

Volume. 10 No. 1  
January - June 2023

ISSN-2349-7912  
E-ISSN-2350-112X  
Subscribers copy  
Not for sale



i-manager's  
Journal on  
Pattern Recognition

Interpretation of Information Handling and Retrieval Applications



## **i-manager's**

# **Journal on Pattern Recognition**

### **About the Journal**

*i-manager's Journal on Pattern Recognition provides innovative ideas for the researchers, academics, professionals and high level engineering students to bring together the current developments in the applications of medicine, robotics, and remote sensing by satellites. Pattern recognition is characterized by automated searches over a large number of observations and huge combinatorial spaces and is concerned with the automatic discovery of regularities in data through the use of computer algorithms.*

*i-manager's Journal on Pattern Recognition is presently in its 10<sup>th</sup> Year. The first issue was launched in 2014.*

*i-manager's Journal on Pattern Recognition is published by i-manager Publications, one of India's leading Academic Journal Publisher, publishing 38 Academic Journals in diverse fields of Engineering, Education, Management and Science.*

### **Why Publish with us**

*i-manager Publications currently publishes academic Journals in Education, Engineering, Scientific and Management streams. All of i-manager's Journals are supported by highly qualified Editorial Board members who help in presenting high quality content issue after issue. We follow stringent Double Blind Peer Review process to maintain the high quality of our Journals. Our Journals target both Indian as well as International researchers and serve as a medium for knowledge transfer between the developed and developing countries. The Journals have a good mix of International and Indian academic contributions, with the peer-review committee set up with International Educators.*

### **Submission Procedure**

*Researchers and practitioners are invited to submit an abstract (200 words) / Full paper on or before the stipulated deadline, along with a one page proposal, including Title of the paper, author name, job title, organization/institution and biographical note.*

*Authors of accepted proposals will be notified about the status of their proposals before the stipulated deadline. All submitted articles in full text are expected to be submitted before the stipulated deadline, along with an acknowledgement stating that it is an original contribution.*

### **Review Procedure**

*All submissions will undergo an abstract review and a double blind review on the full papers. The abstracts would be reviewed initially and the acceptance and rejection of the abstracts would be notified to the corresponding authors. Once the authors submit the full papers in accordance to the suggestions in the abstract review report, the papers would be forwarded for final review. The final selection of the papers would be based on the report of the review panel members.*

### **Format for Citing Papers**

*Author surname, initials (s.) (2023). Title of paper. i-manager's Journal on Pattern Recognition, 10(1), xx-xx.*

### **Copyright**

*Copyright © i-manager Publications 2023. All rights reserved. No part of this Journal may be reproduced in any form without permission in writing from the publisher.*

### **Contact e-mails**

*editor@imanagerjournals.com  
submissions@imanagerpublications.com*

## **i-manager's Journal on Pattern Recognition**

### **Editor-in-Chief**

**Dr. G. R. Sinha**

Professor,  
Myanmar Institute of Information Technology,  
Mandalay, Myanmar.

### **EDITORIAL COMMITTEE**

**Dr. Kyawt Kyawt Htay**

Assistant Lecturer,  
Myanmar Institute of Information  
Technology,  
Mandalay,  
Myanmar.

**Dr. N. Santhiya kumari**

Associate Professor and Head,  
Department of Electronics and  
Communication Engineering,  
Knowledge Institute of Technology,  
Salem, India.

**Dr. H. S. Sheshadri**

Professor,  
Department of Electronics and  
Communication Engineering,  
PES College of Engineering, Mandya,  
Karnataka,  
India.

**Dr. C. Shoba Bindu**

Professor,  
Department of Computer Science  
Engineering,  
Jawaharlal Nehru Technological  
University, Anantapur,  
Andhra Pradesh, India.

**Dr. Smita Selot**

Professor and Head,  
Department of Computer Applications,  
Shri Shankaracharya Technical Campus,  
Chattishgarh,  
India.

**Ms. Rita Dewanjee**

Head of Department,  
Faculty & Staff of MATS School of  
Information Technology, MATS University,  
Chhattisgarh,  
India.

### *Abstracting / Indexing*



## i-manager's Journal on Pattern Recognition

### OUR TEAM

#### Publisher

Joe Winston

Renisha Winston

Editorial Director

Dr. Jeya Shobana S.

Research Head

Dr. Sybi Cynthia

Editorial Head

Jeba Shiny S.

Issue Editor

Ramya R.

Corresponding Editor

Anitha Bennet

Business Head

C. A. Jeffrin Christo

Operations Manager

M. Sajintha

Issue Design

J. S. Joy Robinson

Production Manager

### OUR OFFICES

#### Registered Office

3/343, Hill view,  
Town Railway Nager,  
Nagercoil, Kanyakumari District - 629001  
Ph : 91-4652- 277675  
E-mail : [info@imanagerpublications.com](mailto:info@imanagerpublications.com)

#### Editorial Office

13-B, Popular Building,  
Mead Street, College Road,  
Nagercoil, Kanyakumari District - 629001  
Ph : (91-4652) 231675, 232675, 276675  
E-mail : [editor@imanagerjournals.com](mailto:editor@imanagerjournals.com)

### Join with us



<https://www.facebook.com/imanagerJPR/>



<https://www.facebook.com/imanagerPublishing/>



<https://twitter.com/imanagerpub>

# CONTENTS

## RESEARCH PAPERS

---

- |    |   |
|----|---|
| 1  | DEVELOPMENT OF ANOMALY DETECTOR FOR MOTOR BEARING CONDITION MONITORING USING FAST FOURIER TRANSFORM (FFT) AND LONG SHORT TERM MEMORY (LSTM)-AUTOENCODER<br>By Sulaimon A. Bashir, Oladebo Suliat Jimoh, Idris Mohammed Kolo, Enesi Femi Aminu |
| 16 | NAVIGATION OF EMPTY PARKING SLOTS USING IMAGE PROCESSING<br>By Srujan Panthula, Yaramala John Wesley, Alle Saiteja Wesley, Sudha Mishra, Agnivesh Pandey  |
| 25 | TRACKING MISSING PERSONS USING FACIAL RECOGNITION<br>By Bhavika Gupta, Palak Agarwal, Disha Devalia   |
| 34 | PERFORMANCE ANALYSIS OF PHOTONICS-BASED RF TRANSCEIVER FOR HIGH-SPEED DATA TRANSMISSION<br>By Mayank Kaushik, J. S. Sidana, Shivendra Kumar, P. M. Menghal  |
| 42 | MUSIC RECOMMENDER SYSTEM USING FACIAL FEATURE RECOGNITION<br>By Harsh Mishra, Deep Kothari, Prajwal Jaiswal, Sheetal Rath   |

# EDITORIAL

*i-manager's Journal on Pattern Recognition (JPR), (January – June 2023: Volume – 10, Issue - 1) has five peer reviewed research papers that captivate distinct themes on many characteristics associated with Pattern Recognition.*

*Sulaimon proposed a novel approach for anomaly detection using Fast Fourier Transform (FFT) and Long Short-Term Memory (LSTM)-Autoencoder (AE). A data processing approach based on FFT was developed to pre-process the raw sensor data. This helped to reduce noise and improve the Signal-to-Noise Ratio (SNR). Additionally, an anomaly detection model based on LSTM-Autoencoder was developed and trained on the pre-processed data. The proposed approach was able to detect anomalies at a low threshold and achieved a high accuracy score.*

*Srujan et al. proposed a method to solve the major problem of traffic congestion at parking places. To solve this issue, the navigation of empty parking slots is used. In this research, live information is displayed before drivers enter the parking area. With the help of the display and website, drivers can save a significant amount of time finding an empty parking slot to park their vehicle. The system is based on camera sensors and software, where image processing is used to process the image.*

*Bhavika et al. proposed a system that uses an integrated and centralized database of missing persons using Aadhar card details. The facial recognition system used for this approach is the Deepface which improves the efficiency of locating missing individuals. The system matches the facial features with the data stored in Aadhar cards, providing a reliable means of identification. The integration of technology and data provides quick and more accurate identification, increasing the chances of locating missing persons promptly.*

*Mayank et al. proposed a novel Wavelength-Division Multiplexing (WDM)-based photonics link for a radar demonstrator. The proposed system uses a single pulsed laser as a source to design a transceiver for high-speed data transmission and reception which can generate tuneable radar signals and their echoes, avoiding the up/down conversion of radio frequency and ensuring both high resolution and simulation-based operation. The system is scalable and can be easily adapted to different radar applications.*

*Harsh et al. proposed a system which is capable of identifying the user's emotions and moods and suggesting a list of appropriate songs to improve their mood. Incorporating mood detection offers significant benefits to customers' mental health and personal satisfaction. The objective of this system is to capture the user's image, detect their mood and create a music player system that recommends music based on the user's real-time mood, utilizing a web camera and deep learning algorithms.*

*We thank the authors for their research and the reviewers for their efforts in reviewing the papers and bringing this issue to fruition. Our special thanks to the Editor-in-Chief Dr. G. R. Sinha for his constant support and efforts in further enhancing the quality of the Journal.*

*Hope this issue imparts an enlightening reading experience! Enjoy Reading!*

*Warm regards,*

*Dr. Jeya Shobana S.  
Research Head  
i-manager Publications*

## DEVELOPMENT OF ANOMALY DETECTOR FOR MOTOR BEARING CONDITION MONITORING USING FAST FOURIER TRANSFORM (FFT) AND LONG SHORT TERM MEMORY (LSTM)-AUTOENCODER

By

SULAIMON A. BASHIR \*

OLADEBO SULIAT JIMOH \*\*

IDRIS MOHAMMED KOLO \*\*\*

ENESI FEMI AMINU \*\*\*\*

\*-\*\*\*\* Department of Computer Science, School of Information and Communication Technology, Federal University of Technology, Minna, Nigeria.

