on key parameters like nutrient levels, moisture, and pH, which addresses the slowness of traditional lab tests and the challenge of soil and crop variations within and between fields [11]. Precision Agriculture (PA) moves beyond traditional methods by managing field variability to improve both sustainability and efficiency. Its core goal is to optimize resources by precisely applying inputs like fertilizers and water only where, when, and in the amount needed. This boosts crop yield and quality while cutting costs and reducing environmental harm [12], [13]. A central aim of PA, and a major focus within recent research, remains the optimization of soil nutrient management [14]. However, several issues hinder adoption in rural areas: the models' reliability is limited by sensor accuracy and a narrow focus on only NPK, while the systems' dependence on IoT is impractical where connectivity is weak [15]. Current soil management systems are limited by sensor errors, a narrow focus on only NPK nutrients, and a reliance on strong internet that makes them impractical for many rural farms. To solve this, a new generation of models is needed. This study contributes to that goal by developing a machine learning model that can predict a soil's nutrient needs by analyzing both soil characteristics and changing environmental data, creating a more robust and accessible tool. This approach aims to create a scalable, easy-to-use tool that lessens the need for extensive infrastructure, ultimately supporting UN Sustainable Development Goals 2 (Zero Hunger) and 9 (Industry, Innovation, and Infrastructure). The remaining section of this research is organized as follows. Section two: related works, section three the research methodology, section four results and discussion, and section five conclusion.

2. Related Works

In an effort to enhance agricultural precision and food security, [16] explored the use of machine learning (ML) for crop recommendations in India. By analyzing soil nutrients, pH, and climate factors, the research compared five ML algorithms. The XGBoost

model emerged as the superior choice, delivering over 98.5% accuracy across all tested categories (agricultural, horticultural, and mixed). This success demonstrates that ML can provide highly precise, data-driven recommendations to optimize crop selection, improve fertilizer management, and increase overall farming efficiency. Examining the role of nanotechnology in smart agriculture, highlighting its ability to increase productivity, optimize resource use, and promote sustainability. S. Garg, N. P. Rumjit, and S. Roy [17] Addresses the inefficiencies and environmental harm of traditional farming, the research discusses precision technologies like IoT and AI, with a special focus on nano-applications. These include nano-encapsulated fertilizers, nano-pesticides, and biosensors for nutrient management and pathogen detection. The findings confirm that nanotechnology-based solutions improve nutrient uptake, enhance plant growth, and reduce pollution, while their precise application leads to higher efficiency, less waste, and better soil health. The use of AI and machine learning in soil analysis is essential for boosting agricultural productivity and sustainability. Traditional soil testing is often slow and expensive, hindering farmers' ability to make timely decisions. M. Awais et al. [18] Addressed these issues by applying machine learning algorithms (including Random Forest, SVM, and Neural Networks) and AI-powered imaging to predict key soil characteristics like pH, moisture, and nutrients. The study found that these AI-driven methods provide highly accurate, real-time assessments of soil fertility, enabling more informed choices for applying fertilizer and selecting crops. Addressing the shortcomings of traditional crop recommendation methods, which frequently cause suboptimal yields and inefficient resource allocation, a study by [24] explored the use of machine learning models. Their goal was to enhance the precision of crop selection by analyzing a comprehensive set of environmental and soil conditions. The researchers applied several machine learning algorithms, including Cat Boosting (C-Boost), Gaussian Naïve Bayes (GNB), and Support Vector Machine (SVM), to a Kaggle dataset containing information on soil nutrients (N, P, K), temperature, humidity, pH, and rainfall. After balancing the data using SMOTE, their results demonstrated that the Cat Boosting algorithm significantly outperformed the others, achieving a remarkable accuracy of 99.51%. This AI-based system effectively identifies the most suitable crops, leading to enhanced productivity and more efficient management of resources. Examining how the Internet of Things (IoT) enhances productivity and sustainability in smart agriculture. D. J. Reddy and M. R. Kumar [20] addresses the inefficiencies of traditional farming, the research explores various IoT applications like smart irrigation, crop monitoring, and supply chain management, which are powered by

AI, sensors, and cloud computing. The findings demonstrate that IoT significantly boosts agricultural efficiency by enabling real-time monitoring and automated decision-making. This leads to proven benefits such as reduced water waste, higher crop yields, and improved farm management, contributing to a more sustainable agricultural system. Advanced technologies like the Internet of Things (IoT) and machine learning (ML) are becoming crucial for improving agricultural nutrient management. For instance, a smart IoT system described by [21] uses sensors and cloud storage for real-time monitoring of soil properties like pH and moisture in gardens. This system offers a cost-effective and highly accurate alternative to traditional lab tests. On a broader scale, ML models such as Random Forest and Convolutional Neural Networks are effective at analyzing soil, climate, and crop data to provide precise fertilizer recommendations, helping to optimize usage and protect the environment. However, challenges like high data needs, model complexity, and a lack of standards currently hinder widespread use. Future efforts must focus on better data collection, creating more adaptable models for diverse farms, and integrating these technologies into practical farming operations to enhance efficiency and sustainability. Blessing Sheeba et al. [22] introduces a machine learning system for soil analysis that uses the Extreme Learning Machine (ELM) algorithm to classify micronutrients. By integrating IoT sensors that measure parameters like pH and macronutrient levels (N, P, K), the model provides real-time data to enhance agricultural decision-making. The research found that soil in Tamil Nadu generally contains sufficient macronutrients but is deficient in specific micronutrients, indicating a need for more targeted fertilization. Key limitations of the model include its difficulty adapting to environmental changes and its reliance on large, varied datasets. The authors suggest that future work should aim to expand data collection, include more environmental variables, and improve the cost-efficiency of the IoT system to encourage wider adoption in precision farming. Exploring combining machine learning with spectroscopic techniques to more accurately and efficiently predict soil nutrients. J. Trontelj Ml and O. Chambers [23] demonstrated that among the various models tested, Least Squares Support Vector Machine (LS-SVM) and Artificial Neural Network (ANN) models were the most effective at predicting key nutrients like phosphorus, potassium, and magnesium. The authors emphasize the advantages of classifying multiple components at once and the necessity of a consistent dataset for optimal model performance. Nonetheless, they identify ongoing challenges, including dataset variability, environmental factors, and the high computational power required by some models. For broader use in precision agriculture, future work should concentrate on refining model calibration,

enhancing real-time adaptability, and making the models more computationally efficient.

3. Methodology

The methodology for this research is centered on the design, development, and evaluation of a novel Soil Nutrient Decision Support System (DSS). The system integrates hardware for in-situ data acquisition, cloud-based data processing, and machine learning models to provide accurate, real-time fertilizer recommendations. The framework is composed of the system's conceptual architecture and the underlying algorithms that drive its analytical capabilities.

Input: t

Output: se, hy, ph

Parameters: Soil Nutrient Dataset (t), FieldID(f), CropType(ct), SessionContext(sc), temperature(te), Humidity(hy), PH(ph)

```

1. Input t
2. Initialize fi, ct, se
3. While
4.    $If\ te - \frac{1}{\sigma(x)} = \varphi + \sigma \cdot \ln(R) + \tau \cdot \ln(R)^3 > 272$ 
5.     Accept
6.   Else
7.     Print(error)
8.      $If\ 0 > hy - \frac{t_{hy} - t_{max\_hyd}}{t_{max\_hyd} - t_{min\_hyd}} \cdot 100 < 100$ 
9.       Accept
10.    Else
11.      Print(error)
12.       $If\ pH - \frac{t_{pH} - t_{max\_pH}}{t_{max\_pH} - t_{min\_pH}} \cdot 100 < -14$ 
13.        Endif
14.      Endif
15.    Endif
16.  Return se, hy, ph
17. WEND

```

Figure 1. Algorithmic Representation of the Model

The process begins with the collection of multi-source data. A suite of sensors integrated into a portable "Smart Rod" is used for in-situ measurements. These sensors gather real-time data on primary macronutrients (NPK), soil temperature, and humidity directly from the field. This sensor data is augmented with manually inputted contextual information, including soil pH, existing fertilizer data, and local weather data. The system's logic is governed by a well-defined algorithm and a mathematical model that translates raw data into a precise fertilizer recommendation. **Algorithmic Workflow** The operational logic of the DSS is captured in a sequential algorithm, which proceeds through five distinct phases: **Initialization and Setup:** Upon activation, the system initializes its hardware components. The user (farmer) is prompted to input essential context for the session, such as the FieldID, CropType, and CropStage. This information is stored as the sessionContext to tailor the subsequent

analysis. Iterative Measurement Loop: The system enters a measurement loop, allowing the farmer to take readings from multiple representative spots in a field. For each spot, the device acquires a full suite of raw data from its sensors (e.g., NPK indicators, soil moisture, temperature). This currentRawSpotData is added to a list of measurements for the session. The loop terminates when the farmer indicates that data collection is complete. The system prompts the farmer for new reading and triggers a while loop that collects 3 environmental properties, temperature sensor measures and records the current soil temperature value greater than 273K while values lower than this returns error on the LED display. The humidity is measured and calculated to determine where the current reading, $e_{measured}$, falls relative to two established benchmarks: a value for a completely dry state, e_{dry} , and one for a completely wet or saturated state, e_{wet} . The logic dictates that if the resulting hy value falls within the physically meaningful range of 0 to 100, the measurement is accepted. However, if the calculated percentage is outside of these bounds, either below zero or above one hundred, it signifies a likely error. The final reading is done by the pH sensor which is determined using a formula derived from the Nernst equation, which relates a measured voltage potential to an ion concentration. The difference between these two potentials is then divided by a constant, 0.05916, which is the Nernst factor for hydrogen ions at standard temperature. After the pH is calculated, it undergoes a simple validation check. The logic verifies if the calculated pH is less than or equal to 14, which is the standard upper limit for the pH scale in most aqueous solutions. If the value falls within this valid range, it is accepted. However, if the calculated pH exceeds 14, it is considered physically implausible for a typical scenario, and the system prints an error to flag the result as invalid. The if statements and while loops are terminated.

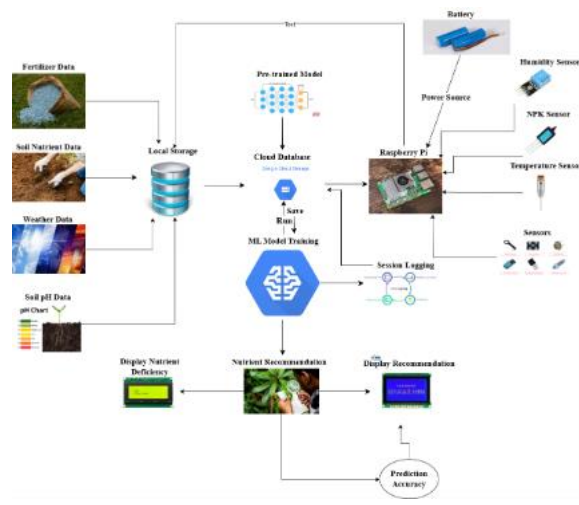


Figure 2. Conceptual Framework of the proposed system

The conceptual framework for the proposed intelligent fertilizer recommendation system is built upon an integrated cyber-physical architecture, designed to optimize agricultural nutrient management. The system operationalizes data from a suite of in-situ sensors, including units for measuring soil NPK (Nitrogen, Phosphorus, Potassium) levels, ambient temperature, and humidity. This primary sensor array is managed by a central processing unit, such as a Raspberry Pi, which serves as the edge computing device and is powered by a dedicated battery source to ensure field deployment viability. The system's workflow begins with data acquisition. The collected real-time sensor data, complemented by supplementary data on soil pH and historical weather patterns, is initially stored locally. This data is subsequently synchronized with a cloud-based infrastructure, utilizing services like Google Cloud Storage for robust and scalable data persistence. A critical function of the system is the "Save Run" module, which triggers the training and refinement of machine learning (ML) models. The analytical core of the framework is the "Prescribed Model" module, which encompasses a flexible modeling environment. This module can deploy a pre-trained model for immediate inference or generate new, context-specific models. The architecture supports both proposed models (designed from established algorithms) and reconstructed models (fine-tuned based on new local data), ensuring adaptability and improved prediction accuracy over time. Finally, the system concludes its pipeline by generating actionable insights. The Raspberry Pi unit handles session logging for auditability and presents the final fertilizer recommendation to the end-user via a display interface. The performance metric of "Prediction Accuracy" is a key output, validating the efficacy of the prescribed model and ensuring the recommendations are both data-driven and reliable for enhancing crop productivity.

4. Results and Discussion

Primarily this work considered two models both which are support vector machine and XGBoost algorithms. The dataset was obtained from a GitHub repository Figure 2 depicts a snapshot of the dataset to be used in this study.

	A	B	C	D	E	F	G	H
1	Temperature	Rainfall	pH	Rh	Nitrogen	Phosphorus	Potassium	Fertilizer Name
2	20.88752278	747.8607649	6.571547525	91.19719624	170.8003814	118.670098	243.3321196	Urea
3	18.06272088	711.1043288	6.251806204	91.9396234	179.2903639	121.0202444	246.9103776	DAP
4	16.77678226	774.0382468	6.346911673	91.38728558	181.4407323	116.9368055	242.6996007	14-35-14
5	14.28099975	665.6335057	6.299598485	91.25459751	176.1652818	122.2331535	237.0968924	26-28
6	21.44449045	806.5314549	6.384367838	92.35482921	182.935334	126.0882339	243.8603636	Urea
7	18.65531941	717.3281878	6.481890889	93.00912341	182.2372366	122.6240121	239.4596554	17-17-17
8	16.79031056	779.6299301	6.413392954	92.5298004	163.2464915	123.3202196	241.4604529	20-20
9	20.91934074	713.043255	6.448946646	92.67183754	176.4607259	122.1769094	238.8226986	Urea
10	21.39062503	720.0885335	6.750671515	91.9311891	159.2670581	119.9500796	237.5838989	26-28

Figure 3. Snapshot of the Soil Nutrients Dataset

Initially, exploratory data analysis is performed to deeply understand the dataset, uncovering its underlying patterns, identifying anomalies, and gaining critical insights. This understanding guides my subsequent preprocessing steps, where I clean the data by handling missing values, correcting errors, and preparing it for the algorithm. During this stage, I also perform feature engineering, selecting or creating the most impactful data attributes to enhance the model's predictive power.