Date Received: 16/05/2023

Date Revised: 25/05/2023

Date Accepted: 05/06/2023

### ABSTRACT

Anomaly detection in motor bearings is a critical task for preventing downtime and ensuring efficient operation. This paper proposes a novel approach for anomaly detection using Fast Fourier Transform (FFT) and Long Short-Term Memory (LSTM)-Autoencoder (AE). A data processing approach based on FFT was developed to pre-process the raw sensor data. This helped to reduce noise and improve the Signal-to-Noise Ratio (SNR). Additionally, an anomaly detection model based on LSTM-Autoencoder was developed and trained on the pre-processed data. The proposed approach was able to detect anomalies at a low threshold and achieved a high accuracy score.

Keywords: Motor Bearing, Anomaly Detection, Deep Learning, Fast Fourier Transform, Long Short Term Memory, Autoencoder.

### INTRODUCTION

In recent times, Artificial Intelligence (AI) has been employed in the detection of anomalies in machines. Some techniques used include the Support Vector Machine (SVM) (Lee et al., 2020), Bayesian Classifier, Neural Networks, and Deep Learning (DL) methods (Jiang et al., 2019). However, the imbalanced nature of the dataset generated from sensors poses challenges in accurate anomaly detection.

According to Chalapathy and Chawla (2019), Artificial Intelligence (AI), including Machine Learning and Deep Learning, is preferred over traditional anomaly detection methods because it can detect anomalies using data that may not be available in traditional methods. This allows for earlier detection of anomalies, which can help

to prevent breakdowns.

Among AI approaches, Deep Learning (DL) has been shown to outperform traditional Machine Learning because of its flexibility in learning and representing data in a hierarchical form in a Neural Network. This allows for better performance in anomaly detection.

As a result, Deep Learning has been used in a variety of applications, including, financial fraud detection (Adewumi & Akinyelu, 2017), Intrusion Detection in cyberspace (Kwon et al., 2019), Big data anomaly detection in the Internet of Things (IoT) (Mohammadi et al., 2018), and Medical Image Analysis (Litjens et al., 2017).

According to Lee et al. (2018a, 2018b), the use of deep learning for anomaly detection has been increasing because of its ability to receive signal data and automatically pre-process it before extracting the features needed for anomaly detection.

Generally, deep learning techniques for anomaly detection require huge amounts of data. It is important to



This paper has objectives related to SDGs



note that the data generated from sensors is often characterized by electrical noise and a low Signal-to-Noise ratio (SNR). This noise can make it difficult for deep learning models to accurately detect anomalies (Ahmad et al., 2020).

Recently, a lot of researchers have done extensive work on anomaly detection. Researchers have observed that large thresholds are often used in anomaly detection systems that use sensor data. This is because sensor data is often characterized by noise and low Signal-to-Noise Ratio (SNR), regardless of the technique used. This low SNR can often lead to false alerts or no alerts of anomalies, and the inability of deep learning models to detect weak signals.

The generation of weak signals by some sensors can make it difficult to detect anomalies. This is because the threshold for detecting anomalies is often set based on the error margin of normalcy. If the threshold is set too high, it may not be possible to detect weak signals, even if they are outside the range of normalcy.

Another challenge in detecting anomalies is the use of time-stamped data. Time-stamped data can be large and difficult to process. This can make it difficult to identify anomalies in a timely manner.

To improve the detection of anomalies, it is important to clean up the data using Digital Signal Processing techniques such as Fast Fourier Transform (FFT) (Jabczyński & Szczesniak, 1995; Vo et al., 2017). This will allow anomalies to be detected at lower thresholds, which indicates improved sensitivity. FFT is a robust technique for data characterized by low Signal-to-Noise Ratio (SNR), and it helps to convert sequence data from the time domain to the frequency domain. This makes feature extraction easier, even for data with low SNR. Additionally, using FFT can aid anomaly detection faster by reducing computational time and acting as a filter to clean up the data.

Therefore, this research proposes the use of Fast Fourier Transform and LSTM-Autoencoder for the faster, more sensitive, and more accurate detection of anomalies in motor bearings.

## 1. Related Work

Khadersab and Shivakumar (2018) presented how FFT and Inverse Fast Fourier Transform (IFFT) were used to analyze bearing failure. They compared healthy vibration data and faulty vibration data using these techniques to accurately assess the bearing failure. In their experiment, a piezoelectric accelerometer sensor was used to collect vibration data. The acquired data was connected to an FFT algorithm called EL-Calc. The FFT signal was then used to generate IFFT and a spectrogram. At the end of the experiment, defects were effectively identified.

According to Sulka et al. (2019), the causes of unwanted vibration and the extent of a fault can be estimated. They used FFT and Short-Time Fourier Transform (STFT) to identify defects in bearings. The difference between these two techniques is that FFT does not use a time window, while STFT uses a time window to analyze a particular vibration signal. After using both techniques, it was observed that FFT shows a larger amplitude in the vibration signal that was analyzed. The amplitude also depends on the severity of the damage.

Abouelanouar et al. (2018) used Wavelet Transform (WT) to detect faults in gears and bearings. They found that WT was a powerful tool for fault detection, even with non-periodic vibrational signals (Hruntovich et al., 2019).

Kanwal et al. (2019) presented a method for detecting tampering in images that used a combination of the Fast Fourier Transform (FFT), local texture descriptors, and a Support Vector Machine (SVM) classifier. The results of their research showed that this method was able to increase the accuracy of tampering detection.

Barot and Kulkarni (2021) reviewed different techniques for detecting anomalies in machines. The research emphasized the need for Digital Signal Processing (DSP) techniques, such as the Fast Fourier Transform (FFT), Wavelet Transform (WT), Discrete Wavelet Transform (DWT), and Wavelet Packet Transform (WPT). They also stressed the importance of denoising and extracting weak signals, which can be done using wavelet-based signal denoising. It also reviewed Artificial Intelligence (AI) methods for diagnosing faults in bearings.

To reduce downtime in industries, Jin et al. (2019) proposed a data-driven approach for bearing prognostics based on the Kolmogorov-Smirnov test, self-organizing map, and unscented Kalman filter, as shown in Figure 1. The first step in their approach is to detect bearing degradation, also known as anomaly detection, by learning from historical data generated by vibration sensors. The second step is to predict the Remaining Useful Life (RUL) of the bearing using a degradation model and an unscented Kalman filter.

Zhang et al. (2019) emphasized the use of the Internet of Things (IoT) for data gathering for detecting anomalies in bearings. This research work aims to ensure adequate maintenance before a complete machine breakdown, exposes different techniques used to track the degradation of the system. These techniques involve the use of signal processing for better performance in degradation tracking. Cyclic Spectral Correlation (CSC) and Cyclic Spectral Coherence (CSCoh) have been proven to be powerful tools for signal processing. Due to the difficulty of obtaining labeled data from the experimental setup in Figure 2, fault detection was performed using semi-supervised learning and Support Data Description (SDD).

Wang et al. (2019) presented a case study that utilized Conditional Based Maintenance (CBM) for centrifugal pumps as part of a safety program for critical water systems used to mitigate fire hazards. This was done to monitor the operational condition of centrifugal pumps. Vibration data was gathered to investigate if CBM could identify different faults in the centrifugal pump.

Figure 3 shows the workflow of the research, which involved determining the best practices or approaches

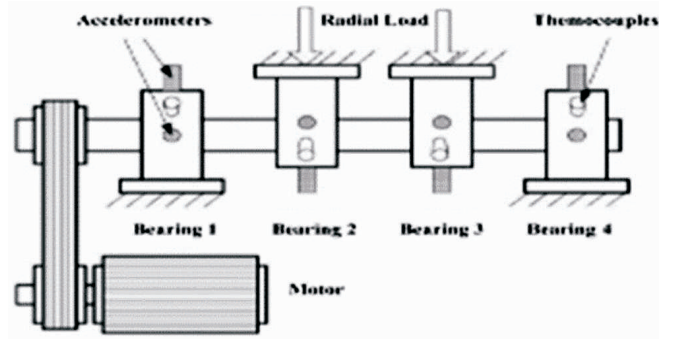


Figure 2. Experimental Set Up for Data Gathering

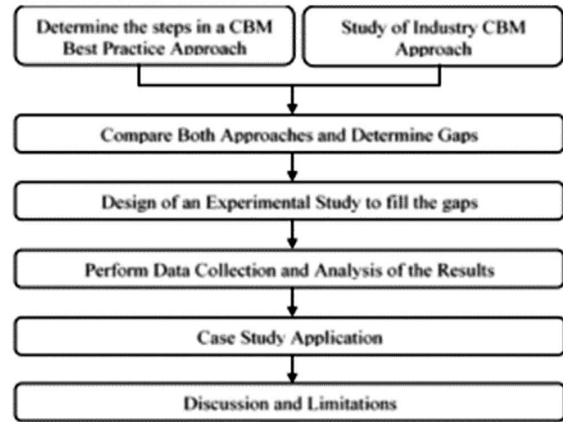


Figure 3. Work Flow of the Research

for CBM in the lab compared to the CBM approach in industry.

The advent of smart meters and the Internet of Things (IoT) has improved the gathering of data, which is instrumental to the detection of anomalies in industrial machines. According to Pittino et al. (2020), Machine Learning has been used to extract information from datasets, the important aspect of anomaly detection is that it aids in the detection of bearing faults. This helps to detect malfunctioning bearings before the machines completely break down.

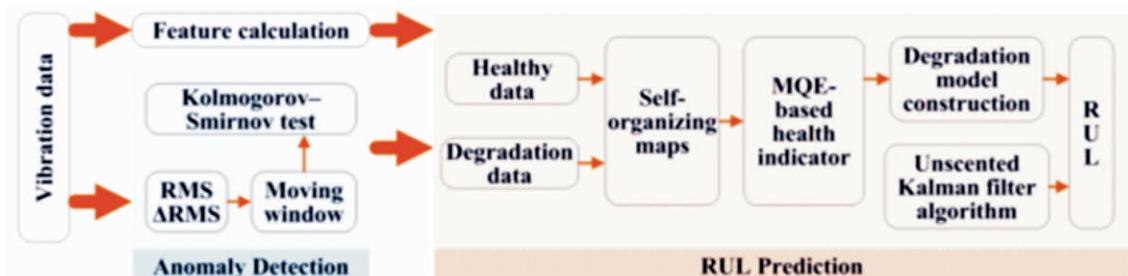


Figure 1. Data Driven Approach for Bearing Prognostic

Saeki et al. (2019) used visualization techniques, specifically Convolutional Neural Networks (CNNs), for the detection of motor bearing anomalies. In this work, a vibration dataset was captured and predictions were made to state the health status of the bearings. These predictions were compared with the analysis done by experts. Results showed that the technique is helpful in providing useful information about the health status of motor bearings.

Schmidt and Heyns (2019) state that anomaly detection can be used to identify localized faults in gears, even when historical fault data is not available. Discrepancy Analysis, Continuous Wavelet Transform (CWT), and Principal Component Analysis (PCA) are used to determine the divergence of the gear. Bayesian Data Analysis is then used to infer the presence of a localized anomaly. By identifying localized faults early, preventive maintenance can be performed to extend the life of the machine and avoid costly downtime.

Asakura et al. (2020) used Mahalanobis–Taguchi System (MTS), a method for quantifying the extent of damage in logical systems, to detect anomalies in those systems. They applied the technique to a large-scale vertical transfer system and developed calculations to achieve proper values for the technique based on simple excitation using a shaker.

Cooper et al. (2020) worked on anomaly detection in milling tools using Generative Adversarial Networks (GAN) and acoustic data. It uses K-Nearest Neighbors (KNN) to detect anomalies in machineries.

Nath (2020) disclosed that abrupt changes in sequential data are a major problem in anomaly detection. To solve this problem, low latency anomaly detection based on Quick Change Detection (QCD) is needed for effective detection. QCD minimizes the delay in detecting anomalies observed in sequential data. This is useful because in most models, the post-change distribution model may not be available. However, QCD has been used for bearing fault detection in turbines.

Sohaib and Kim (2018) proposed a method for detecting faults in bearings using bi-spectrum analysis and

Convolutional Neural Networks (CNNs). First, the bi-spectrum of the vibrational signal was extracted. Next, a CNN based on a stochastic optimization function was used to extract the interclass of the bi-spectra. The proposed method was able to detect faults more accurately than the previous work.

Pandarakone et al. (2018) proposed an online bearing fault detection method based on deep learning. Fast Fourier transform (FFT) was used to perform spectral analysis on data generated from the load current of the stator coil. This helped to extract features from the data, which were then used to train a convolutional neural network (CNN). The CNN was able to classify different types of bearing faults, including single scratch (SS) and full scratch (FS). The average accuracy of the system was 88.17%. Abid et al. (2019) presented a technique to aid the detection of faults in bearings and the extent of the damage.

The technique Optimized Stationary Wavelet Packet Transform (OP-SWPT), is an advanced form of the Digital Signal Processing Technique Wavelet Transform. The authors used Fast Fourier Transform (FFT) to analyze the signal in the frequency domain, Short Time Fourier Transform (STFT) is used to identify the location of the extracted features, and Wavelet Transform (WT) for non-stationary signals. The results showed that the technique was effective in detecting different types of bearing faults. It was also reported that the technique detected faults faster than other methods.

Egaji et al. (2020) emphasized the need to use available data, usually vibration data, to avoid downtime in industries. It suggests that Digital Signal Processing methods such as FFT could be used to improve the output of the data, which is often noisy due to sensor inefficiency.

In Boniol et al. (2020), features were extracted from the data. These features were then used to train a Neural Network. To make detection easier, Principal Component Analysis (PCA) was used to reduce the dimension of the data from 24 to 1-dimensional space. The output from PCA was then used as input to a regression model, which reconstructs the input. The error between the input and the

reconstructed output reveals anomaly detection. The regression models used in this research were Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN).

Industrial fans are used to aid cooling and ensure the proper functioning of industrial machines. Gong et al. (2018) presented an online solution to detect anomalies in industrial fans. To achieve this, they used acoustic signals with an intelligent prediction integrated system that is connected to the internet. Furthermore, Glowacz (2019) used Acoustic Signal Enhancement Filter and Adaptive Kalman Filter for feature extraction and detection.

Sohaib and Kim (2019) presented a method for detecting anomalies in machine bearings using complex envelope spectra and a Stacked Sparse Autoencoder-based Deep Neural Network (DNN) (Liu & Gryllias, 2020). To overcome the fluctuations in shaft speed, a fault diagnostic scheme was developed. The use of complex envelope spectra made the detection process easier.

The literature review shows that researchers have used either Digital Signal Processing (DSP) or Machine Learning (ML) to improve motor bearing detection (Vo et al., 2017, Wei et al., 2019). However, no studies have been found that use both the Fast Fourier transform (FFT) and Long Short-Term Memory (LSTM)-Autoencoders. Additionally, no other researchers have improved the performance of LSTM-Autoencoders for detecting motor bearing anomalies.

## 2. Methodology

The proposed technique in this research includes the use of National Aeronautics and Space Administration (NASA) data on motor bearings, the use of the Fast Fourier Transform (FFT) and a Long Short-Term Memory (LSTM)-Autoencoder to achieve anomaly detection. Python was used as a tool to write the code. The workflow of the proposed system is shown in Figure 4.

In this research, the data was collected, pre-processed with pandas, and then processed with FFT to aid anomaly detection. The structured data characterized by normalcy was then fed into the proposed model and

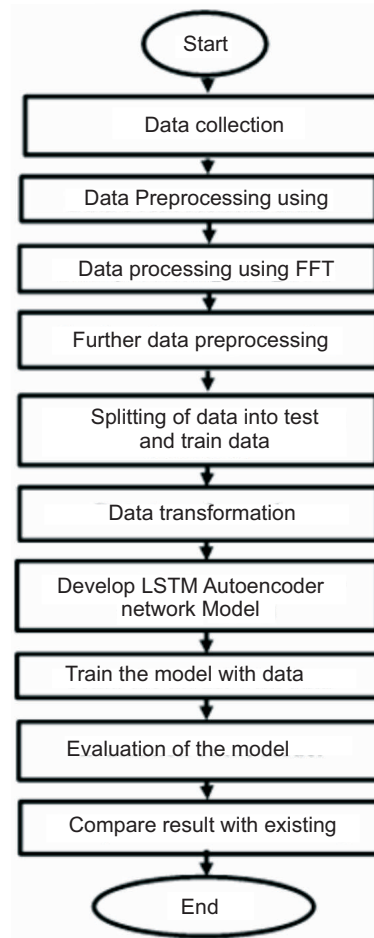


Figure 4. Proposed Model Diagram

trained. Afterwards, the proposed model was tested with abnormal data and the performance of the model was then evaluated using two metrics threshold, also known as anomaly score, and accuracy.