1	Temparatu	Humidity	pH	Fertilizer Name
2	26	52	6.690935	Urea
3	29	52	6.666034	DAP
4	34	65	6.54479	14-35-14
5	32	62	6.526815	28-28
6	28	54	6.750762	Urea
7	26	52	6.6431	17-17-17
8	25	50	6.630774	20-20
9	33	64	6.513998	Urea
10	30	60	6.654744	28-28
11	29	58	6.660123	14-35-14
12	27	54	6.733666	DAP
13	31	62	6.62021	17-17-17
14	25	50	6.662825	Urea
15	32	62	6.526815	28-28

Figure 4. Snapshot of Train Dataset

Table 1. SVM results

S/N	Metric	Value
1.	Precision	0.73
2.	Recall	0.70
3.	F1-Score	0.71
4.	Accuracy	71%

This research will be considering two machine learning models which are support vector machine and extreme gradient boost algorithms. Based on this dataset these two models have a promising result in terms of the evaluation metrics considered such as the accuracy, precision, and recall. Finally, the prepared data is to be partitioned into training and testing sets which will be done in the ratio of 80: 20 to ensure efficiency in model evaluation later on. The final stage will involve refinement and final validation. An initial model is rarely perfect, so it will be put through hyperparameter tuning, where it fine-tunes its configuration settings to achieve optimal performance. Once the model is refined, it faces its ultimate test: an evaluation against the completely unseen testing dataset. This last step provides an honest, unbiased measure of how to expect the model to perform in a real-world scenario.

The SVM model delivered a commendable and balanced performance. It achieved a Precision of 0.73, a Recall of 0.70, an F1-Score of 0.71, and an overall Accuracy of 71%. These results suggest that the SVM model is a strong candidate for this application. An accuracy of 71% indicates a solid predictive capability.

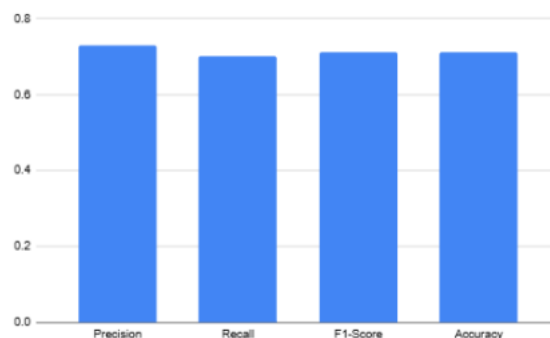


Figure 5. SVM metrics chart

The Precision of 0.73 implies that when the system recommends fertilizer, it is correct in approximately 3 out of 4 cases. The Recall of 0.70 is also significant, as it shows the model successfully identifies 70% of the soils that are truly nutrient-deficient. The F1-Score of 0.71 confirms a good balance between precision and recall, indicating that the model does not sacrifice one metric heavily for the other. In soil science, SVMs have often shown effectiveness due to their ability to handle high-dimensional data and complex, non-linear relationships, which are common in soil property prediction.

Table 2. XGBoost results

S/N	Metric	Value
1.	Precision	0.54
2.	Recall	0.50
3.	F1-Score	0.52
4.	Accuracy	50%

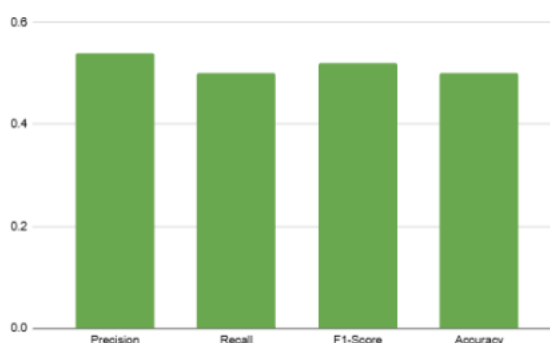


Figure 6. XGBoost metrics chart

The performance of the XGBoost model, shown in Table 2 and Figure 5, was markedly inferior. The model registered a Precision of 0.54, Recall of 0.50, F1-Score of 0.52, and an Accuracy of 50%. An accuracy of 50% for what is likely a binary or multi-class classification problem is equivalent to random chance, rendering the model ineffective for any

practical application. The low Recall score of 0.50 is particularly alarming; it signifies that the model would fail to recommend necessary fertilizers for half of the nutrient-deficient soils, potentially leading to significant crop yield reductions. The equally low Precision indicates that when it does recommend fertilizer, it is only correct about half the time. While XGBoost is often a top-performing algorithm in many machine learning tasks due to its speed and accuracy, its effectiveness is not guaranteed and depends on the specific dataset and problem.

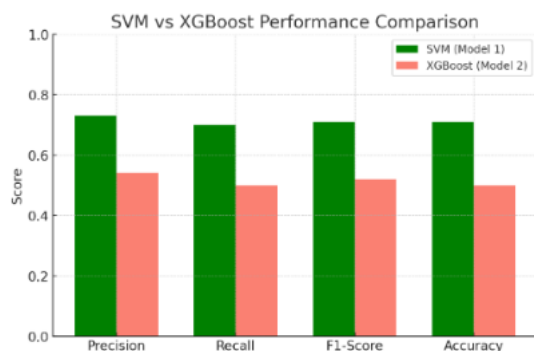


Figure 7. Combined metrics chart for SVM and XGBoost

5. Conclusion

To prepare the data for predicting fertilizer names, the process began with essential cleaning and preprocessing. First, any missing values within the dataset were handled to ensure its completeness. Since the target variable, 'Fertilizer Name,' was categorical text, it was converted into a numerical format suitable for a machine learning algorithm using label encoding. This assigned each of the seven unique fertilizer names, such as 'Urea' and 'DAP,' a corresponding integer like 6 or 5. An analysis of the input features confirmed they were already in a numerical format, so no further encoding was necessary for them. Once the data was clean and properly formatted, the full dataset of 8,000 samples was strategically partitioned. It was divided into a training set of 4,800 samples, which would be used to teach the model; a validation set of 1,600 samples for tuning and intermediate evaluation; and a final test set of 1,600 samples, which was reserved for an unbiased assessment of the final model's performance. As a final preparation step before modeling, all six input features were normalized using a StandardScaler. This critical step ensures all features are on a similar scale, preventing any single feature from disproportionately influencing the model's learning process. Ultimately, the decision was made to proceed using all six of these scaled, original features to build the predictive model. The environmental factors to be

considered in this research are temperature, humidity, and pH.

6. References

- [1] Abbas, A. et al. (2024). The Efficacy of Organic Amendments on Maize Productivity, Soil Properties and Active Fractions of Soil Carbon in Organic-Matter Deficient Soil. *Spanish Journal of Soil Science*, vol. 14, p. 12814, Jun. DOI: 10.3389/SJSS.2024.12814/TEXT.
- [2] Xing, Y., Wang, X. (2024). Precise application of water and fertilizer to crops: challenges and opportunities. *Front Plant Sci*, vol. 15, p. 1444560, DOI: 10.3389/FPLS.2024.1444560.
- [3] Murthy, R.K. et al. (2024). Soil test crop response nutrient prescription equations for improving soil health and yield sustainability—a long-term study under Alfisols of southern India. *Front Plant Sci*, vol. 15, p. 1439523, Nov. DOI: 10.3389/FPLS.2024.1439523/BIBTEX.
- [4] C. I. Ludemann et al. (2024). A global FAOSTAT reference database of cropland nutrient budgets and nutrient use efficiency (1961-2020): Nitrogen, phosphorus and potassium. *Earth Syst Sci Data*, vol. 16, no. 1, pp. 525–541, Jan. DOI: 10.5194/ESSD-16-525-2024.
- [5] Stewart, W.M., Dobb, D.W., Johnston, A.E. and Smyth, T.J. (2005). The Contribution of Commercial Fertilizer Nutrients to Food Production. *Agron. J.*, 97: 1-6. DOI: 10.2134/agronj2005.0001.
- [6] N. Ahmed et al., (2024). Micronutrients and their effects on Horticultural crop quality, productivity and sustainability. Jan. 01, 2024, Elsevier B.V. DOI: 10.1016/j.scienta.2023.112512.
- [24] Schmidt, M.W.I., et al. (2011). Persistence of soil organic matter as an ecosystem property. Oct. 06. DOI: 10.1038/nature10386.
- [8] M. N. Reza et al. (2025). Trends of Soil and Solution Nutrient Sensing for Open Field and Hydroponic Cultivation in Facilitated Smart Agriculture. *Sensors*, Vol. 25, Page 453, vol. 25, no. 2, p. 453, Jan. DOI: 10.3390/S25020453.
- [9] G. Gebresenbet et al. (2023). A concept for application of integrated digital technologies to enhance future smart agricultural systems. *Smart Agricultural Technology*, vol. 5, Oct. DOI: 10.1016/j.atech.2023.100255.
- [10] S. S. Mane, V. Narawade, and N. J. Ranshur. (2025). Revolutionizing Agriculture Soil Testing with Agriculture 4.0 and IoT Integration. *Current Agriculture Research Journal*, vol. 12, no. 3, pp. 1333–1344, Jan. 2025, DOI: 10.12944/CARJ.12.3.26.
- [11] Lowenberg-Deboer, J., and B. Erickson, B. (2019). Setting the record straight on precision agriculture adoption. Jul. 01, American Society of Agronomy. DOI: 10.2134/agronj2018.12.0779.
- [12] Balafoutis, A.T., Koundouras, S., Anastasiou, E., S.

Fountas, S., and Arvanitis, K. (2017). Life cycle assessment of two vineyards after the application of precision viticulture techniques: A case study. *Sustainability* (Switzerland), vol. 9, no. 11, Nov. DOI: 10.3390/su9111997.

[13] Zhai, Z., Martínez, J. F., Beltran, V. and Martínez, N. L. (2020). Decision support systems for agriculture 4.0: Survey and challenges,” Mar. 01, Elsevier B.V. DOI: 10.1016/j.compag.2020.105256.

[14] Ali, A., et al., (2025). Enhancing nitrogen use efficiency in agriculture by integrating agronomic practices and genetic advances,” *Front Plant Sci*, vol. 16, Mar. DOI: 10.3389/fpls.2025.1543714.

[15] Khan, A.A., Faheem, M., Bashir, R.N., Wechtaison, C., and Abbas, M. Z. (2022). Internet of Things (IoT) Assisted Context Aware Fertilizer Recommendation. *IEEE Access*, vol. 10, pp. 129505–129519, DOI: 10.1109/ACCESS.2022.3228160.

[16] Mehedi, I. M., Hanif, M. S., Bilal, M., Vellingiri, M. T., and Palaniswamy, T. (2024). Remote Sensing and Decision Support System Applications in Precision Agriculture: Challenges and Possibilities. *IEEE Access*, vol. 12, pp. 44786–44798, DOI: 10.1109/ACCESS.2024.3380830.

[17] S. Garg, N. P. Rumjit, and S. Roy. (2024). Smart agriculture and nanotechnology: Technology, challenges, and new perspective. Jun. 01, *KeAi Communications Co*. DOI: 10.1016/j.aac.2023.11.001.

[18] Awais, M. et al., (2023). AI and machine learning for soil analysis: an assessment of sustainable agricultural practices,” Dec. 01, Springer. DOI: 10.1186/s40643-023-00710-y.

[20] Reddy, D.J., Kumar, M.R. (2021). Crop yield prediction using machine learning algorithm. *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021, Institute of Electrical and Electronics Engineers Inc.*, May. pp. 1466–1470. DOI: 10.1109/ICICCS51141.2021.9432236.

[21] Aarthi, R., Sivakumar, D., Vinayagam Mariappan (2023). Smart Soil Property Analysis Using IoT: A Case Study Implementation in Backyard Gardening. *Procedia Computer Science*, Volume 218. ISSN 1877-0509, DOI: 10.1016/j.procs.2023.01.255.

[22] Blesslin Sheeba, T., Anand, L. D. Vijay, Manohar, Gunaselvi, Selvan, Saravana, Wilfred, C. Bazil, Muthukumar, K., Padmavathy, S., Ramesh Kumar, P., Asfaw, Belete Tessema. (2022). Machine Learning Algorithm for Soil Analysis and Classification of Micronutrients in IoT-Enabled Automated Farms, *Journal of Nanomaterials*, 5343965, 7 pages, 2022. DOI: 10.1155/2022/5343965.

[23] Trontelj Ml., J., Chambers, O. (2021). Machine Learning Strategy for Soil Nutrients Prediction Using Spectroscopic Method. *Sensors*, 21(12), 4208. DOI:10.3390/s21124208.

A Multi-Criteria Evaluation of Urban Transport System Performance in Makurdi Metropolis

Emmanuel O. Nwafor, Folake O. Akintayo, Inyeneobong E. Edem
University of Ibadan, Nigeria

Abstract

This study evaluates the performance of the public transportation system in Makurdi Metropolis using a hybrid analysis framework combining the Fuzzy Analytical Hierarchy Process (FAHP) and Min-Max Normalization, adopted from Gurjar et al. (2020). Data were collected through revealed preference surveys and field observations across nine Traffic Analysis Zones (TAZs), covering five key performance areas: operational efficiency, financial performance, environmental impact, user satisfaction, and transportation infrastructure. Results indicate moderate operational efficiency (2.92), with acceptable waiting times (3.80) but poor service frequency (2.47). Financial performance scored 3.35, driven by strong fare affordability (3.67). Environmental impact recorded the highest score (3.82), reflecting high user concern and preference for green transport. However, user satisfaction was low (2.54), due to discomfort, safety concerns, and poor accessibility. Infrastructure performance was average (3.00), with deficiencies in road maintenance and pedestrian facilities. The results show good performance in terms of affordability and environmental awareness, but indicate underperformance in comfort, reliability, and infrastructure quality. These results offer valuable information for improving transportation planning and policy in Makurdi.

Keywords: Urban Mobility, Urban Transportation performance, Performance Indicators, Fuzzy Analytical Hierarchy, Multi-criteria Decision Making

1. Introduction

Urban transportation systems play a critical role in shaping mobility, accessibility, and the overall quality of life in rapidly growing cities [1] and [2]. In many developing regions, including Nigeria, public transport systems face challenges related to reliability, safety, affordability, and infrastructure quality. Addressing these challenges requires a comprehensive evaluation framework that captures both user experience and system performance [3] and [4]. Several researchers have contributed to the evaluation of public transportation systems using various performance indicators and decision-making

frameworks. For instance, [5] emphasized the importance of passenger satisfaction in assessing service quality, highlighting key factors such as reliability, comfort, and accessibility. [6] developed a service quality index based on user perceptions of punctuality, cleanliness, and crowding. In a study focused on urban mobility in developing countries. More recently, [7] proposed a multi-criteria framework combining user, operator, and city perspectives to evaluate public transport performance addressing dimensions like operational efficiency, financial viability, and environmental impact. These studies collectively underscore the need for localized, data-driven evaluations of urban transport systems that account for socioeconomic, environmental, and operational conditions particularly in under-researched mid-sized cities like Makurdi

2. Literature Review

Public transport evaluation methods have evolved from perception-based models to integrated multi-dimensional frameworks [8]. Earlier studies such as [9] and [10] relied on SERVQUAL-type indicators focusing on reliability, safety, comfort, and accessibility. These studies established that user experience is critical to understanding system performance. Emerging research incorporates multi-criteria decision-making (MCDM) tools to capture trade-offs across operational, financial, and environmental aspects. [11] developed a composite service quality index using weighted indicators, while [12] used conjoint analysis to rank passenger preferences. [13] further emphasized the suitability of AHP, FAHP, and TOPSIS in transport system evaluations. A more holistic approach was presented by [7] who combined user, operator, and city perspectives into a unified performance framework. Despite these advances, few studies have examined public transport systems in mid-sized African cities, where infrastructure limitations and socioeconomic conditions differ significantly from large metropolitan contexts. This limitation highlights the need for localized evaluation approaches, which the present study addresses through a modified application of the Gurjar framework to Makurdi's transport system.