### 2.1 Data Collection

The data used in this study is sensor readings obtained from vibration sensors fastened to motor bearings by NASA Acoustics. These data were generated from four bearings that were run until failure under constant load. The data, which was taken at 10-minute intervals contains 20,800 data points per bearing read at a sampling rate of 20 kHz.

Figure 5 shows the part of the data collected in files named with time and date stamp. Figure 6 shows the unorganized data before processing.

### 2.2 Data Pre-Processing

The vibration data for the four motor bearings was

2004.02.12.10.52.39
2004.02.12.11.02.39
2004.02.12.11.12.39
2004.02.12.11.22.39
2004.02.12.11.32.39
2004.02.12.11.42.39

Figure 5. Raw Data Collected with Date and Time Stamp

1	0.112	-0.010	-0.005	0.022
2	0.049	-0.012	-0.046	0.020
3	-0.027	-0.017	-0.012	0.098
4	-0.110	0.029	0.112	0.056
5	0.007	-0.024	-0.044	0.054
6	0.122	0.051	-0.007	-0.012
7	0.134	0.007	0.034	-0.017
8	-0.015	-0.027	0.002	0.027
9	-0.168	-0.037	-0.107	0.042
10	-0.061	-0.090	0.020	0.046
11	0.012	-0.059	0.100	0.012
12	0.027	0.061	-0.068	-0.032
13	-0.056	0.002	-0.081	0.022
14	-0.115	-0.049	-0.044	0.056
15	0.015	0.042	-0.046	0.005
16	0.100	0.049	-0.125	-0.020
17	0.088	0.076	0.020	-0.015

Figure 6. Unorganized Data Before Processing

collected. The data was tagged with dates of collection and stored in separate files. Pandas was used to merge the four files into a single dataset as shown in Figure 7. The dataset was inspected to verify that there were no missing data. The data type and shape were checked to ensure that the data conformed to the format required by the deep learning algorithm.

### 2.3 Data Pre-Processing using FFT

Fast Fourier Transform (FFT) is a technique that employs the transformation of time sequence signals to its frequency domain so as to extract latent behavior of the signal source. In other words, the FFT decomposes an N-point time domain signal into N-time domain signals, each composed of a single point. Afterwards, the FFT algorithm calculates the N-frequency spectra corresponding to the N-time domain signals. Lastly, the N-spectra are then synthesized into a single frequency spectrum. This algorithm utilizes Equation 1.

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{i2\pi kn}{N}} \quad (1)$$

	Bearing 1	Bearing 2	Bearing 3	Bearing 4
2004-02-12 10:52:39	0.060236	0.074227	0.083926	0.044443
2004-02-12 11:02:39	0.061455	0.073844	0.084457	0.045081
2004-02-12 11:12:39	0.061361	0.075609	0.082837	0.045118
2004-02-12 11:22:39	0.061665	0.073279	0.084879	0.044172
2004-02-12 11:32:39	0.061944	0.074593	0.082626	0.044659

Figure 7. Dataset of the Motor Bearing After Pre-Processing

where N is the size of the domain, and  $X_k = X_0, X_1, \dots, X_{N-1}$  is converted to another sequence number  $x_n = x_0, x_1, \dots, x_{N-1}$  is the signal. In other words,  $x_n$  is a sinusoid with frequency of  $k/n$ , which is a cross correlation sequence of  $X_k$ . Furthermore, it must be noted that the 'N' points must be in the form of  $2^n$ . This means the N-time points must be within this range, otherwise it would not be able to capture the whole data. During the decomposition, the levels of decomposition is given as  $\log_2 N$ . For example, a 16-point signal, also represented as  $(2^4)$ , is broken down into x stages where  $x = \log_2 16$ . This results in a breakdown into 4 stages. During the breakdown, the original samples are re-ordered via bit reversal. Table 1 shows a sample data re-ordering.

The pseudocode for the Fast Fourier Transform is as follows,

```

//For general case let the input G of any case have a
sequence
//G=(a_0, a_1, ..., a_{N-1})
//Note that N is a power of 2. Also, we want to return
output values of H//given as:
// H = A(x) = \sum_{j=0}^{N-1} a_j x^j
//where H is a polynomial similar to Equation 1,
evaluated at Nth root of //unity and a is the coefficient
of the polynomial.
if N=1 then return {a_j}
if N>1 then //calling the Fourier transforms
recursively.
(s_0, s_1, ..., s_{N/2-1}) = FFT((a_0, a_2, ..., a_{N-2})w^2) // this deals
with even sequence
(s'_0, s'_1, ..., s'_{N/2-1}) = FFT((a_1, a_3, ..., a_{N-1})w^2) // this deals
with an odd sequence.
    
```

```

for j=0 to  $\frac{N}{2}-1$ 
     $r_j = s_j + w_N^j s'_j$ 
     $r_{j+\frac{N}{2}} = s_j - w_N^j s'_j$  // the negative sign is from
the odd.
// Note that w is the primitive Nth root of unity if
 $w^0, \dots, w^{N-1}$  // are root of the unity
return ( $r_0, r_1, \dots, r_{N-1}$ )
end for
end if
end if
// How long will it take to do this computation?
 $T(N) = 2T(N/2) + O(N)$ 
// N is the size of the problem and  $O(N)$  is the order of N
as solved in the // Equation 1.
// =  $O(N \log N)$ 
// which is much better if the system is to run for  $O(N^2)$ .

```

## 2.4 Further Data Pre-Processing

Further pre-processing step was done to normalize the data. The normalization process ensures that all data points are converted to numbers between 0 and 1. LSTM models typically use data in a three-dimensional tensor format, which is important to reshape the data from its current two-dimensional format (date-time stamp,

Normal Sample	Binary Representation	Decomposed Outcome	Binary Representation
0	0000	0	0000
1	0001	8	1000
2	0010	4	0100
3	0011	12	1100
4	0100	2	0010
5	0101	10	1010
6	0110	6	0110
7	0111	14	1110
8	1000	1	0001
9	1001	9	1001
10	1010	5	0101
11	1011	13	1101
12	1100	3	0011
13	1101	11	1011
14	1110	7	0111
15	1111	15	1111

Table 1. Data Re-Ordering Table

feature) to a three-dimensional format (date-sample, time-sample, features). Figure 8 shows the shape of the dataset that was originally collected.

## 2.5 Data Splitting

To use an Autoencoder network for anomaly detection, the data must be split into training and test data. There are two common approaches which are as follows,

- *Labeled Data:* The entire dataset can be characterized with both normal and abnormal data. It can then be split into training and test data using a traditional ratio, such as 60% training data and 40% test data. Some researchers use a 70/30 split. The ratio of the split is based on the researcher's judgement.
- *Unlabeled Data:* Only the normal data can be used to detect anomalies. This approach is used when the data is unlabeled. The faulty data is used as the test data.

In this research, the training dataset was obtained by plotting all the data to get the percentage of both normal and faulty signals.

## 2.6 Model Description

### 2.6.1 Autoencoder

Autoencoders are neural networks that learn to replicate its inputs at the output. It can be used for a variety of tasks, including dimensionality reduction, anomaly detection, and feature extraction.

In anomaly detection, autoencoders are used to identify data points that are significantly different from the rest of the data. This is done by comparing the output of the autoencoder to the input data. If the output is significantly different from the input, then the data point is likely to be an anomaly.

The anomaly score is calculated by measuring the difference between the output of an autoencoder and the input data. Data points with high anomaly scores are more likely to be anomalies.

Data points with anomaly scores that are higher than the

**Dataset shape: (982, 4)**

Figure 8. Shape of the Dataset Originally Collected in Two Dimension

threshold are flagged as anomalies.

## 2.6.2 Long Short-Term Memory Network (LSTM)

LSTM is a type of Recurrent Neural Network (RNN) that helps to retain long-term dependencies between data points in a sequence (Nguyen et al., 2020). This is achieved through the use of three control gates, such as the input gate, the output gate, and the forget gate. These gates are implemented using Sigmoid Neural Network layers and pointwise multiplication. LSTMs are typically implemented as a chain of repeated modules, each of which contains the three control gates.

LSTMs can read data sequentially as vector  $S = \{s_1, s_2, s_3, \dots, s_t, \dots\}$ , where  $S \in \mathbb{R}^k$  represents the vector reading of  $x$  with  $x$  dimensions at time  $t$ . This Neural Network consists of three gates that helps to store relevant information and forget irrelevant ones. These gates are the forget gate, input gate and output gate.

## 2.6.3 LSTM-Autoencoder

An autoencoder is a neural network that can be used to detect anomalies by compressing large amounts of data into a smaller vector space. However, using an Autoencoder alone can be challenging for time series data, as the data can become very large over time, and the computation required to train and use the model can be prohibitive.

To address this challenge, LSTM-Autoencoders can be used. LSTMs are a type of recurrent neural network that are well-suited for processing time series data. Autoencoders can learn to remember long-term dependencies in data, which allows it to identify the most important features to remember. This reduces the amount of data that needs to be processed, which can significantly improve the performance of the model.

The architecture of an LSTM-Autoencoder is shown in Figure 9. The input data, which is a time series data  $x_u, x_d$  is first fed into the autoencoder. The autoencoder compresses the data into a smaller vector  $z$ . This compressed data is then fed into the LSTM. The LSTM learns to remember the important features of the data and forget the irrelevant features. This reduces the amount of data that needs to be reconstructed, which

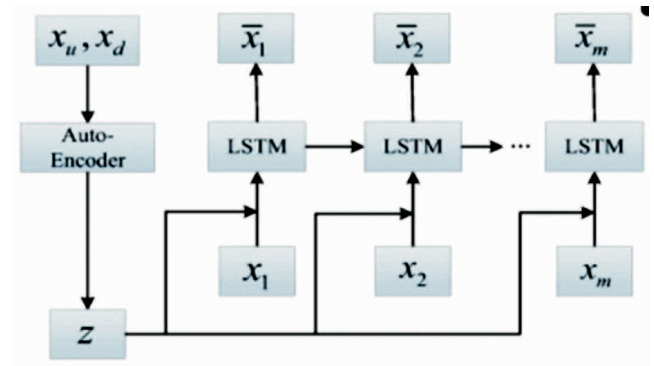


Figure 9. Architecture of LSTM-Autoencoder

further improves the performance of the model.

During reconstruction, a threshold is selected. If the error between the reconstructed data and the original data is above the threshold, an anomaly is detected.

$$L = \frac{1}{2} \sum_s \|s - s'\| \times \|s - s'\| \quad (2)$$

Equation 2 shows the LSTM-Autoencoder model. The pseudocode for the LSTM-Autoencoder is as follows,

*LSTM-Autoencoder Algorithm*

```

//INPUT: the training set Xu, //Xu is the dataset
//OUTPUT: prediction result X.
//Preprocess the data Xu to give Xuu
//Split Xuu to AutoEncoder training set XAE and XAD
validation dataset.
//Initialize the weight matrices of AutoEncoder
randomly.
//Put XAE into AutoEncoder.
if L(X, Y) < d then
    Calculate the error L(X, Y)
    Use the back propagation to train the AutoEncoder.
else
    End the training.
end if
//Generate the characteristics of the input dataset Zt.
for t = 0 to epoch do
    Put Zt into the LSTM, and do for forward propagation.
    Generate output
    Calculate error.
    
```

```

Use the back propagation to update parameters.
Use forward propagation to update network status ht,
end for
//Add LSTM after the encoder of AutoEncoder to form
AE-LSTM.
//Fine-tuning the whole network, training initialization
parameters.
//Input XAD test data in AE-LSTM to generate the
predicted value X.
Return X.

```

## 2.6.4 Development of LSTM-Autoencoder Network with FFT

The model used in this research is the LSTM-Autoencoder with FFT. The LSTM-Autoencoder was used to aid the detection of anomalies in time series data. The FFT technique was also used to extract latent features of a signal in the frequency spectrum. The pseudocode for the model algorithm is as follows,

```

Model Algorithm (LSTM-Autoencoder with FFT)
//INPUT: the training set Xu, // Xu is the dataset
//OUTPUT: prediction result X.
//Preprocess the data Xu to give Xuu
//process the data using FFT to give Xuu'
//Split Xuu' to AutoEncoder training set XAE and XAD
validation //dataset.
//Initialize the weight matrices of AutoEncoder
randomly.
//Put XAE' into AutoEncoder.
if L(X, Y) < d then
Calculate the error L(X, Y)
Use the back propagation to train the AutoEncoder.
else
End the training.
end if
//Generate the characteristics of the input dataset Zt.
for t = 0 to epoch do
Put Zt into the LSTM, and do forward propagation.
Generate output

```

```

Calculate error.
Use the back propagation to update parameters.
Use forward propagation to update network status ht,
//INPUT: the training set Xu, // Xu is the dataset
//OUTPUT: prediction result X.
//Preprocess the data Xu to give Xuu
//Process the data using FFT to give Xuu'
//Split Xuu' to AutoEncoder training set XAE and XAD
validation //dataset.
//Initialize the weight matrices of AutoEncoder
randomly.
//Put XAE' into AutoEncoder.
if L(X, Y) < d then
Calculate the error L(X, Y)
Use the back propagation training the AutoEncoder.
else
End the training.
end if
//Generate the characteristics of the input dataset Zt.
for t = 0 to epoch do
Put Zt into the LSTM, and do for forward propagation.
Generate output
Calculate error.
Use the back propagation to update parameters.
Use forward propagation to update network status ht,
end for
//Add LSTM after the encoder of AutoEncoder to form
AE-LSTM.
//Fine-tuning the whole network, training initialization
parameters.
//Input XAD' test data in AE-LSTM to generate the
predicted value X.
Return X.

```

## 2.6.5 Training and Testing

The model was trained with normal ball bearing data using 100 epochs and a batch size of 10. Afterwards, it was tested with data characterized by anomalies. The

training loss and loss distribution were computed and plotted to serve as a guide for selecting a threshold. After the threshold was selected, the bearing failure plot was presented.

### 3. Results and Discussion

The data was pre-processed, converted from the time domain to the frequency domain, and features were extracted. Then, an LSTM-Autoencoder network was built and the loss was computed. The threshold was computed from the loss distribution, and the anomaly was observed graphically.

#### 3.1 Data Description

The data used was obtained from the NASA repository. The data contains four files, each of which contains 20,800 data points read at a sampling rate of 20 kHz. Figure 10 shows the description of the dataset. The data set has 982 rows and 4 columns. The mean of the dataset is between 0.048 and 0.081. The standard deviation is within 0.009 to 0.040. The minimum value is 0.0007. The maximum value is 0.453.

#### 3.2 Data Pre-Processing

The first step in the pre-processing stage is to merge the dataset. The dataset is currently stored in multiple files, each of which is labeled with a date and time stamp. Merging the files combines them into a single file, as shown in Figure 11. This allows the data to be used in the proposed model.

After the files have been merged, it is observed that the data was captured with date and time stamps. This means that the data is a sequence of data points, which

	Bearing 1	Bearing 2	Bearing 3	Bearing 4
count	982.000000	982.000000	982.000000	982.000000
mean	0.080951	0.078543	0.081351	0.047830
std	0.040200	0.011789	0.011607	0.009549
min	0.001168	0.000767	0.000716	0.001699
25%	0.060773	0.074240	0.076829	0.043951
50%	0.062021	0.075206	0.078187	0.044524
75%	0.083277	0.077458	0.080575	0.048130
max	0.453335	0.161016	0.151299	0.119047

Figure 10. Description of the Dataset

can be used for LSTM-Autoencoders. LSTM-Autoencoders are a type of neural network that is specifically designed to learn from sequential data.

By merging the dataset and converting it into a sequence of data points, it can be easier for the LSTM-Autoencoder to learn from the data. This will improve the performance of the model and make it more accurate.

The data was divided into training data and test data. The training data is the data of the normal working bearing. The split was done so that the validation dataset would capture both normal and anomalous data, allowing anyone to observe the point at which the anomaly occurred. While the training data is mainly normal data, it is not possible to observe the normal data and the point of abnormality from the raw data. Additionally, since the data may be characterized by noise, Fast Fourier Transform (FFT) was employed to clean up the data and to observe the point suspected as anomalous.

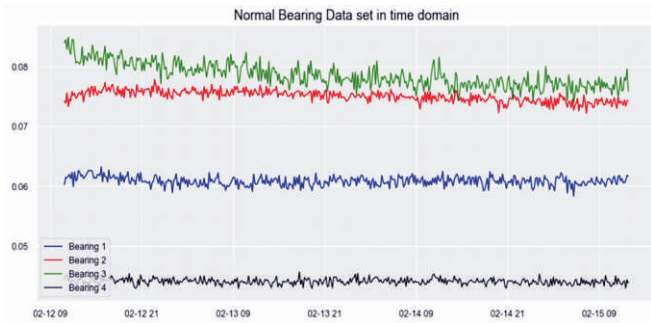
#### 3.3 Data Transformation using FFT

Figure 12(a) shows a graphical representation of the normal data (train data set) in the time domain. The time domain is a representation of the data as a function of time. In this case, the data is a series of numbers that represent the values of the signal at different points in time. Figure 12(a) shows that the data is relatively random, with no clear patterns. This makes it difficult to identify any anomalies in the data.