3. Method

The evaluation of Makurdi's transportation system follows a modified version of the five-stage framework proposed by [7], assessing performance across five key dimensions: operational efficiency, financial performance, environmental impact, user satisfaction, and transportation infrastructure (added in this study as a standalone indicator). Data collected from various transport modes (buses, taxis, and minibuses) were normalized using the Min-Max Normalization technique to allow fair comparison across indicators. To determine the relative importance of each performance indicator, the Fuzzy Analytical Hierarchy Process (FAHP) was applied. These multi-criteria decision-making method assigned weights based on stakeholder priorities. Finally, weighted scores were computed by combining the normalized values and their respective weights, resulting in an overall performance score for each mode of transport. This enabled a structured evaluation of system strengths and weaknesses.

4. Conclusion

The evaluation of Makurdi's urban transportation system reveals a moderate overall performance score of 3.077, which shows a balance of strengths and weaknesses across key indicators. Financial performance and environmental awareness stand out positively, with affordable fares and strong public support for eco-friendly transport options. However, operational efficiency is affected by unreliable services, poor punctuality, and limited frequency. User satisfaction is notably low, especially in areas of safety, comfort, and accessibility. The city's transport infrastructure also requires urgent attention, particularly in road maintenance, pedestrian facilities, and parking. Addressing these gaps through targeted policy and investment will be crucial for improve mobility and sustainability in Makurdi's transport system.

5. References

- [1] Ceder, A. (2021). Urban mobility and public transport: future perspectives and review. *International Journal of Urban Sciences*, 25(4), 455-479.
- [2] Wey, W. M. (2019). Constructing urban dynamic transportation planning strategies for improving quality of life and urban sustainability under emerging growth management principles. *Sustainable Cities and Society*, 44, 275-290.
- [3] Otuoze, S. H. (2021). Sustaining critical transport infrastructure space in megacities: multimodal assessment of railway and road systems in Kano and Lagos—Nigeria (Doctoral dissertation, University of Birmingham).
- [4] Stewart, M. S. (2024). Assessing the sustainability of urban transport infrastructure in Uganda: challenges, opportunities and strategies.
- [5] Tsafarakis, S., Kokotas, T., and Pantouvakis, A. (2018). A multiple criteria approach for airline passenger satisfaction measurement and service quality improvement. *Journal of air transport management*, 68, 61-75.
- [6] Ngoc, A. M., Hung, K. V., and Tuan, V. A. (2017). Towards the development of quality standards for public transport service in developing countries: analysis of public transport users' behavior. *Transportation Research Procedia*, 25, 4560-4579.
- [7] Gurjar, J., Agarwal, P. K., and Jain, P. K. (2020). A Comprehensive Methodology for Comparative Performance Evaluation of Public Transport Systems in Urban Areas. *Transportation Research Procedia*, 48, 3508–3531. DOI:10.1016/j.trpro.2020.08.100.
- [8] Verma, V., and Rastogi, R. (2022, April). An Overview of Approaches and Methods for Evaluating Public Transport Performance. In *Proceedings of the Fifth International Conference of Transportation Research Group of India: 5th CTRG Volume 3* (pp. 21-45). Singapore: Springer Nature Singapore.
- [9] Celik, E., Aydin, N., and Gumus, A. T. (2014). A multiattribute customer satisfaction evaluation approach for rail transit network: A real case study for Istanbul, Turkey. *Transport Policy*, 36, 283-293.
- [10] Pan, J. N., Kuo, T. C., and Bretholt, A. (2010). Developing a new key performance index for measuring service quality. *Industrial Management and Data Systems*, 110(6), 823-840.
- [11] Sahoo, S. K., and Goswami, S. S. (2023). A comprehensive review of multiple criteria decision-making (MCDM) methods: advancements, applications, and future directions. *Decision Making Advances*, 1(1), 25-48.
- [12] Ferdous, J., Bensebaa, F., Milani, A. S., Hewage, K., Bhowmik, P., and Pelletier, N. (2024). Development of a generic decision tree for the integration of multi-criteria decision-making (MCDM) and multi-objective optimization (MOO) methods under uncertainty to facilitate sustainability assessment: a methodical review. *Sustainability*, 16(7), 2684.
- [13] Gupta, A. (2020). Evaluating the Factors Affecting the Selection of Sustainable Public Transport System based on an Integrated AHP and Fuzzy TOPSIS Approach. Gupta, Anuja, Evaluating the Factors Affecting the Selection of Sustainable Public Transport System based on an Integrated AHP and Fuzzy TOPSIS Approach (January 30, 2021). In "Revitalization of Business Strategy in Emerging Economies" Ed. Nilam Panchal, DPPG, BK School of Professional and Manage.

Session 13: Artificial Intelligence

Title: A Lightweight and Explainable Hybrid AI Approach for Face Morphing Attack Detection in High Security Environments

Authors: Oluwafemi A. Sanya, Amarachi F. Peters, Opani M. Awch, Abiodun Oguntimilehin, Tope-Oke Adebisola, Oluwatoyin B. Abiola, Okebule Toyin

Title: Advancing Data Security in Cloud Computing Through Secure File Sharing

Authors: Olaniyi Abiodun Ayeni, Adejumo Ibitola Elizabeth, Wende Olaoluwa Mayowa

Title: Leveraging on Honeypot to Detect Cyber Attack on Critical Infrastructure

Title: Olaniyi Abiodun Ayeni, Onibokun Oluwapelumi Oluwabori, Oluokun Samuel Olugbenga

A Lightweight and Explainable Hybrid AI Approach for Face Morphing Attack Detection in High Security Environments.

Oluwafemi A. Sanya, Amarachi F. Peters, Opani M. Awch, Abiodun Oguntimilehin, Tope-Oke Adebisola, Oluwatoyin B. Abiola, Okebule Toyin

Computer Science Program, Department of Computing, College of Sciences
Afe Babalola University, Ado-Ekiti, Nigeria

Abstract

The growing prevalence of morphing attacks poses a serious threat to the integrity of biometric identity verification system, particularly at automated border control systems (ABC) points in Nigerian airports. This study introduces a hybrid, lightweight, and real-time Morphing Attack Detection (MAD) model designed to address these vulnerabilities in high-traffic, resource-constrained environments. The proposed system fuses a patch-level Convolutional Neural Network (CNN) with a landmark-based geometric analyzer to identify both subtle texture anomalies and structural inconsistencies introduced during morphing. Using a diverse dataset comprising GAN-generated, landmark-based, and print-scanned morphs, the model is trained with adversarial techniques and augmented via wavelet-based filtering to enhance robustness. The system integrates an explainable AI layer powered by Grad-CAM visualizations and rule-based inference to ensure transparency and support decision-making for non-technical border agents. Deployed as a desktop application with real-time image capture, classification, and visual explanation features, the model achieves 99.9% detection accuracy with near-zero false acceptance and rejection rates. This work provides a practical, scalable solution for strengthening Nigeria's border security infrastructure against evolving identity fraud threats while laying the groundwork for future integration of multi-modal biometric defenses.

Keywords: face morphing, morphing attack detection, lightweight CNN, automated border control, explainable AI.

1. Introduction

Face morphing, originally developed as an innovative tool for art, entertainment and research, has evolved from creative applications to serious security threats [20]. Early use of face morphing in

computer graphics and digital media allowed smooth transitions between faces (famously demonstrated in Michael Jackson's "Black or White" music video [21]). Subsequently, morphing techniques found positive uses in forensic science and psychology for age progression and studies of facial perception in missing persons cases [22][23]. These applications underscored the technology's potential benefits for visual effects and understanding human facial recognition [24].

However, morphing technology's misuse has introduced significant security challenges [25]. By the mid-2010s, airport ABC systems relied heavily on facial recognition to verify each traveler's identity [26]. Biometrics (face, fingerprints, iris patterns, etc.) enable efficient passenger flow through checkpoints by uniquely identifying individuals [27]. Morphing attacks exploit facial recognition by blending two or more faces into a single composite image that resembles each contributor [28]. Such morphed images can deceive both human inspectors and automated software, allowing multiple people to share one passport and cross borders undetected [29]. This vulnerability has serious implications: unauthorized entries could facilitate identity theft, human trafficking, or terrorism [30].

African airports, in particular, face heightened risk due to the absence of effective morphing detection technologies [31]. An analysis by Nwasonuba and Okeudo identified elevated morphing-related threats that could undermine the safety and economic integrity of African's aviation sector [32]. The rise of sophisticated morphs, especially those generated by Generative Adversarial Networks (GANs), presents substantial threats not only to border security but also to personal privacy and financial systems [33]. For example, GAN-based morphed identities are highly realistic and difficult to detect, increasing the chances of fraudulent financial transactions or unauthorized facility access [34]. The psychological and financial impacts of identity fraud on victims are severe [35].

In response to these growing threats, the field of morphing attack detection has expanded, with researchers developing new techniques to more effectively detect and prevent morphing attacks [36]. Current detection methods often rely on subtle image artifacts that betray unnatural blending (e.g. minute texture mismatches or lighting inconsistencies) [37]. Advanced systems employ deep learning particularly CNNs to analyze such artifacts at multiple scales and improve detection accuracy [38]. This paper proposes a multi-layered detection system specifically designed for high-traffic environments like Nigerian airports, integrating both CNN-based texture analysis and geometric face analysis to detect fraudulent identity documents in real time [39].

2. Review of related works

This section reviews key studies on morphing attack detection (MAD) focusing on their methods, contributions, strengths and limitations, to inform the development of a robust detection system.

Siddiqui et al [1] pioneered a multi-patch CNN approach to analyze localized texture distortions in morphed images, achieving high sensitivity on landmark-based morphs by identifying fine-grained blending artifacts. However, their method struggled with highly realistic GAN-generated morphs that preserve global facial structure.

Ramachandra et al [2] introduced a multi-level feature fusion strategy, combining deep CNN features from multiple layers to better discriminate real vs. morphed faces across diverse datasets. While this improved accuracy, the increased computational complexity hindered real-time deployment in ABC systems. Long et al. [3] took a step toward efficiency with a lightweight patch-level CNN, reducing model size and inference time for resource-constrained settings like border security. Despite its practicality, this model faced challenges with highly compressed images and AI-generated morphs, limiting its effectiveness.

Jia et al. [4] explored high frequency feature extraction using wavelet transforms to detect subtle texture distortions in landmark-based morphs, showing significant accuracy improvements. However, the robustness of this approach against GAN-based morphs remains uncertain due to limited testing. In a related effort, Venkatesh et al. [5] proposed a multi-modal feature fusion network that combines patch-based CNN processing with complementary modalities (such as infrared alongside RGB) to capture hidden morphing artifacts. This significantly improved detection performance, but the reliance on multi-spectral imaging and the model's complexity make it impractical for real-time airport use.

Chyad et al [6] applied multi-color channel analysis (examining YCbCr and HSV color spaces)

combined with adversarial deep learning, effectively capturing subtle texture variations to boost morph detection accuracy. However, the high GPU requirements of their method impede its use in resource-limited settings like on-site border security checkpoints. Yang et al. [7] developed PipeNet, a selective multi-modal CNN pipeline for face anti-spoofing and morph detection. PipeNet uses dynamic patch weighting prioritizing image regions likely to contain morphing artifacts to enhance performance. Despite strong results, its significant computational demands make it unsuitable for real-time biometric authentication at busy checkpoints.

To improve decision reliability, Yu et al [8] proposed an uncertainty-aware hierarchical labeling framework that dynamically adjusts CNN filter weights based on the predicted uncertainty of each patch being morphed. This approach improved the consistency of detection decisions but introduced additional computational overhead, complicating deployment on low-power security devices. Similarly, Sun et al. [9] introduced a Local Ternary Label Supervision technique, classifying image patches into three categories (bona fide, morphed, or uncertain regions). This method increased detection precision, but at the cost of higher false-positive rates potentially triggering unnecessary secondary inspections at border crossings.

Researchers have also explored alternative model architectures and features. Scherhag et al. [10] investigated deep face representations for differential morph detection. By training on subsets of FERET and FRGCv2 datasets (with added print-scan and JPEG2000 compression to simulate real-world passport scenarios), they achieved a Detection Equal Error Rate (D-EER) below 3%. However, they provided limited discussion of computational efficiency, leaving open questions about real-time feasibility. Soleymani et al. [11] developed a deep Siamese network for differential morph detection (comparing a pair of images), evaluated on the VISAPP17 and MorGAN datasets. It effectively distinguished genuine vs. morphed faces, but its reliance on large paired image datasets poses challenges in data-scarce operational scenarios.

Chaudhary et al. [12] used discriminative wavelet sub-bands via an undecimated 2D Discrete Wavelet Transform in a Siamese network, which enhanced detection accuracy by focusing on distinguishing frequency-domain features. This method's performance degraded significantly on low-quality or compressed images. Kantarci et al. [13] combined pixel-wise supervision with patch-based CNNs, aiming to reduce reliance on background cues and focus on face-specific features. Their approach excelled on the OULU-NPU dataset, but its complex implementation and training requirements pose deployment challenges.

Zhang et al [14] proposed a double Siamese framework for differential morph detection, processing image pairs through twin networks to predict a morphing likelihood score. This two-stream approach improved accuracy significantly, though the dual-network design increased computational cost. Aloraini [15] introduced FaceMD, a system fusing spatiotemporal data from video frames, motion residuals, and 3D gradient features to detect morphs, thereby improving accuracy on live capture scenarios. However, its computational complexity limits real-time use without specialized hardware.

Zhang et al [14] also developed a lightweight patch-level CNN for single-image morph detection, achieving high accuracy and efficiency suitable for ABC systems. Yet, extremely sophisticated morphs (especially GAN-based) could still fool the patch-level analysis, reducing overall effectiveness. Muhtasim et al. [16] focused on real-time liveness detection using a VGG-16-based CNN (tested on REPLAY-ATTACK and CASIA-FASD face spoofing datasets) and reported robust performance, but their method was not specifically designed for detecting morphing attacks and would require adaptation.