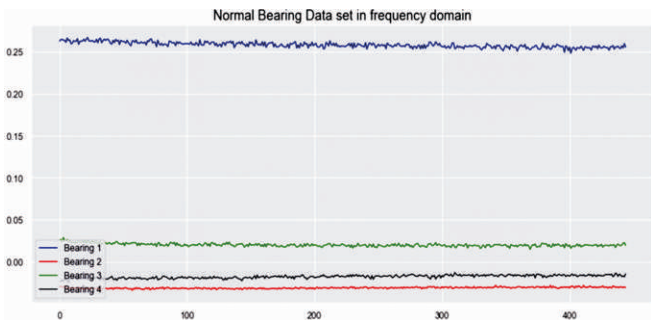
Figure 12(b) shows a graphical representation of the train data set in the frequency domain. The frequency domain is a representation of the data as a function of frequency. In this case, the data is a series of numbers that represent the amplitudes of the different frequencies that are present in the signal. Figure 12(b) shows that the data is

	Bearing 1	Bearing 2	Bearing 3	Bearing 4
2004-02-12 10:52:39	0.060236	0.074227	0.083926	0.044443
2004-02-12 11:02:39	0.061455	0.073844	0.084457	0.045081
2004-02-12 11:12:39	0.061361	0.075609	0.082837	0.045118
2004-02-12 11:22:39	0.061665	0.073279	0.084879	0.044172
2004-02-12 11:32:39	0.061944	0.074593	0.082626	0.044659

Figure 11. Data Representation of the Four Bearings Merged into One File



(a)



(b)

Figure 12. Graphical Representation of the Normal Data (a) in Time Domain (b) in Frequency Domain

less random in the frequency domain, with a few clear peaks. These peaks correspond to the frequencies that are most present in the signal. This makes it easier to identify any anomalies in the data, as they will stand out from the rest of the data.

Figures 13(a) and 13(b) show the graphical representation of the abnormal data in both time domain and frequency domain.

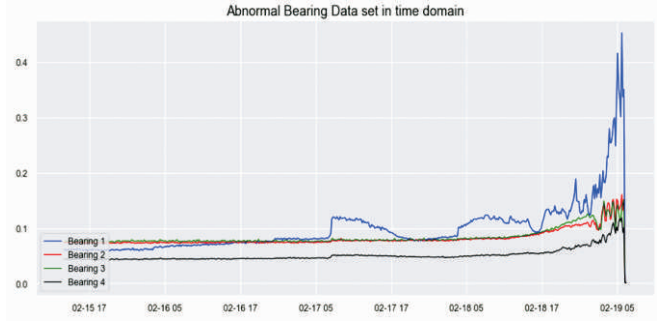
### 3.4 Further Pre-Processing

The LSTM-Autoencoder requires three-dimensional data as input. Therefore, the two-dimensional data shown in Figure 8 needs to be converted into three dimensions. This was done in time domain by repeating each data point three times, as shown in Figure 14.

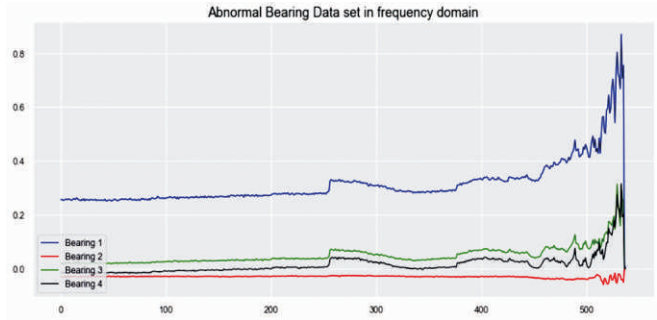
### 3.5 LSTM-Autoencoder Network Used

After the data was reshaped as required by the LSTM-Autoencoder, the model was built and the data was fed into it. The summary of the LSTM-Autoencoder network is shown in Figure 15.

The LSTM network has 200 nodes in the first layer, 25 nodes in the second layer, and 200 nodes in the third layer. This



(a)



(b)

Figure 13. Graphical Representation of the Abnormal Data (a) in Time Domain (b) Frequency Domain

Training data shape: (445, 1, 4)  
Test data shape: (538, 1, 4)

Figure 14. Shape of the Data Converted into Three-Dimension

```
Model: "model_3"
```

Layer (type)	Output Shape	Param #
input_3 (InputLayer)	(None, 1, 4)	0
lstm_11 (LSTM)	(None, 1, 200)	164000
lstm_12 (LSTM)	(None, 25)	22600
repeat_vector_3 (RepeatVecto	(None, 1, 25)	0
lstm_13 (LSTM)	(None, 1, 25)	5100
lstm_14 (LSTM)	(None, 1, 200)	180800
time_distributed_3 (TimeDist	(None, 1, 4)	804
-----		
Total params:		373,304
Trainable params:		373,304
Non-trainable params:		0

Figure 15. The Summary of the LSTM-Autoencoder Model configuration was chosen to achieve a lower threshold.

### 3.6 Loss Model

To evaluate the performance of the model, the loss is calculated. This was done when the network was trained for 100 epochs. Figure 16 represents the loss model and shows how much the training data differs from the validation LSTM data.

### 3.7 Loss Distribution

This parameter is needed to calculate the threshold. Figure 17 shows that the threshold can be below 0.126.

### 3.8 Anomaly Detection

Figure 18 shows that the model was able to detect anomalies at the selected threshold. The red line represents the threshold, and any signal above the red line is considered an anomaly.

### 3.9 Comparison of Results

In this paper, a new method for anomaly detection using FFT and LSTM-Autoencoder was proposed. The proposed method was compared to existing methods, such as Virtual Spectrum Imaging (VSI), Artificial Neural Network

(ANN), and Stacked Denoising Autoencoder. The results showed that the proposed method was able to detect weak signals more accurately than the existing methods. Additionally, the proposed method was able to achieve a higher accuracy than the existing methods.

The researchers analyzed each dataset (RM2, RM3, RM4, and RM5) separately and observed different anomaly scores. For example, the anomaly score for RM2 was significantly higher than the anomaly scores for RM3, RM4, and RM5. This suggests that the proposed method is more sensitive to anomalies in some datasets than the others. Figures 19 and 20 show the anomaly and accuracy scores of the proposed system.

### Conclusion and Recommendations

This research focuses on detecting anomalies in motor bearings. Although several research studies have been conducted to find techniques that would better assist in anomaly detection, the goal of this research is to detect anomalies in weak signals. To achieve this, the Long Short-Term Memory (LSTM)-Autoencoder (AE) with Fast Fourier Transform (FFT) was used. FFT was used to clean up and transform the data to reduce complex multiplications, speed up computation, and identify if an anomaly exists,