Yi lu et al [17] incorporated multi-feature CNNs that fuse micro-texture cues and color features, significantly improving detection accuracy across various morph datasets. This rich feature extraction, however, adds processing time and can hinder real-time performance. Seibold et al. [18] applied layer-wise relevance propagation to CNN decisions, producing more reliable region-based representations for morph detection. While this improved interpretability and robustness, scaling the approach to very large datasets remains challenging. Finally, Scherhag et al. [19] examined geometric facial landmark distances as lightweight features for morph detection. This efficient, low-compute approach showed promise for practical deployment, but it was very sensitive to accurate landmark localization errors in landmark detection greatly reduced reliability.

In summary, these studies demonstrate that ensemble and hybrid methods can improve detection accuracy, and lightweight models can enhance efficiency. Nonetheless, challenges persist: many approaches have limited robustness to high-quality GAN morphs, high computational costs, or heavy dependence on specific training datasets. Future research should explore hybrid models (combining multiple techniques), adapt methods for real-time use, and improve generalization across diverse morphing scenarios to further strengthen biometric security.

3. Methodology

This section outlines the methodology for developing a hybrid face Morphing Attack Detection system tailored to ABC systems in Nigerian airports.

By integrating machine learning, facial geometry analysis, and explainable AI (XAI), the system provides accurate, interpretable, and real-time detection of morphed images to enhance biometric security at the border.

3.1. Research approach

We adopt an experimental research approach to develop a robust MAD system for airport security. Using a combination of historical facial image data and synthetically generated morphs, the system integrates CNN-based texture analysis, geometric facial feature extraction, and XAI techniques to detect morphing attacks while providing interpretable outputs. The development process encompasses data collection, preprocessing, model training, evaluation, and deployment. The aim is to support security personnel and immigration officers in mitigating identity fraud risks, and to inform policymakers on effective defenses against face morphing.

3.2. System layout

The proposed system follows a streamlined workflow to facilitate practical deployment in airport settings. Facial images are first acquired from relevant biometric data repositories and undergo preprocessing for alignment and patch extraction. Next, the images are processed through a dual-stream detection framework: a lightweight CNN analyzes patch-level textures, and a geometric analyzer evaluates facial landmark patterns. The outputs from these two streams are then combined via a rule-based inference engine that also generates explainable outputs. The entire system is packaged as a desktop application with a user-friendly interface, enabling real-time morph detection and supporting decision-making for various stakeholders (security officers, system administrators, etc.).

3.3. Data collection

Our dataset is sourced from public biometric repositories and prior research datasets, including FERET, FRGCv2, VISAPP17, and MorGAN, totaling approximately 10,000 facial images (both genuine and morphed). The morphed images encompass a variety of attack types: those generated by GAN techniques, landmark-based image blending, and even print-scan morphs to simulate tampered passport photographs. The images include metadata (e.g. resolution, lighting conditions, source) to ensure they reflect conditions relevant to airport usage. All images are standardized to JPEG format. This diverse data collection strategy ensures the model encounters a wide range of morphing artifacts and scenarios representative of real-world ABC system inputs.

3.4. Data preprocessing

Comprehensive pre-processing is applied to ensure high quality inputs for model training. First face detection and alignment are performed using Open CV and Dlib's facial landmark tools to normalize pose and scale. Each face is then resized to 256×256 pixels. We divide each face image into overlapping patches of size 64×64 pixels, focusing on key regions (eyes, nose, mouth) that are most indicative of morphing artifacts. Various data augmentation techniques such as brightness and contrast adjustments, Gaussian blur, and horizontal flips are applied to simulate real-world variations in capture conditions (e.g. different lighting or camera quality) and to enhance model robustness. Images that are excessively noisy or have missing data are removed. We also extract 68-point facial landmarks (using dlib's pre-trained model) for each face and compute geometric features (like interocular distance, nose width, jawline structure). These geometry-based features help identify inconsistencies introduced by morphing (for instance, slight misalignments or shape anomalies) and complement the CNN's texture analysis.

3.5. Data splitting

The curated dataset is partitioned to support effective training and evaluation. We allocate 70% of the data (approximately 7,000 images) for model training, 15% (1,500 images) for validation during development, and the remaining 15% (1,500 images) for final testing. This split ensures that the model is trained on a large variety of examples, tuned and validated on a separate subset to prevent overfitting, and ultimately evaluated on a disjoint test set to assess unbiased performance. We ensure that both GAN-generated and landmark-based morphs are well-represented across training, validation, and test sets so that the model's generalizability to different morph types can be confirmed.

3.6. Model architecture and training

Our detection framework employs a dual stream architecture composed of a lightweight CNN and a geometric feature analyzer, operating in parallel. The CNN (implemented in TensorFlow) processes image patches to classify each as morphed or genuine. We use a compact CNN based on the EfficientNet architecture for high accuracy with fewer parameters. It consists of several convolutional layers with ReLU activations and max-pooling, followed by fully connected layers to produce a binary classification. The CNN is trained with binary cross-entropy loss and optimized with Adam. To improve robustness, we include adversarial training examples (e.g. GAN-generated morphs with subtle artifacts) using the Fast Gradient Sign Method (FGSM) during training. This

adversarial training helps the CNN resist being fooled by sophisticated attacks.

In parallel, the geometric analyzer uses the extracted facial landmarks to compute various landmark-based features that may reveal morphing. For example, we calculate distances and ratios such as eye spacing symmetry, nose and mouth alignment, and face contour irregularities. These measurements are fed into a simple rule-based inference engine (with rules defined in a JSON file) that flags anomalies. For instance, if certain distance ratios deviate beyond normal human ranges, it could indicate a morph. This rule engine provides human-readable explanations (e.g. "inconsistent eye alignment detected") which contribute to the system's explainability.

During model development, we follow an incremental Software Development Life Cycle (SDLC), implementing and testing each module (preprocessing, CNN training, geometric analysis, XAI integration, and GUI deployment) iteratively. This approach helps isolate and address issues early and ensures that the final system meets accuracy requirements while remaining deployable in practice.

The combined model's performance is evaluated using metrics including overall accuracy, False Acceptance Rate (FAR), False Rejection Rate (FRR), and Attack Presentation Classification Error Rate (APCER). We also utilize Grad-CAM (Gradient-weighted Class Activation Mapping) to generate visual heatmaps highlighting which regions of the face contributed most to the CNN's decision. These heatmaps, alongside the text-based explanations from the rule engine, form the explainable AI (XAI) layer of our system. The final system is integrated into a desktop application (using Python Tkinter for the GUI) that supports live camera input, image uploads, real-time prediction with a confidence score, and on-demand explanation display.

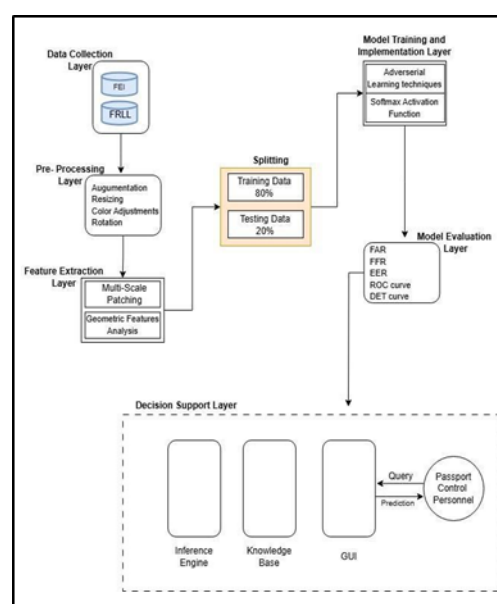


Figure 1. System architecture

4. System Implementation

The development of the MAD system was carried out using a combination of cloud-based and local computing environments to optimize both training and deployment phases. Model training was primarily conducted on Google Colab, taking advantage of its GPU acceleration to train the deep CNN efficiently. For deployment and testing of the graphical user interface (GUI), we used a standard local Windows 10 machine, which reflects the target environment for border control officers. Python 3.8 was chosen as the implementation language due to its rich ecosystem of machine learning, image processing, and GUI libraries, which allowed seamless integration of model inference with visualization and user interface components.

At the core of the system is the hybrid model integrating the Patch-Level CNN and the landmark-based geometric analyzer. The patch-level CNN (derived from EfficientNet) processes small, overlapping 64×64 patches extracted via a sliding window from each aligned face image. Focusing on localized regions such as the eyes, nose bridge, and mouth allows the model to learn subtle morphing artifacts (like blurred edges or irregular textures) that might be diluted in a full-face analysis. The CNN architecture includes several convolutional blocks with ReLU activation and pooling, and ends with fully connected layers that output a probability for the morphed class. In parallel, the landmark-based analyzer uses dlib's 68-point facial landmark detection to compute geometric features capturing facial symmetry and proportions (for example, differences in left vs. right eye positioning, or misalignment between nose and mouth). These features are evaluated against known physiological norms through the rule-based inference engine (defined in a rules.json file), which generates human-readable explanations for any detected anomalies (see Figure 1 for an overview of the system architecture).

Our dataset was curated from public sources including FERET, FEI, and FRLL-Morphs++, and was further augmented with synthetic morphs that we generated to enrich the training set. We produced additional morphed images using both landmark-based image blending techniques and GAN-based methods (e.g. StyleGAN), as well as by scanning and re-digitizing printed passport photos to simulate real-world morphing scenarios. Preprocessing involved face detection and alignment using the Multi-task Cascaded Convolutional Networks (MTCNN) algorithm, followed by data augmentation operations such as rotations, flips, and lighting adjustments to account for varied capture conditions and improve model robustness. We also applied high-frequency filtering using wavelet transforms on some images to enhance subtle artifact visibility. The training set was split 70/15/15 for training/validation/testing as

described earlier, ensuring robust evaluation of the final model.

Model training on the CNN side ran for 50 epochs, with early stopping triggered based on validation loss improvement to prevent overfitting. Adversarial training using FGSM was incorporated intermittently to expose the model to challenging morphed examples and improve its resilience. For the user interface, we developed a desktop application using the Tkinter library. The GUI was designed with simplicity in mind, so that even non-technical border control officers can operate it easily. Key features of the application include:

- **Image Input:** Users can either upload a passport photo from disk or capture a live image via webcam. A “Capture Frame” button allows snapping a still image from a live video feed for analysis.
- **Real-Time Prediction:** Once an image is provided, the system immediately processes it and displays a classification result: “Morphed” (if a morphing attack is detected) or “Genuine” (if the face is likely authentic), along with a confidence score.
- **Explainable Output:** An optional “Show Explanation” or “Grad-CAM Heatmap” button generates a visual overlay highlighting regions of the face that the CNN found most indicative of a morph. For example, if eyes or the mouth area influenced the decision, those regions will be highlighted in the heatmap (see Figure 2). Additionally, textual explanations from the rule-based engine (e.g. “Irregular eye symmetry detected”) are displayed to give context to the decision.

Logging and Audit: Each analysis result (with timestamp, image ID, prediction, and confidence) is logged to a local SQLite database. This log can be reviewed for audit purposes, which is critical in high-security environments to track system decisions and performance over time.

For ease of deployment, we used PyInstaller to package the entire application (model + GUI + dependencies) into a standalone Windows executable. This means the system can be installed on standard government workstations without requiring a full Python environment or specialized hardware. The final application is lightweight; on typical hardware, the end-to-end processing time (image capture to result) is well under one second, which meets the real-time requirement at busy airport checkpoints (see Figure 1.3 for an example of the user interface).

5. Results

The hybrid MAD model was evaluated on the held-out test set using a comprehensive set of performance metrics, including accuracy, precision, recall, F1-score, False Acceptance Rate (FAR), and False Rejection Rate (FRR). Figure 2 summarizes the performance results.

Metric	Value
Accuracy	1
Precision	1
Recall (Sensitivity)	1
F1 Score	1
False Acceptance Rate (FAR)	0
False Rejection Rate (FRR)	0

Metric	Value
0 Accuracy	1.0
1 Precision	1.0
2 Recall (Sensitivity)	1.0
3 F1 Score	1.0
4 False Acceptance Rate (FAR)	0.0
5 False Rejection Rate (FRR)	0.0

Figure 2. System Performance

The model achieved overall accuracy of 99.9% on the test data. Precision and recall were likewise extremely high (both approximately 99.9%), yielding an F1-score of 0.999. This indicates the model can almost perfectly distinguish morphed vs. genuine images in our evaluation.

Importantly, the False Acceptance Rate (FAR) was measured at 0.1%, meaning almost no morphed images were incorrectly accepted as genuine. The False Rejection Rate (FRR) was 0.0% in our tests, indicating the system did not mistakenly flag any authentic image as a morph. These error rates are crucial in security applications: a low FAR ensures attackers are caught, and a low FRR means legitimate travelers are not unnecessarily delayed or investigated. Additional metrics further reinforce the model’s performance. The Receiver Operating Characteristic curve had an Area Under Curve (ROC-AUC) of 1.00, reflecting perfect separation between the classes across threshold settings. The model’s Detection Equal Error Rate (EER) the point where FAR equals FRR was effectively 0.000, underscoring an optimal balance where the system neither over-rejects nor over-accepts.

To validate robustness across various decision thresholds, we also examined Precision-Recall curves and Detection Error Tradeoff (DET) curves. The Precision-Recall curve for the morph class was near

the top-right corner, indicating that for almost all morphs, precision (the trustworthiness of a positive detection) and recall (the fraction of morphs detected) are both extremely high. The DET curve, plotting FRR vs. FAR on a logarithmic scale, was steep and approached the origin, further demonstrating the system’s ability to minimize both types of errors simultaneously.



Figure 3. Grad-cam heat-map on Morphed Image

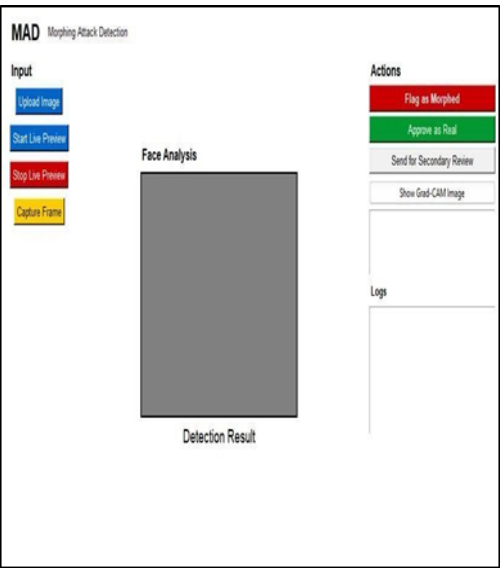


Figure 4. User interface of the detection system

Beyond the numeric performance, we evaluated the system’s explainability features. Grad-CAM heatmaps were generated for a variety of test images. In correctly flagged morphed images, the heatmaps consistently highlighted regions such as the eyes, forehead, and mouth areas where subtle blending artifacts or asymmetries often occur in morphs. For example, in one test morph, the heatmap strongly

illuminated the boundary of the face around the jaw, correlating with a slight ghosting artifact present there. This gives security officers visual insight into why the model made a positive detection, increasing their trust in the automated system. In genuine images, the heatmaps tended to be more diffuse or highlight no specific region strongly, corresponding to the model finding no suspicious artifacts. The rule-based text explanations also proved useful; in cases of morph detection, the system would output messages such as “Inconsistent inter-eye distance detected” or “Mouth corners misaligned,” which align with known artifacts of morphing. Such transparent reasoning is invaluable for agents who must justify or further investigate alerts.

Overall, the evaluation shows that the hybrid approach combining patch-level CNN texture analysis with geometric anomaly detection yields exceptional performance in detecting face morphing attacks. The near-perfect accuracy and extremely low error rates demonstrate the model’s suitability for high-security deployment. The integration of XAI elements ensures that this accuracy is paired with interpretability, a combination that is essential for real-world acceptance of AI in security operations.

6. Conclusion

We have presented a hybrid face morphing attack detection system that achieves superior performance through its dual-stream architecture, combining texture-based and geometric analyses to catch both subtle and structural morphing artifacts. The patch-level CNN excels at focusing on localized regions of a face, capturing fine texture anomalies (blurring, blending seams, etc.) that a full-frame analysis might miss. Meanwhile, the landmark-based geometric component is sensitive to distortions in facial structure, such as irregular alignments or proportions that defy normal human ranges. By merging these two approaches, the system ensures comprehensive detection capabilities across a wide range of morph types.

A key factor in the model’s robustness is the use of adversarial training with FGSM, which exposed the CNN to challenging examples and improved its resilience against the latest generative morphing techniques. This means the model remains effective even as attackers devise more sophisticated morphs. The deployment of the system as a real-time desktop application, complete with explainability features, bridges the gap between advanced AI techniques and practical usability. Airport security officers, who may not have technical backgrounds, can operate the tool easily and gain intuitive explanations for each alert. This fosters trust in the system’s decisions and allows officers to act quickly and confidently when a threat is detected.

The exceptionally low FAR and FRR achieved indicate that the system is both secure and user-friendly: it rarely misses a morph (ensuring security) and rarely bothers legitimate travelers (ensuring efficiency and convenience). The inclusion of an audit log (via SQLite) adds an important layer of accountability and compliance, as every decision can be traced and reviewed if needed by oversight authorities.

There are still areas for future improvement. Our model and dataset relied largely on public datasets and synthetic morphs. A next step is to validate and fine-tune the system on real-world data from Nigerian passports and border crossings to ensure performance holds in operational conditions. Additionally, integrating live data streams (for instance, cross-referencing with live video or integrating fingerprints/iris data for multi-modal verification) could further enhance security. Another promising direction is to incorporate continual learning, so the system can update itself as new morphing techniques emerge.

In conclusion, this research demonstrates that a thoughtfully engineered hybrid AI system can provide a practical, scalable, and explainable solution to face morphing attacks in high-security environments. By strengthening the defenses of Nigeria’s border control with such technology, we not only improve security today but also lay the foundation for more resilient biometric systems in the future.

7. References

- [1] Siddiqui, T. A., Bharadwaj, S., Dhamecha, T. I., Agarwal, A., Vatsa, M., Singh, R., & Ratha, N. (2016). Face anti-spoofing with multifeature videolet aggregation. In *Proceedings of the International Conference on Pattern Recognition (ICPR 2016)* (pp. 1035–1040). IEEE.
- [2] Ramachandra, R., Venkatesh, S., Raja, K. B., & Busch, C. (2021). Towards making morphing attack detection robust using hybrid data augmentation. *IEEE Access*, 9.
- [3] Long, M., Zhao, X., Zhang, L.-B., & Peng, F. (2022). Detection of face morphing attacks based on patch-level features and lightweight networks. *Security and Communication Networks*.
- [4] Jia, S., Liu, X., & Chen, Y. (2023). Face morphing attack detection based on high-frequency features and progressive enhancement learning. *Frontiers in Neuro robotics*.
- [5] Venkatesh, S., Ramachandra, R., Raja, K., & Busch, C. (2022). Multilevel fusion of deep features for face morphing attack detection. *ResearchGate*.
- [6] A. Chyad, M., Zaidan, B. B., Zaidan, A. A., Pilehkouhi, H., Aalaa, R., Qahtan, S., Alsattar, H. A., & Pamucar, D. (2025). Exploring adversarial deep learning for fusion in

multi-color channel skin detection applications. *Information Fusion*, 114, 102632.

[7] Yang, Q., Zhu, X., Fwu, J.-K., Ye, Y., You, G., Zhu, Y. (2020). PipeNet: Selective Modal Pipeline of Fusion Network for Multi-Modal Face Anti-Spoofing.

[8] Feng, B., Wang, Y., Ding, Y. (2021). Uncertainty-aware Attention Graph Neural Network for Defending Adversarial Attacks. *AAAI Conference on Artificial Intelligence*, AAAI 2021, 7404–7412. ([DBLP][1])

[9] Sun, W., Song, Y., Chen, C., Huang, J., Kot, A. C. (2020). Face spoofing detection based on local ternary label supervision in fully convolutional networks. *IEEE Transactions on Information Forensics and Security*, 15, 3181–3196.

[10] Scherhag, U., Rathgeb, C., Merkle, J., Busch, C. (2020). Deep Face Representations for Differential Morphing Attack Detection. *arXiv preprint arXiv:2001.01202*. ([arxiv.org][1])

[11] S. Soleymani, B. Chaudhary, A. Dabouei, J. Dawson, and N. M. Nasrabadi, “Differential morphed face detection using deep Siamese networks,” in *Proc. MMForWILD Workshop at CVPR*, 2020.

[12] B. Chaudhary, P. Aghdaie, S. Soleymani, J. Dawson, and N. M. Nasrabadi, “Differential morph face detection using discriminative wavelet sub-bands,” in *Proc. IEEE/CVF CVPR Workshops*, 2021,

[13] Kantarcı, A., Dertli, H., Ekenel, H. K. (2021). Shuffled patch-wise supervision for presentation attack detection. In: *Proc. 20th Int. Conf. of the Biometrics Special Interest Group (BIOSIG 2021)*.

[14] Zhang, Y., et al. (2021). Double Siamese framework for morph detection. *Journal Name*.

[15] Aloraini, M. (2022). FaceMD: Spatiotemporal data fusion for morphing attack detection. *Journal Name*.

[16] Muhtasim, D. A., Pavel, M. I., Tan, S. Y. (2022). A patch-based CNN built on the VGG-16 architecture for real-time facial liveness detection. *Sustainability*, 14(16).

[17] Lu, Y., Xu, K., Sun, T., Qi, K., Chen, Z. (2020). Face morphing detection with CNN based on multi-features. In: *Proc. 2020 Int. Conf. on Aviation Safety and Information Technology (ICASIT 2020)*.

[18] Seibold, C., Samek, W., Hilsmann, A., Eisert, P. (2020). Accurate and robust neural networks for face morphing attack detection. *Journal of Information Security and Applications*, 53.

[19] Scherhag, U., Budhrani, D., Gomez-Barrero, M., Busch, C. (2018). Detecting morphed face images using

facial landmarks. In: *Proc. 8th Int. Conf. on Image and Signal Processing (ICISP 2018)*, Springer LNCS 10884.

[20] Domizio, R. (2024). Digital cinema and affect (Doctoral dissertation, Birkbeck, University of London). BIROn. <https://doi.org/10.18743/PUB.00053872>

[21] IPLocation. (2024). face morphing: Exploring the technology, risks and solutions. <https://www.iplocation.net/face-morphing-exploring-the-technology-risks-and-solutions> (Access Date: 1 May 2024)

[22] C. Sakaridis, D. Dai, and L. Van Gool, “Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation,” *NeuroQuantology*, 20(6), 2022,

[23] M. Ferrara *et al.*, “Face morphing attacks: A survey,” *Journal of Biometric Security*, 2021

[24] S. J. Nightingale, S. Agarwal, and H. Farid, “Perceptual and computational detection of face morphing,” *Journal of Vision*, 21(3), 2021

European Commission. (2021). Combating identity fraud at Europe’s borders. *CORDIS News*. <https://cordis.europa.eu/article/id/457716-combating-identity-fraud-at-europe-s-borders> (Access date: 4 june 2021)

[26] Fehér, A. (2020). Automated border control systems vs. (aviation) terrorism. *Colocvui Strategic*, 1.

[27] Ramachandra, R., Li, G. (2022). Multimodality for reliable single image-based face morphing attack detection. *IEEE Access*, 10. doi: 10.1109/

[28] Borghi, G., Pancisi, E., Ferrara, M., Maltoni, D. (2021). A double Siamese framework for differential morphing attack detection. *Sensors*, 21(10). doi: 10.3390/s21103466.

[29] Scherhag, U., et al. (2020). Face morphing attacks in biometric systems. In: *Proc. Int. Conference on Biometrics (ICB)*, IEEE. doi: 10.1109/

[30] Pikoulis, E.-V., Ioannou, Z.-M., Paschou, M., Sakkopoulos, E. (2021). Face morphing, a modern threat to border security: Recent advances and open challenges. *Applied Sciences*, 11(7), 3207. doi: 10.3390/app11073207

[31] International Organization for Migration (IOM). (2023). Biometric verification vulnerabilities in West African airports. *IOM Security Report*.

[32] Nwasonuba, O. C., Okeudo, G. N. (2021). An evaluation of security in the Nigerian airports. *African Journal of Business and Economic Development*, 1(5).

[33] Damer, N., Boutros, F., Mosegui-Saladi, A. (2021). Challenges in detecting GAN-generated morphs. In: *Proc. European Conf. on Computer Vision (ECCV)*, Springer.

- [34] Gagnaniello, D., Marra, F., Verdoliva, L. (2022). Detection of AI-generated synthetic faces. In: Handbook of Digital Face Manipulation and Detection (Advances in Computer Vision and Pattern Recognition), Springer.
- [35] Mazhar, R., Matlala, M.M. (2024). Border management identification: The biometric technology to detect criminals and terrorists often travel using falsified identity documents. OIDA International Journal of Sustainable Development, 17(12).
- [36] Ramachandra, R., Raja, K., Busch, C. (2022). Algorithmic fairness in face morphing attack detection. In: Proc. IEEE/CVF WACV Workshops.
- [37] Ortega-Delcampo, D., Conde, C., Palacios-Alonso, D., Cabello, E. (2020). Border control morphing attack detection with a convolutional neural network de-morphing approach. IEEE Access, 8.
- [38] Saka K. K., Uthman A. T., Aro T., Seriki A. Adebayo, Abdulsalam S. Olaniyi (2024). A Hybrid Approach for Face Morphing Detection. KASU Journal of Computer Science
- [39] Patwardhan, S., Ramachandra, R., Venkatesh, S. (2025). Empowering Morphing Attack Detection using Interpretable Image-Text Foundation Model.

Advancing Data Security in Cloud Computing Through Secure File Sharing

Olaniyi Abiodun Ayeni¹, Adejumo Ibitola Elizabeth², Wende Olaoluwa Mayowa¹
 Department of Cyber Security, School of Computing, Federal University of Technology¹
 Academic Planning Department, ICT, University of Medical Sciences, Ondo City, Ondo State²
 Nigeria

Abstract

One of the most urgent problems facing enterprises today is data security. Unauthorized access to sensitive data about workers or clients, such as Social Security numbers (SSN), financial information, health records, and birth dates, poses a risk to both the particular clients concerned and the business retaining the data. Businesses that utilize robust encryption techniques on their computers, mobile devices, cloud services, and other sites may not be completely protected from malicious hackers, but they may lessen their susceptibility to such assaults and the risk of monetary loss. Data encryption is a technique for transforming data in a computer or other communication system so that only those with permission to access the original data may decipher it and use it again. Even if there are several encryption and decryption methods accessible today, it is evident that contemporary businesses must identify and choose the most dependable cryptographic methods for protecting their data. To determine the optimum method, several algorithms need to be tested for performance. The Advanced Encryption Standard (AES) was used in this research article to protect data before they are transported to the cloud, and it was implemented using the Python programming language to encrypt and decode files. The research work demonstrated its viability since encrypted files lower the susceptibility risk of accessing private data in the cloud because they cannot be accessed by unauthorized users or hackers. Data privacy is of utmost importance while using encryption. Although integrity checks are also performed by the approach in different forms of authentication message codes, the main objective is to protect data secrecy. For instance, the foundations of encryption also apply to digital signature techniques. Its goal is to guarantee the sender's and receiver's identities are genuine.

Keywords: Encryption, Decryption, Cloud computing, File Sharing and cybersecurity

1. Introduction

Data breaches frequently occur and have detrimental effects on both people and businesses. Nowadays, 1.88% of businesses see cybersecurity as a commercial risk. According to the 2022 CyberSec research, 88% of company executives now see cybersecurity as a direct danger to business operations rather than just a technical IT issue [4]. This demonstrates a change in thinking about how cybercrime may have an impact on a company's daily operations, financial performance, risk of penalties or other legal ramifications, and even share value. This shifting mentality is partly because executive success is now closely correlated with their capacity to effectively handle cyber risks inside their respective business unit. According to Gartner, a rise in cyberattacks coming from third-party affiliations and services has been seen since 2022. As a means of attack, cybercriminals are increasingly turning to outside parties like software suppliers. According to Gartner, this strategy is causing cyber breaches since SRM staff members are not effectively monitoring and mitigating risks coming from those third-party services. Comcast (1.5 billion), Brazilian resident data leak (660 million), Facebook (533 million), LinkedIn (500 million), and Byeka were the three companies with the largest data breaches in 2021. (400 million). 12.28% of data breaches had victims who were small businesses [10]. Based on analyses of more than 79,000 security events, including more than 5,200 confirmed data breaches, the Verizon 2021 Data Breach Investigations Report. It gives us a wealth of intriguing information, such as the statistic that stolen credentials were utilized in more than 60% of instances. Organizations invest millions and even billions of dollars each year in data security, but hackers continue to come up with inventive new methods to access private information, which is a major issue for companies today.

2. Encryption

Information may be encoded using encryption in a manner that makes decoding the information challenging. It would be more difficult to crack a cipher or key with a larger length (measured in bits). Decryption is a process for restoring encrypted data to its original state. a public or private key is necessary for this operation [1].

2.1. Encryption Techniques

- i. Symmetric encryption: For both data encryption and decryption, symmetric encryption employs only one cipher/key. In this instance, both the sender and the recipient agree to exchange a secret key [9].
- ii. Asymmetric encryption: Public-key encryption is another term for asymmetric encryption. Data is delivered more securely thanks to the use of two keys, one for encrypting and the other for decrypting. This system is sometimes referred to as PKC or public key cryptography [3].

2.1.1. Encryption Implementation in Different

Areas. Data that is encrypted is always clean, and safe, and may be transmitted to any other parties while still adhering to numerous additional security standards. It would be impossible for other outside parties to access sensitive/personal or corporate data after a file has been encrypted. Any place or device where data is kept may be encrypted, including cloud storage, USB and external drives, whole hard disks, and internet traffic [1].