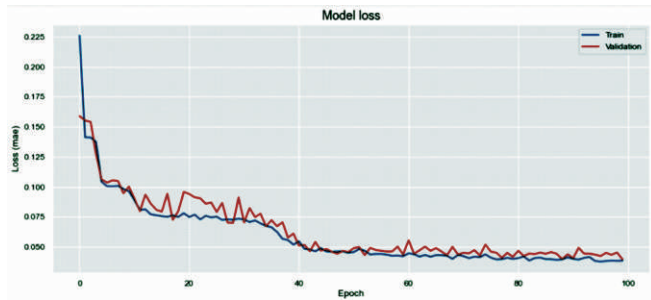


Figure 16. Loss Model

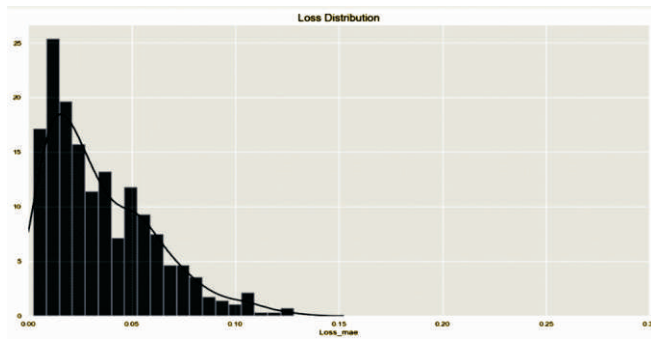


Figure 17. Loss Distribution

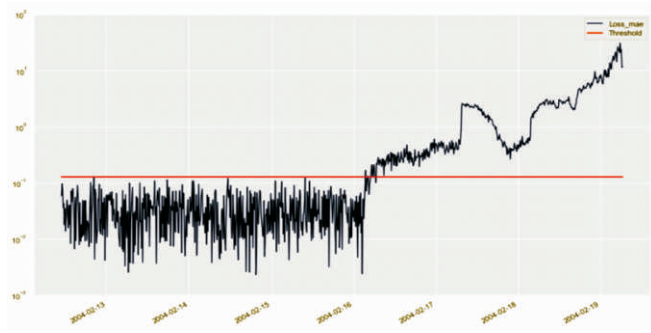


Figure 18. Anomaly Detection

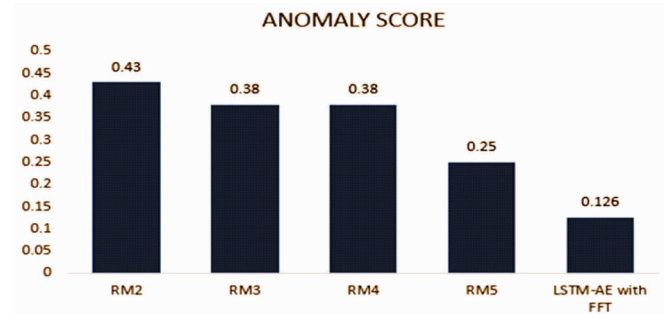


Figure 19. Anomaly Score of Bearing Data

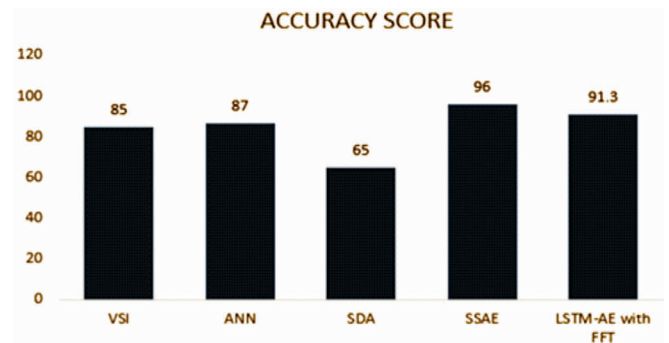


Figure 20. Accuracy Score

while the LSTM-Autoencoder was used to validate and detect the anomaly before it occurs. As a result of using Fast Fourier Transform (FFT), the LSTM-Autoencoder technique was able to detect anomalies at a low threshold of 0.126, compared to all other techniques that have been used. The research has been able to improve the sensitivity of detecting anomalies in motor bearings. Better sensitivity was achieved with the use of FFT and LSTM-Autoencoder. Sparse Stack Autoencoder with FFT can improve early detection of anomalies by lowering the threshold and improving accuracy.

## Future Work

In the future, in addition to detection anomaly localization can be achieved through the use of Machine Learning and statistical models. This will help to expedite the maintenance process.

## References

- [1]. Abid, F. B., Sallem, M., & Braham, A. (2019, March). Optimized SWPT and decision tree for incipient bearing fault diagnosis. In *2019 19<sup>th</sup> International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)* (pp. 231-236). IEEE. <https://doi.org/10.1109/STA.2019.8717197>
- [2]. Abouelanouar, B., Elamrani, M., Elkihel, B., & Delaunois, F. (2018). Application of wavelet analysis and its interpretation in rotating machines monitoring and fault diagnosis - A review. *International Journal of Engineering & Technology*, 7(4), 3465-3471.
- [3]. Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-Inspired based credit card fraud detection techniques. *International Journal of System Assurance Engineering and Management*, 8, 937-953. <https://doi.org/10.1007/s13198-016-0551-y>
- [4]. Ahmad, S., Styp-Rekowski, K., Nedelkoski, S., & Kao, O. (2020, December). Autoencoder-based condition monitoring and anomaly detection method for rotating machines. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 4093-4102). IEEE. <https://doi.org/10.1109/BigData50022.2020.9378015>
- [5]. Asakura, T., Yashima, W., Suzuki, K., & Shimotou, M. (2020). Anomaly detection in a logistic operating system using the Mahalanobis–Taguchi method. *Applied Sciences*, 10(12), 4376. <https://doi.org/10.3390/app10124376>
- [6]. Barot, A., & Kulkarni, P. (2021). Technological evolution in the fault diagnosis of rotating machinery: A review. *IOSR Journal of Mechanical and Civil Engineering*, 18(2), 9-26.
- [7]. Boniol, P., Linardi, M., Roncallo, F., & Palpanas, T. (2020, April). Automated anomaly detection in large sequences. In *2020 IEEE 36<sup>th</sup> International Conference on Data Engineering (ICDE)* (pp. 1834-1837). IEEE. <https://doi.org/10.1109/ICDE48307.2020.00182>
- [8]. Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*. <https://doi.org/10.48550/arXiv.1901.03407>
- [9]. Cooper, C., Zhang, J., Gao, R. X., Wang, P., & Ragai, I. (2020). Anomaly detection in milling tools using acoustic signals and generative adversarial networks. *Procedia Manufacturing*, 48, 372-378. <https://doi.org/10.1016/j.promfg.2020.05.059>
- [10]. Egaji, O. A., Ekwevugbe, T., & Griffiths, M. (2020, July). A data mining based approach for electric motor anomaly detection applied on vibration data. In *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)* (pp. 330-334). IEEE. <https://doi.org/10.1109/WorldS450073.2020.9210318>
- [11]. Glowacz, A. (2019). Fault diagnosis of single-phase induction motor based on acoustic signals. *Mechanical Systems and Signal Processing*, 117, 65-80. <https://doi.org/10.1016/j.ymssp.2018.07.044>
- [12]. Gong, C. S. A., Lee, H. C., Chuang, Y. C., Li, T. H., Su, C. H. S., Huang, L. H., ... & Chang, C. H. (2018). Design and implementation of acoustic sensing system for online early fault detection in industrial fans. *Journal of Sensors*, 1-15. <https://doi.org/10.1155/2018/4105208>
- [13]. Hruntovich, N. V., Kapanski, A. A., Petrov, I. V., & Kostyleva, E. E. (2019). Vibration diagnostic of electric motor roller bearings. In *E3S Web of Conferences*, 124, 02008. EDP Sciences. <https://doi.org/10.1051/e3sconf/201912402008>
- [14]. Jabczyński, J., & Szczesniak, A. (1995). Digital

processing of Doppler signals using fast Fourier transform. *Optica Applicata*, 25(4), 281-289.

[15]. Jiang, W., Hong, Y., Zhou, B., He, X., & Cheng, C. (2019). A GAN-based anomaly detection approach for imbalanced industrial time series. *IEEE Access*, 7, 143608-143619. <https://doi.org/10.1109/ACCESS.2019.2944689>

[16]. Jin, X., Que, Z., Sun, Y., Guo, Y., & Qiao, W. (2019). A data-driven approach for bearing fault prognostics. *IEEE Transactions on Industry Applications*, 55(4), 3394-3401. <https://doi.org/10.1109/TIA.2019.2907666>

[17]. Kanwal, N., Girdhar, A., Kaur, L., & Bhullar, J. S. (2019, April). Detection of digital image forgery using fast Fourier transform and local features. In 2019 *International Conference on Automation, Computational and Technology Management (ICACTM)* (pp. 262-267). IEEE. <https://doi.org/10.1109/ICACTM.2019.8776709>

[18]. Khadersab, A., & Shivakumar, S. (2018). Vibration analysis techniques for rotating machinery and its effect on bearing faults. *Procedia Manufacturing*, 20, 247-252. <https://doi.org/10.1016/j.promfg.2018.02.036>

[19]. Kwon, D., Kim, H., Kim, J., Suh, S. C., Kim, I., & Kim, K. J. (2019). A survey of deep learning-based network anomaly detection. *Cluster Computing*, 22, 949-961. <https://doi.org/10.1007/s10586-017-1117-8>

[20]. Lee, G. Y., Kim, M., Quan, Y. J., Kim, M. S., Kim, T. J. Y., Yoon, H. S., ... & Ahn, S. H. (2018a). Machine health management in smart factory: A review. *Journal of Mechanical Science and Technology*, 32, 987-1009. <https://doi.org/10.1007/s12206-018-0201-1>

[21]. Lee, H., Li, G., Rai, A., & Chattopadhyay, A. (2020). Real-time anomaly detection framework using a support vector regression for the safety monitoring of commercial aircraft. *Advanced Engineering Informatics*, 44, 101071. <https://doi.org/10.1016/j.aei.2020.101071>

[22]. Lee, K., Kim, J. K., Kim, J., Hur, K., & Kim, H. (2018b, July). CNN and GRU combination scheme for bearing anomaly detection in rotating machinery health monitoring. In 2018 *1<sup>st</sup> IEEE International Conference on Knowledge Innovation and Invention (ICKII)* (pp. 102-105). IEEE. <https://doi.org/10.1109/ICKII.2018.8569155>

[23]. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A.,

Ciampi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>

[24]. Liu, C., & Gryllias, K. (2020). A semi-supervised Support Vector Data Description-based fault detection method for rolling element bearings based on cyclic spectral analysis. *Mechanical Systems and Signal Processing*, 140, 106682. <https://doi.org/10.1016/j.ymsp.2020.106682>

[25]. Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018). Deep learning for IoT big data and streaming analytics: A survey. *IEEE Communications Surveys & Tutorials*, 20(4), 2923-2960. <https://doi.org/10.1109/COMST.2018.2844341>

[26]. Nath, S. (2020). *Low Latency Anomaly Detection with Imperfect Models* (Doctoral dissertation, University of Arkansas).