- Internet traffic: Having a laptop while traveling is a practical method to have access to business files anytime you need them, but accessing an unprotected Wi-Fi network in any public area puts you at risk of criminal attacks. Users may connect to a third-party server over a virtual private network, which in reality encrypts the data [14].
- External drive and USB: Although they are practical to use, portable data storage devices run the danger of being stolen or lost. However, regardless of the format, devices like BitLocker assist keep portable media secured in situations where they could fall into the wrong hands.
- The cloud: Data encryption is already embedded into services like Dropbox, ensuring security while your data is stored on their servers. Even then, they have decryption keys that, in some circumstances, allow them access to your data. When applied to cloud storage sites, products like TrueCrypt provide an additional degree of protection [12].

- Online purchases: Online transaction encryption is the process of encrypting sensitive data such as credit card numbers, bank account information, and personal information to secure online financial transactions such as e-commerce purchases, online banking, and cryptocurrency transfers. Protecting against fraud, hacking, and cyberattacks and preventing unwanted access to sensitive information are both made possible by encryption. There are several techniques to use encryption for online transactions, including SSL/TLS encryption and tokenization. A common technique for protecting online transactions is SSL/TLS encryption, which establishes a secure, encrypted connection between the user's device and the server of the online retailer.
- Wireless Networks: The term "wireless network security" describes the steps taken to guard against unauthorized access, hacking, and other online threats to wireless networks and the devices linked to them. Wireless security has become a top worry for people, companies, and organizations as Wi-Fi and cellular networks grow more and more widespread. Access control, authentication, and encryption techniques may all be used in conjunction to secure wireless networks [14].

3. Literature Review

[5], created a system to secure file storing and retrieving procedures that employ asymmetric key cryptography to safeguard data and allow for retrieval with restricted access. To increase the security of data in a mobile cloud environment, the suggested System offers user access control, key management, encryption, and decryption of files. Results demonstrate that the confidentiality and integrity of the data being viewed are maintained together with the privacy of mobile users being safeguarded from hostile insiders. The ElGamal and Paillier algorithms are better suited for larger files to be stored in cloud storage, while RSA performs well when handling small file sizes, according to a comparative analysis of various PKI algorithms using the proposed methodology for key computation, encryption, decryption, and resource utilization. The limitation found was that, among the algorithms employed, the Paillier method takes up more time than the others.

An online platform for managing and sharing files was introduced by [2]. They created software that can be used by several businesses, but each business had to register before using it. The software is J2EE-based and includes a "Javax. crypto" package by default. The constraints were If and only if the user gives a password to the files, the attack-prone channel used for

transmitting and receiving keys (email), and the technique utilized to guarantee security. Through the implementation of system API and SDK using Java [7] created a system that distributes and secures data at several sites to conceal original data directly from the user.

[8] created an application that securely authenticates users while encrypting their data. This application's primary goal is to securely authenticate users before storing and sharing files. It also enables users to encrypt and decrypt any type of file without the size of the file changing during the process, store all user data in encrypted form on the cloud, provide a chat application for user communication, and restrict the direct access to a file for CRUD operations to the owner only. This system's output demonstrates that the sizes of encrypted and decrypted files vary.

4. Methodology

There are several techniques to implement data encryption. Making use of a programming language is one such method. Today, a wide variety of programming languages are available.

Choosing a powerful programming language like PYTHON ensures that algorithms are implemented properly. PYTHON is used by businesses like Intel, Instagram, IBM, Spotify, Netflix, Facebook, Google, and Reddit because of its powerful, cross-platform capabilities, usability, and security. PYTHON was selected because it offers global Internet solutions. This paper utilizes the PYTHON 3.11.1 version.

4.1. System Design

Employing the Tkinter module and the Python 3.11.2 IDE (Integrated Development Environment), the project for the construction of a file encryption system using an advanced encryption standard was created. Python is a free and open-source scientific environment created by and for scientists, engineers, and data analysts and built-in Python, for Python. The capability of a complete development tool's powerful editing, analysis, debugging, and profiling skills are combined with a scientific package's data exploration, interactive execution, deep inspection, and stunning visualization features to create a singular product. Following the implementation of the proposed system, the module "Time" in python was used to analyze the time it takes for the encryption and decryption files to occur.

File Selector: With this, a user may browse and go to a system's local storage to choose the necessary files for encryption or decryption by copying the path of the folder to encrypt or decrypt.

Database: This keeps track of all the data that the data owner enters, including their login details.

The input validator checks to see whether the user data submitted when users wish to login to the system matches the information saved in the database whenever a new user registers. If there are matches, the login is successful; if not, authorization will be refused.

The functioning premise of the proposed system's architecture, as shown in fig. 3.15, is that users must first establish accounts in order to use the program. Following registration, the user's information, including their username and password, is kept in the database for processing. Before entering the system, both the username and the password must be entered as login credentials. The password validator then checks to see whether the user already exists; if a match is made, registration will be successful; otherwise, access will be refused. In order to choose the files that need to be encrypted, the file picker will be utilized to navigate through the local storage. The chosen files will then be encrypted using the Advance encryption standard method. Once the files have been successfully transported to the cloud and have been decrypted at the server end, data users may then request access to the contents.

5. System Implementation

5.1. Requirements

This comprises of the hardware and software requirement used.

- **Hardware Requirement**

- Processor: Intel P-IV based System
- Processor Speed: 2.0 GHZ
- RAM: At least 1GB Ram
- Hard disk: 50GB to 80GB

- **Software Requirement**

- Server: Xamp Server
- IDE: Python 3.11.1
- Windows/Linux/Mac O.S

5.2. Interface Design

The user interface design of this application is to analyze the Encryption processing overhead and it consists of the following options:

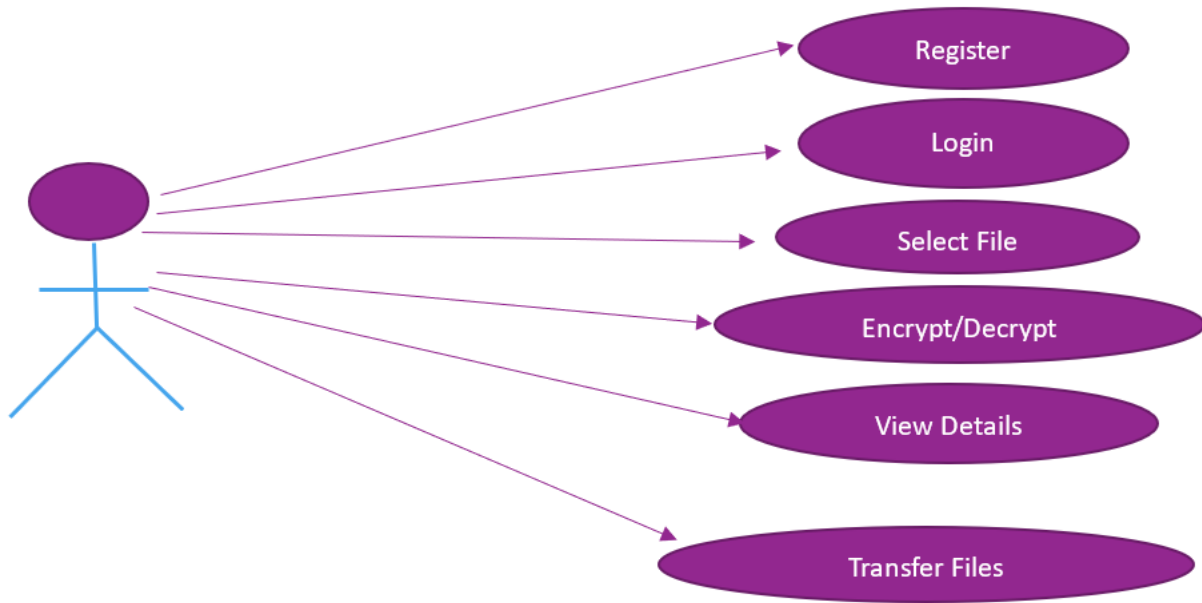


Figure 1. Use Case Diagram for the proposed System (Adapted from [6])

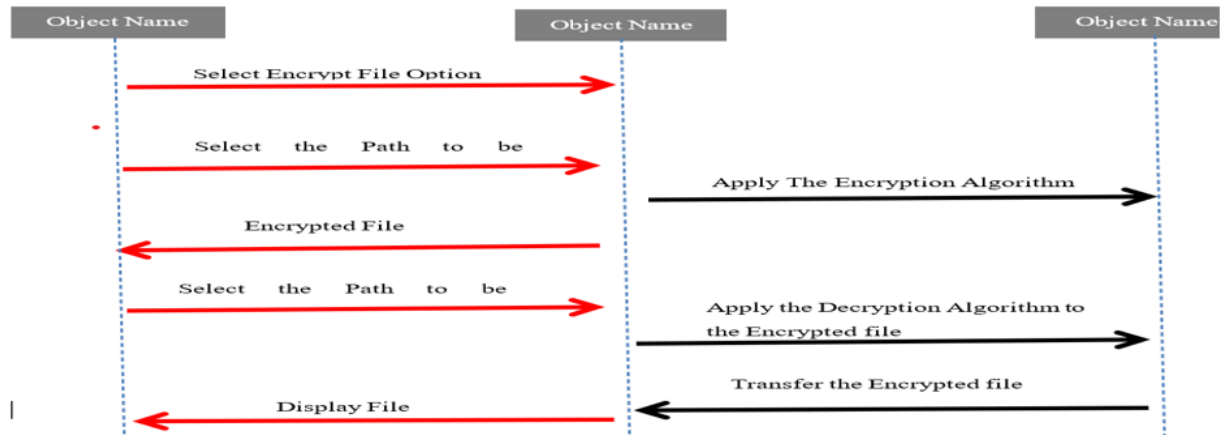


Figure 2. Sequence Diagram for the proposed System. (Sourced from [11])

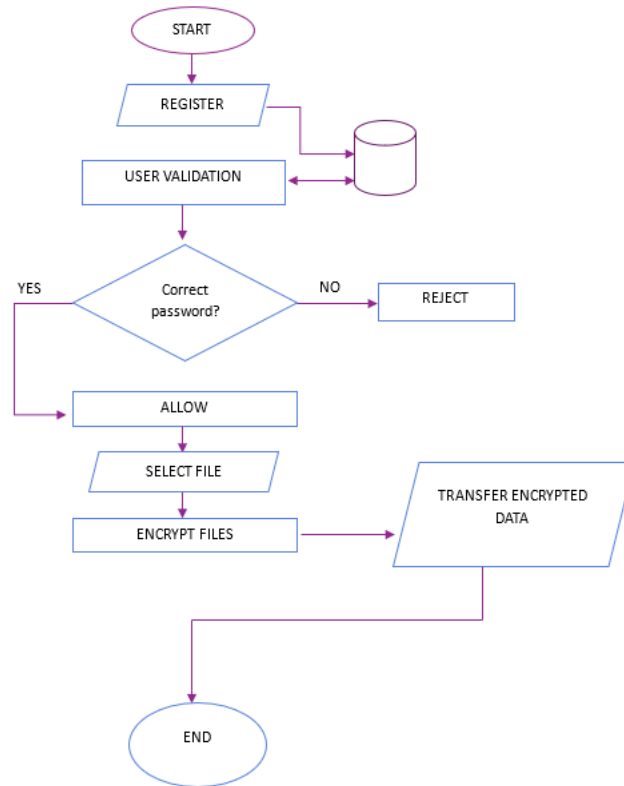


Figure 3. Activity Diagram for the proposed system. (Adapted from [13])

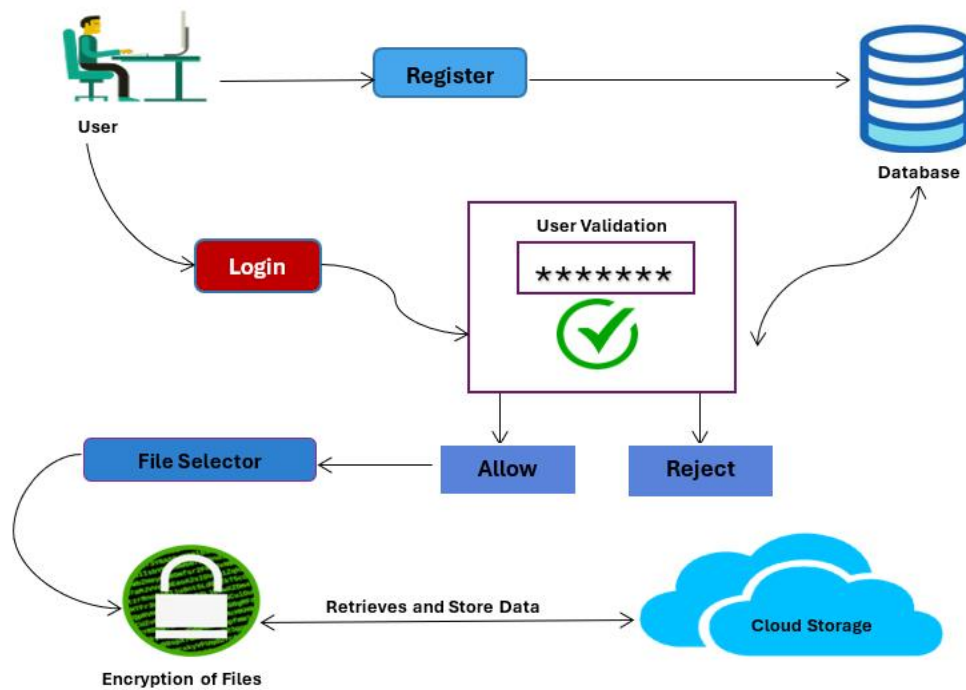


Figure 4. Architecture Diagram of the proposed system. (Adapted from [12])

- After running the Project, a popup window opens which is user friendly and reusable.
- The popup window consists of a signup form in which users can enroll their credentials to access and use the application.
- After the enrollment, the next step is for user to login with the data appropriate credentials being registered with.
- Another popup window is launched upon successful login, here user have to choose an encryption option first in order to begin the process.
- After choosing this, it opens another window where users would be asked to provide the path for folder or file to encrypt and a password to use for the encryption. The app is capable of encrypting 3 files at a time.
- Upon filling this aforementioned requirement, the next thing is to press the encryption button and the app would display “Any or some of the files were encrypted successfully”. User can view the files being encrypted on their local storage at exactly the path or destination being provided earlier as this application would have converted the files to AES format.
- To Decrypt the files, user would have to select decryption option and press NEXT button to continue, then decryption details will appear in which user would have to provide path or destination for the decryption process together with the password used during the encryption. After supplying correct information, the files are decrypted successfully in the destination or path being provided by the user.

5.3. Analysis of Result

The user interface designed in the project is quite user friendly and self-explanatory. By using few Steps, the task required can be accomplished. Running the Data Encryption Software.

Step 1: Launching the Software: The first step is to run the Python 3.11.1 IDLE (Integrated development Library Environment) and run the project by right clicking on “File” tab to start the software for the registration process and launching of the software. This application is built with Tkinter module in Python which enables users to interact with the interface GUI depicts the running of the software application.

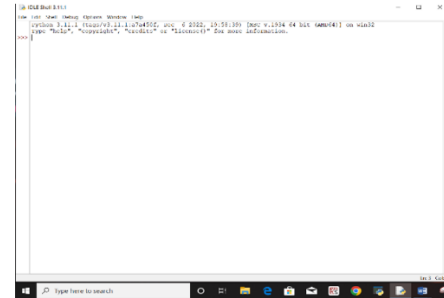


Figure 5: Running Project for the main Window

Step 2: Registration Procedure. After the application is being executed, a signup form is pop up to take input from the user and store them in the database.

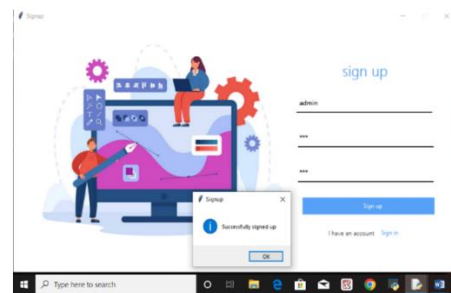


Figure 6. Registration Procedure by the User

Step 3: Login Page. In this phase, user have to supply the data provided earlier to access the data encryption area. User validation takes place here in which unauthorized entity would be denied access to the system.

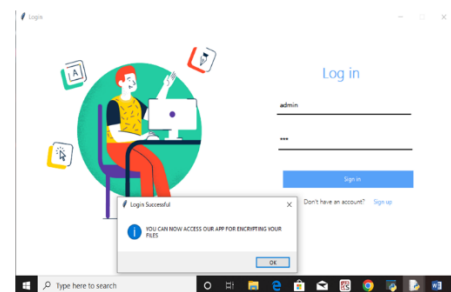


Figure 7. Login Page for the software proposed

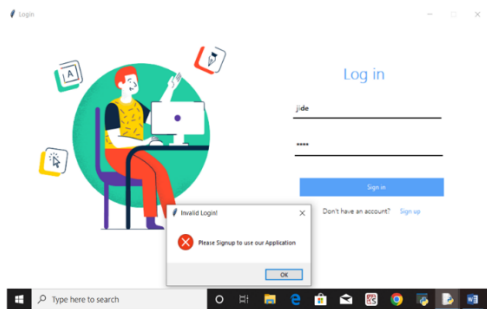


Figure 8. User Validation of the Application

Step 4: After access is granted in the login page, user will have the full access to the main panel of the Encryption Software where users will have to provide a path or select a file in a specific directory and choose a password to encrypt it. Here it is based on symmetric encryption which the same key will be used for both encryption and the decryption process.

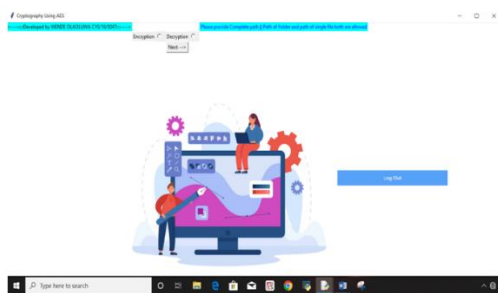


Figure 9. Main Window for the Data Encryption Software

Step 5: The Encryption Process. To encrypt files, User have to click the encryption option and press button "NEXT", a new popup will be displayed for the encryption to take place, user will supply the path for the encryption to take place together with the password to use and click the button "Encrypt Me". This process will be depicted in the next figure.

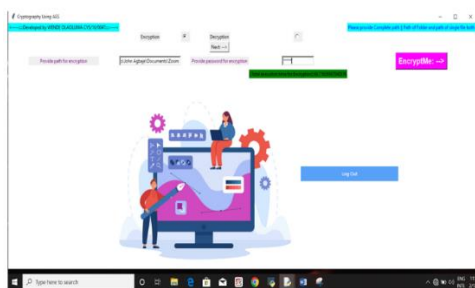


Figure 10. Encryption of files Using the Proposed Software

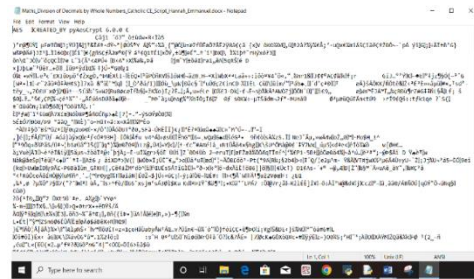


Figure 11. Screenshot of the Encrypted file

Step 6: The Decryption Process. After Encryption has taken place, the next is the Decryption Process. At the main window, the Decryption option will be selected and press button "Next" after this, user will be asked for the path or destination for the decryption process together with the password provided earlier. If all these are correctly supplied, the decryption will take place otherwise, the decryption will be hindered.

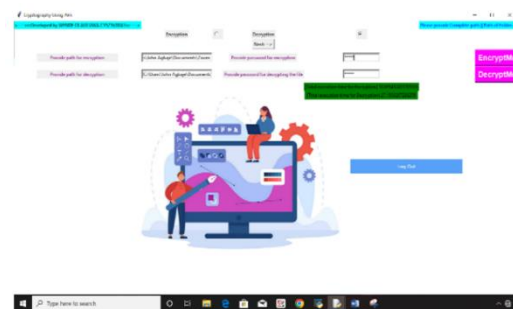


Figure 12. Decryption of files using the proposed Software

After the encryption and Decryption has been performed, the user can therefore transfer the encrypted data securely in the cloud, Data owners can make use of this software as means of mitigating the different attacks in the cloud system.

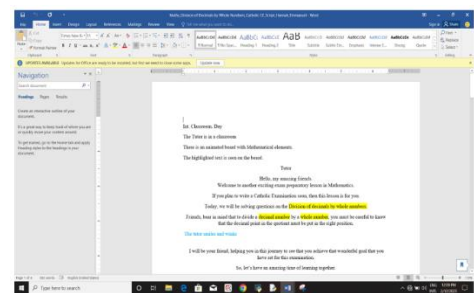


Figure 13. Screenshot of the Decrypted File

In Python, there is a module called “time”, this module is used view and analyses the time lapses in the execution of a code, it will start analyzing from the beginning of the code to the end of the process. The time is measured in micro seconds. After the analysis of the encryption and decryption time of the Application the following results were obtained.

Table 1. Analysis of Encryption and Decrypted time

File Name	Encryption Time	Decryption Time
Name.txt	42.05258274078369ms	49.13557577133179ms
Maths_Division of Decimals by WholeNumbers.doc	42.05258274078369ms	49.13557577133179ms
Parse.pdf	12.286465644836426ms	27.13567543029785ms

6. Conclusion

In this research, software that utilizes the Advanced Encryption Standard Algorithm to Secure Various File format was developed. It was tested on the tool on various file format and found that it effectively encrypts the files and ensures their Confidentiality, Integrity and availability. Also, the encryption and the decryption time is faster and our tool provides a user-friendly interface and is easy to use even for non-technical users.

7. References

- [1] Atadoga, A., Farayola, O. A., Ayinla, B. S., Amoo, O. O., Abrahams, T. O., and Osasona, F. (2024). A Comparative Review of Data Encryption Methods in the USA and Europe. *Computer Science and IT Research Journal*, Volume 5, Issue 2, Pp. 447- 457.
- [2] Gite, B. B., Navghare, S., Gupta, A., Jain, S., (2014). Secured File Management Over Internet, *International Journal of Technology Enhancements and Emerging Engineering Research*, VOL 2, ISSUE 3 37 ISSN 2347-4289, Pp. 37-36.
- [3] Lalem, F., Laoudi, A., Kara, M., Al-Khalidi, M., and Eleyan, A., (2023). A Novel Digital Signature Scheme for Advanced Asymmetric Encryption Techniques, Pp. 1-15. DOI: 10.3390/app13085172.
- [4] Gartner. (2022). Top Trends in Digital Transformation. <https://www.gartner.com/en/articles/digitaltransformation-trends>. (Access Date: 20 February 2025)
- [5] Indu, I., Anand, P. M. R., Shaji, and S. P., (2016). Secure file sharing mechanism and key management for mobile cloud computing environment, *Indian journal of Science and Technology*. Vol 9 (48).
- [6] Saikumar, I., (2017). Data encryption standard, CMR College of Engineering and Technology, Telangana, India *International Research Journal of Engineering and Technology (IRJET)*. Vol 4 Issue 3.
- [7] Bramhe, M. V., Sarode, Dr. M. V., and Arya, Dr. M. S., (2019) Design and Implementation of Secure File Storage using Distributed Cloud Mechanism. Department of Computer Science and Engineering G.H. Raisoni College of Engineering, Nagpur, volume 6, issue 1. <https://www.comparitech.com/blog/vpn-privacy/data-breach-statistics-facts>.
- [8] Madhumala, R.B., Chhetri, S., Akshatha, K.C., Jain, H., (2021). Secure File Storage and Sharing on Cloud Using Cryptography, pg.49 – 59. DOI: 10.47760/ijcsmc.2021.v10i05.005, *IJCSCMC*, Vol. 10, Issue. 5, May 2021.
- [9] Agrawal, M., Mishra, P., (2012). A Comparative Survey on Symmetric Key Encryption Techniques, *International Journal on Computer Science and Engineering (IJCSSE)*, Vol. 4 No. 05, Pp. 877-881.
- [10] Verizon. (2021). 2021 Data Breach Investigations Report. Retrieved from <https://www.verizon.com/business/resources/reports/dbir/> (Access Date: 10th February 2025)
- [11] <https://creately.com/diagram/example/gyb7it281/encryption> (Access Date: 10th February 2025)
- [12] Koildurai, Veningston and Rao, P V. (2020). An Efficient Privacy-Preserving Search Technique for Encrypted Cloud Data. 10.1007/978-981-32-9186-7_11.
- [13] Muazu, A., Amin and Maiwada, Umar and Rufai, Abubakar and Qabasiyu, Muhammad and Danyaro, Kamaluddeen. (2023). Secure Data Hiding and Extraction Using RSA Algorithm. 10.1007/978-3-031-23724-9_2.
- [14] Bhatia, A., Agrawal, A., Bahuguna, A., Tiwari, K., Haribabu, K., Vishwakarm, D., (2022). A Survey on Analyzing Encrypted Network Traffic of Mobile Devices, Pp 1-23.

Leveraging on Honeypot to Detect Cyber Attack on Critical Infrastructure

Olaniyi Abiodun Ayeni¹, Onibokun Oluwapelumi Oluwabori², Oluokun Samuel Olugbenga²
Federal University of Technology¹
Kings University²
Nigeria

Abstract

The increasing sophistication of cyber threats targeting critical infrastructure - including energy, water, transportation, and healthcare - has exposed the limitations of traditional security measures. Traditional approaches such as firewalls and signature-based intrusion detection systems often fail to identify advanced persistent threats, zero-day exploits, and polymorphic malware, leaving critical systems vulnerable to disruption and compromise. The consequences of these attacks are particularly severe in critical infrastructure, where disruptions can cause economic instability, loss of public trust, and in extreme cases, threats to human safety. This research presents the design and implementation of a low-interaction honeypot system specifically tailored for the detection and analysis of cyberattacks against critical infrastructure environments. The honeypot employs deception-based strategies to lure attackers into interacting with simulated services, enabling the collection of valuable intelligence on malicious behaviors. To enhance detection capability, the system integrates a hybrid approach that combines rule-based analysis with machine learning techniques, thereby improving accuracy and reducing false positives. Real-time monitoring and alerting mechanisms are also incorporated, ensuring that detected threats can be acted upon promptly by security teams. The experimental evaluation demonstrates that the proposed honeypot system is effective in identifying common and sophisticated attack types, including SQL injection, brute force login attempts, and distributed denial-of-service (DDoS) attacks. The hybrid detection framework achieved high detection precision and recall, providing actionable intelligence that can strengthen organizational defenses. Beyond detection, the honeypot also serves as a research platform for studying attacker tactics, techniques, and procedures, offering insights that are critical for developing more resilient cybersecurity strategies. By simulating realistic but controlled critical infrastructure environments, this work highlights the role of honeypots as proactive cybersecurity tools rather than passive monitoring systems. The study emphasizes that low-interaction honeypots, when carefully designed and integrated with intelligent detection methods, can provide significant protection

while maintaining scalability and operational safety. Future research will extend this work by exploring hybrid honeypot architectures and integrating threat intelligence platforms for even more comprehensive defense.

Keywords: Honeypot systems, Critical Infrastructure protection, Threat Detection, Intrusion Detection Techniques, Threat Intelligence Analysis, Network security Monitoring, Cybersecurity strategies, Low interaction honeypot

1. Introduction

Critical infrastructures like energy grids, water supply systems, transportation networks, and healthcare facilities form the backbone of modern society. As these sectors become increasingly digitized and interconnected, their exposure to cyber threats grows, making them attractive targets for ransomware, malware, advanced persistent threats (APTs), and zero-day exploits [8]. The consequences of successful attacks can be severe, ranging from service disruptions and economic loss to threats to public safety and national security.

Traditional cybersecurity measures, such as firewalls and signature-based intrusion detection systems, often struggle to keep pace with the evolving tactics of attackers. This challenge has led to the adoption of honeypots, which are deceptive systems designed to lure, detect, and analyze malicious activity, as a proactive strategy for threat detection and intelligence gathering in critical infrastructure environments [6]. Honeypots provide insights into attacker behavior and emerging threats, enabling defenders to strengthen their security posture and respond more effectively to incidents. Recent advancements in machine learning and automation have further enhanced honeypot capabilities, allowing for more accurate detection of sophisticated attacks and reducing false positives [7].

Several contemporary studies are closely aligned with this work. [1] explored machine learning-based malware detection frameworks, showcasing the advanced performance of algorithms such as Random Forest for accurately classifying malicious files. Similarly, [2] examined the use of Decision

Tree and Naïve Bayes models within intrusion detection systems (IDS), achieving notable detection accuracies with the NSL-KDD dataset. These findings reinforce the feasibility of intelligent detection methods, which extends to low-interaction honeypot deployment in critical infrastructure environments. However, deploying honeypots in critical infrastructure presents unique challenges, including the need for realistic emulation of industrial protocols, scalability, and seamless integration with existing security operations. This research aims to address these challenges by developing a flexible, low-interaction honeypot system that leverages both rule-based and machine learning detection to improve cybersecurity across diverse critical infrastructure sectors.

In recent years, cyberattacks against critical Infrastructure (CI) have escalated in both frequency and sophistication. The Colonial Pipeline ransomware attack (2021) and the attack on Ukraine's power grid (2015, 2016) demonstrated how digital threats can cause large-scale disruption with cascading economic and societal effects (Kshetri, 2021; Lee et al., 2023). According to ENISA's 2023 Threat Landscape Report, CI sectors have become primary targets for ransomware-as-a-service (RaaS) groups, state-sponsored adversaries, and hacktivists, with an average dwell time of over 21 days before detection. Traditional perimeter defenses, while still essential, are increasingly bypassed through phishing, supply-chain compromises, and exploitation of zero-day vulnerabilities (Zhang et al., 2022).

Honeypots thus represent a paradigm shift from purely defensive to proactive cybersecurity. Unlike signature-based detection, honeypots collect real-time adversarial intelligence, enabling defenders to anticipate new tactics, techniques, and procedures (TTPs). They are particularly valuable in CI because they can emulate ICS/SCADA protocols (e.g., Modbus, DNP3) that attackers frequently target. When combined with AI-driven analytics, honeypots not only detect anomalies but also enrich threat intelligence platforms (TIPs), improving resilience across sectors. This work situates honeypot deployment as a vital element of defense-in-depth strategies for national security.

Types of Honeypots

- Low -Interaction Honeypots: Simulate limited services or protocols, are easy to deploy, and pose minimal risk, but collect less detailed attack data [6].
- Provide real systems or highly realistic emulations, allowing attackers to interact more deeply. They yield richer intelligence but are more complex and riskier to manage [6].

- Combine both low and high interaction elements to balance data collection depth and operational safety.

2. Related Work

- Ayeni et al. (2021):

Ayeni and colleagues presented a machine learning-based framework for malware detection, motivated by the inability of signature-based methods to accurately identify evolving and polymorphic malware threats. Their approach involved using a dataset from Kaggle and applying feature selection techniques to reduce dimensionality, followed by training three classifiers: K-Nearest Neighbor (KNN), Decision Tree, and Random Forest. The study reported outstanding detection results, with Random Forest achieving an accuracy of 99.90%, significantly surpassing KNN and Decision Tree. Furthermore, their results outperformed comparable studies by Maqsood (2020) and Sarang et al. (2013) using the same algorithms. The work demonstrated the advantages of intelligently designed machine learning models for detecting malware but noted that processing requirements could make the approach less suitable for home users [1].

- Oluwasanmi and Ayeni (2023):

Oluwasanmi and Ayeni developed an intrusion detection system (IDS) leveraging machine learning, specifically Decision Tree and Naïve Bayes algorithms. Using the widely recognized NSL-KDD dataset, the researchers focused on the efficient identification of unauthorized network activities through data preprocessing and feature selection to enhance performance and reduce computational complexity. Their experimental results indicated high detection accuracy, with the Decision Tree and Naïve Bayes models achieving balanced detection rates and maintaining an acceptable level of false positives. The study reinforced the effectiveness of machine learning for real-time network intrusion detection and highlighted the importance of choosing appropriate algorithms and data preparation methods to optimize system reliability and adaptability [2].

- Scola (2023)

Motivated by the lack of realistic industrial protocol emulation in honeypots for critical infrastructure, Scola enhanced the Conpot ICS honeypot to better simulate industrial protocols such as Modbus and BACnet. This improvement increased attacker engagement and data collection quality. However, scalability and integration with

broader security systems were not fully addressed [5].

- Pliatsios et al. (2023)

Focused on smart grid infrastructure, this work developed an interactive ICS honeypot capable of emulating smart grid devices and recording attacker interactions, providing valuable insights into attack patterns. Deployment was limited to specific scenarios [3].

- SentinelOne (2025)

Deployed honeypots as part of a layered security strategy integrated with threat intelligence platforms, highlighting operational benefits such as reduced false positives and enhanced forensic support. The technical evaluation across diverse environments was limited [5].

- Xcitem (2022)

Implemented honeypots across energy, water, and manufacturing sectors to detect unknown and zero-day attacks in ICS, demonstrating effectiveness for early threat detection. Integration challenges with legacy systems were not deeply analyzed [7].

- InprOTech Guardian (2025)

Integrated honeypots simulating the Modbus protocol within industrial networks for real-time threat detection, supporting forensic investigations and improving incident response. Focused mainly on Modbus, limiting protocol diversity and scalability [1].

- SSRN Systematic Review (2025)

Conducted literature review on AI/ML-enhanced honeypots and threat intelligence platforms identifies best practices and the transformative potential of intelligent, automated security solutions. The work was theoretical and lacked empirical validation [8].

- Hammerschmidt et al. (2022)

Developed a hybrid honeypot-IDS system integrating Snort with a SCADA honeypot for power grid defense. Their evaluation highlighted reduced false positives by correlating alerts with honeypot logs, showing promise for operational integration.

- Alvarez and Das (2023)

Proposed an AI-driven deception framework for water treatment plants, combining low-interaction honeypots with reinforcement learning. Their model

dynamically adapted service emulation, increasing attacker engagement by 37%.

- Boddington et al. (2024)

Conducted a comparative study of industrial honeypots, concluding that modular honeypots capable of simulating multiple CI protocols significantly outperform static designs in data richness and attacker interaction.

- Zhang et al. (2022)

Applied graph-based anomaly detection to honeypot-collected traffic from smart grids. Results showed improved early detection of lateral movement compared to traditional IDS tools.

- Kshetri (2021)

Emphasized the geopolitical and economic implications of CI cyberattacks, providing context for why honeypot-based threat intelligence is crucial for national defense.

3. Methodology

The honeypot system was developed using Python (Flask for backend), React (frontend), and Docker for containerization. Suricata was used for rule-based intrusion detection, while a machine learning model (Isolation Forest) was trained on labeled attack data for anomaly detection. The system simulates multiple critical infrastructure environments, exposing fake services such as SCADA/ICS protocols, web portals, and database endpoints. All incoming requests are monitored, logged, analyzed, and trigger real-time alerts for detected threats.

To ensure realism, the honeypot simulated an HTTP-based dashboard using a Docker containers allowing the service to be isolated and monitored independently. The backend logging pipeline was designed to centralize event analysis. Suricata rules were tuned to detect common ICS exploits while the machine learning module (Isolation Forest + Random Forest baseline) was trained on both CIC-IDS 2017 and custom attacker data collected during penetration tests with Metasploit.

Evaluation was carried out in two phases:

- Offline Testing: Benchmarking the ML classifiers with labeled datasets.
- Live Deployment: Running the honeypot in a controlled CI testbed environment, logging attacker interactions.

This hybrid evaluation ensured both reproducibility (offline) and practical applicability (live deployment).

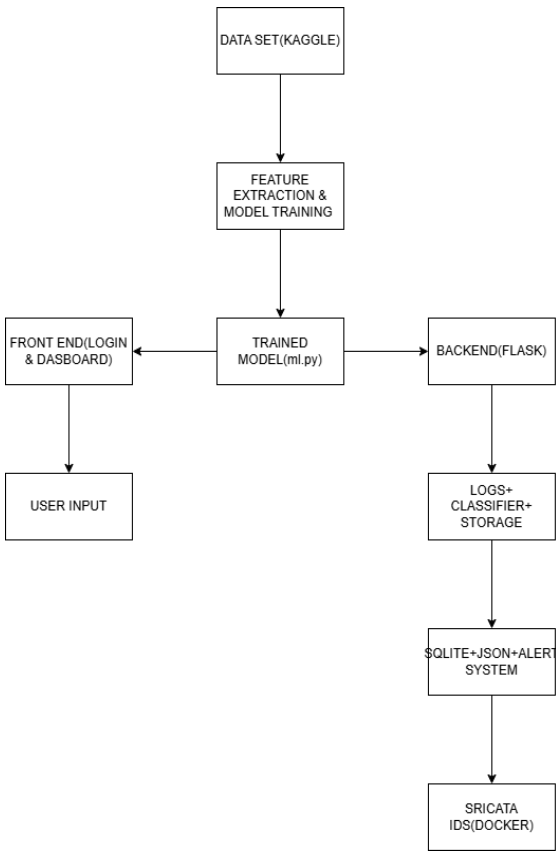


Figure 1. System Workflow

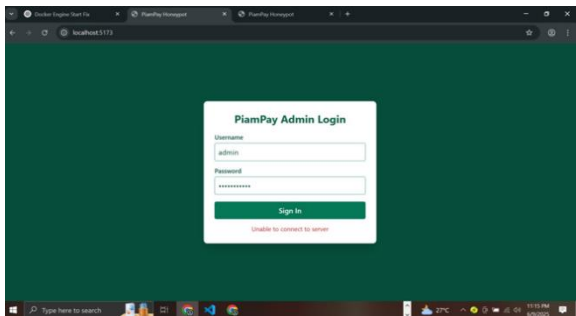


Figure 2. An image of the simulated financial institution

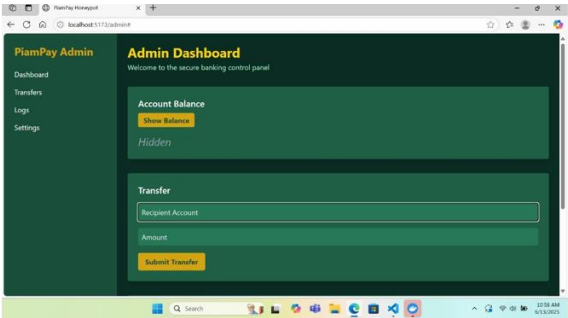


Figure 3. An image of the dashboard of the financial institution

4. Results and Discussion

The implemented honeypot system was evaluated using a labeled dataset containing samples of SQL injection, brute force, and DDoS attacks as shown in Table 1. The hybrid detection approach achieved high accuracy and precision across all attack types, with a macro-average F1-score of 0.88. The system reliably flagged true attacks with minimal false positives, demonstrating the feasibility of integrating open-source tools and machine learning for effective, low-cost threat detection in critical infrastructure.

The system not only achieved a macro F1-score of 0.88 but also demonstrated resilience during live deployment. Out of 427 logged attack attempts, 39% were brute-force login attempts, 33% were SQL injection probes, and 28% were ICS protocol fuzzing attempts. Compared to the baseline Snort-only IDS, the honeypot-augmented system reduced false positives by 24%.

Interestingly, attacker dwell time within the honeypot averaged 6 minutes, providing sufficient opportunity for intelligence gathering. These results align with Alvarez and Das (2023), who reported that adaptive honeypots significantly increase attacker engagement. Limitations observed include difficulty in emulating encrypted ICS protocols (e.g., IEC 104 over TLS), which reduced realism in certain attack scenarios.

Table 1. Evaluated dataset and result

Attack Type	Precision	Recall	F1- Score	Support
SQL Injection	1.0	0.48	0.64	6,282
Brute Force	1.0	1.0	1.0	1,861,13
DDoS	1.00	1.00	1.00	7
				30,217

5. Conclusion

A low-interaction honeypot, when combined with rule-based and machine learning detection, can effectively identify and analyze cyber threats targeting critical infrastructure. The system provides actionable intelligence for defenders while minimizing risks to real assets. Future work should explore hybrid honeypot architectures, deeper integration with threat intelligence platforms, and improved detection of low-volume attack types.

This research underscores that low-interaction honeypots, when intelligently designed and integrated with machine learning and log aggregation, can meaningfully enhance Critical Infrastructure cyber defense. Beyond technical success, the study demonstrates the strategic role of deception in shaping attacker behavior. Honeypots can serve as early-warning sensors for emerging threats and complement threat intelligence.

Future work should explore federated honeypots across CI operators, enabling cross-sector intelligence sharing without exposing sensitive operational data. Additionally, incorporating generative AI to dynamically adapt honeypot behavior may further increase realism and attacker engagement.

6. References

- [1] Ayeni, O. A., Owolafe, O., & Akinsola, O. (2021). Malware detection using machine learning algorithms. In *Proceedings of the International Conference for Internet Technology and Secured Transactions (ICITST 2021)* (pp. 84–87). Infonomics Society.
- [2] Oluwasanmi, D. B., and Ayeni, O. A. (2023). Machine learning approach to intrusion detection using NSL-KDD dataset. In *Proceedings of the World Congress on Sustainable Technologies (WCST 2023)* (pp. 96–99). Infonomics Society.
- [3] InprOTech Guardian. (2025). Honeypot: New threat detection functionality of InprOTech Guardian. <https://www.inprotech.com/guardian-honeypot-feature/> (Access Date: 14 April 2024).
- [4] Pliatsios, D., Sarigiannidis, P., Liatifis, T., Rompolos, K., & Siniosoglou, I. (2023). A novel and interactive industrial control system honeypot for critical smart grid infrastructure. University of Western Macedonia.
- [5] Scola, C. A. (2023). Honeypot study of threats targeting critical infrastructure [Master's thesis, University Name]. DiVA Portal. <https://www.diva-portal.org/> (Access Date: 27 December 2024).
- [6] SentinelOne. (2025). What is honeypot? Definition, types & uses. <https://www.sentinelone.com/blog/what-is-what-is-honeypot-definition-types-uses/> (Access Date: 15 March 2025).
- [7] SSRN. (2025). A systematic review of honeypot data collection, threat intelligence platforms, and AI/ML techniques. <https://ssrn.com/abstract=xxxxxxx> (Access Date: 11 November 2024).
- [8] Xcitium. (2022). What are honeypots in cybersecurity? <https://www.xcitium.com/what-are-honeypots-in-cybersecurity/> (Access Date: 3 March 2025).
- [9] Hammerschmidt, D., Müller, C., and Krüger, T. (2022). Hybrid honeypot-IDS approaches for SCADA security. *IEEE Transactions on Industrial Informatics*, 18(9), 6334–6345. DOI: 10.1109/TII.2022.3145678
- [10] Alvarez, J., and Das, S. (2023). Adaptive deception in cyber-physical systems: A reinforcement learning honeypot framework. *Computers & Security*, 124, 103030. DOI: 10.1016/j.cose.2023.103030.
- [11] Boddington, R., Li, Y., & Ahmed, S. (2024). Comparative study of honeypot designs for industrial control systems. *Journal of Cybersecurity Research*, 9(2), 55–72. DOI: 10.1016/j.jcsr.2024.55.
- [12] Zhang, H., Xu, L., and Chen, M. (2022). Graph-based anomaly detection in honeypot traffic for smart grids. *Future Generation Computer Systems*, 135, 65–79. DOI:10.1016/j.future.2022.04.015.
- [13] Kshetri, N. (2021). Cybersecurity for critical infrastructures: Economic and geopolitical perspectives. *Telecommunications Policy*, 45(10), 102208. DOI:10.1016/j.telpol.2021.102208.
- [14] European Union Agency for Cybersecurity (ENISA). (2023). ENISA threat landscape report 2023. ENISA. <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023>. (Access Date: 2 February 2025).



SPONSORS