[27]. Pandarakone, S. E., Masuko, M., Mizuno, Y., & Nakamura, H. (2018, September). Deep neural network based bearing fault diagnosis of induction motor using fast Fourier transform analysis. In 2018 *IEEE Energy Conversion Congress and Exposition (ECCE)* (pp. 3214-3221). IEEE. <https://doi.org/10.1109/ECCE.2018.8557651>

[28]. Pittino, F., Puggli, M., Moldaschl, T., & Hirschl, C. (2020). Automatic anomaly detection on in-production manufacturing machines using statistical learning methods. *Sensors*, 20(8), 2344. <https://doi.org/10.3390/s20082344>

[29]. Saeki, M., Ogata, J., Murakawa, M., & Ogawa, T. (2019, June). Visual explanation of neural network based rotation machinery anomaly detection system. In 2019 *IEEE International Conference on Prognostics and Health Management (ICPHM)* (pp. 1-4). IEEE. <https://doi.org/10.1109/ICPHM.2019.8819396>

[30]. Schmidt, S., & Heyns, P. S. (2019). Localised gear anomaly detection without historical data for reference density estimation. *Mechanical Systems and Signal Processing*, 121, 615-635. <https://doi.org/10.1016/j.ymsp.2018.11.051>

[31]. Sohaib, M., & Kim, J. M. (2018). Reliable fault

diagnosis of rotary machine bearings using a stacked sparse autoencoder-based deep neural network. *Shock and Vibration*, 2018, 1-11. <https://doi.org/10.1155/2018/2919637>

[32]. Sohaib, M., & Kim, J. M. (2019). Fault diagnosis of rotary machine bearings under inconsistent working conditions. *IEEE Transactions on Instrumentation and Measurement*, 69(6), 3334-3347. <https://doi.org/10.1109/TIM.2019.2933342>

[33]. Sulka, P., Sapietova, A., Dekys, V., & Stalmach, O. (2019). Vibration analysis and comparison of the damaged and undamaged rolling ball bearing. *Transportation Research Procedia*, 40, 511-518. <https://doi.org/10.1016/j.trpro.2019.07.074>

[34]. Vo, Q., Fang, F., Zhang, X., & Gao, H. (2017). Surface recovery algorithm in white light interferometry based on combined white light phase shifting and fast Fourier

transform algorithms. *Applied Optics*, 56(29), 8174-8185. <https://doi.org/10.1364/AO.56.008174>

[35]. Wang, Z., Zhou, Y., & Li, G. (2019, November). Anomaly detection for machinery by using Big Data Real-Time processing and clustering technique. In *Proceedings of the 3<sup>rd</sup> International Conference on Big Data Research* (pp. 30-36). <https://doi.org/10.1145/3372454.3372480>

[36]. Wei, W., Wu, H., & Ma, H. (2019). An autoencoder and LSTM-based traffic flow prediction method. *Sensors*, 19(13), 2946. <https://doi.org/10.3390/s19132946>

[37]. Zhang, S., Ye, F., Wang, B., & Habetler, T. G. (2019). Semi-supervised learning of bearing anomaly detection via deep variational autoencoders. *arXiv preprint arXiv:1912.01096*. <https://doi.org/10.48550/arXiv.1912.01096>

---

## ABOUT THE AUTHORS

Sulaimon A. Bashir, Department of Computer Science, School of Information and Communication Technology, Federal University of Technology, Minna, Nigeria.

Oladebo Sullat Jimoh, Department of Computer Science, School of Information and Communication Technology, Federal University of Technology, Minna, Nigeria.

Idris Mohammed Kolo, Department of Computer Science, School of Information and Communication Technology, Federal University of Technology, Minna, Nigeria.

Enesi Femi Aminu, Department of Computer Science, School of Information and Communication Technology, Federal University of Technology, Minna, Nigeria.

# NAVIGATION OF EMPTY PARKING SLOTS USING IMAGE PROCESSING

By

SRUJAN PANTHULA \*

YARAMALA JOHN WESLEY \*\*

ALLE SAITEJA WESLEY \*\*\*

SUDHA MISHRA \*\*\*\*

AGNIVESH PANDEY \*\*\*\*\*

\*-\*\*\*\*\* Department of Information Technology, School of Engineering & Technology, Guru Ghasidas Vishwavidyalaya (Central University), Bilaspur, India.

Date Received: 02/03/2023

Date Revised: 17/04/2023

Date Accepted: 24/04/2023

## ABSTRACT

Navigation of empty parking slots has been proposed to solve the major problem of traffic congestion at parking places, such as malls and theatres. In other words, it can also reduce time consumption and costs. When there is no information about free parking slots, drivers face the issue of spending time to find a suitable parking spot. In multi-story buildings or malls, it is even more difficult to find an available parking. To solve this issue, the navigation of empty parking slots is used. In many parking areas, there is only an automated entrance to enter the vehicle into the parking area, but there is no software that can display live video to consumers to show whether there is an available space. Additionally, there is no website or application that can show live footage of the parking area, indicating whether it is empty or occupied. In this research live information is displayed before drivers enter the parking area. With the help of the display and website, drivers can save a significant amount of time finding an empty parking slot to park their vehicle. The system is based on camera sensors and software, where image processing is used to process the image.

Keywords: Sensor, Parking, Slot, Open CV, Image Processing.

## INTRODUCTION

The main goal of this paper is to alleviate the stress of finding a free parking slot. Highly rushed areas like cinema halls often experience heavy traffic in the parking areas. Year by year, the population is increasing along with the number of cars purchased. For example, in large cities like Hyderabad and Mumbai, the number of cars are increasing rapidly, leading to a significant traffic. Research shows that cars are no longer a luxury item in daily life, and people often face the same issue of finding free parking slots while going to functions or malls. Nowadays, finding a free parking slot is a tough task. Our

system is a state-of-the-art approach that aims to assist drivers in efficiently locating vacant parking slots in busy parking areas. With the rapid expansion of urban areas and the increasing number of vehicles on the road, finding a convenient parking spot can be a daunting task (Alsaferi et al., 2018). This system utilizes the power of image processing to accurately detect and analyze parking slots in real-time, making the process of locating an empty spot quicker and more convenient. In this way, the system not only saves time for drivers but also optimizes the utilization of available parking spaces (Jeffrey et al., 2012). This technology holds a great promise for improving the overall parking experience and reducing the environmental impact of urban transportation.

## 1. Navigation of Empty Parking Slots

There have been numerous technological advancements



This paper has objectives related to SDGs



that have saved time and made people's lives easier. However, the parking management system still remains traditional, manual, and time-consuming. It requires human effort to find a vacant parking spot by searching the entire parking area. This is because early arrivals have the advantage of parking wherever they want without any restrictions, leading to a lack of discipline and order (Farooqi et al., 2019).

This has become a significant problem for end-users, who do not want to spend a lot of time searching for parking spots. Furthermore, with the rapid population growth and increasing use of private vehicles, the demand for parking spaces has skyrocketed, and a more organized parking system can provide a solution to this issue (Suhr & Jung, 2016). Even though governments are trying to improve transport systems and infrastructure, there is currently a shortage of available parking spaces in public places like stadiums, markets, hospitals, shopping malls, and airports. The number of cars on the road is expected to increase significantly from 841 million cars in 2008 to more than 1.6 billion cars in 2035, reflecting the rapid industrial growth worldwide (Thangam et al., 2018). The slow development of urban planning has worsened the problem of finding vacant parking spaces in modern cities, especially during peak periods and festivals (Abel et al., 2021; Gupta et al., 2020).

According to the 2022 survey, the number of cars in India has increased from 2012 to 2022, as shown in Figure 1. Based on this data, it can be inferred that the number of vehicles will continue to grow in the coming years.



Figure 1. Number of Cars Increased in India (2012- 2022)

However, the number of parking slots available at universities, malls, and other public places remains limited. This lack of information about parking areas makes it difficult for drivers to find an empty parking slot. This problem becomes more pronounced in multi-story buildings where attendants may not be present to guide drivers towards an empty slot. As a result, drivers may face difficulties finding a free parking slot, which can cause traffic jams or accidents in parking areas.

## 2. Literature Review

Alsafery et al. (2018) proposed a smart car parking system that will assist users in solving the issue of finding a parking space and minimize the time spent searching for the nearest available car park. In addition, it provides users with road traffic congestion status. Moreover, the proposed system collects the raw data locally and extracts features by applying data filtering and fusion techniques to reduce the transmitted data over the network. After that, the transformed data is sent to the cloud for processing and evaluation using machine learning algorithms.

Kaarthik et al. (2017) proposed an intelligent parking system through image processing. In this systematic approach, the image processing technique can be used to identify the free, empty parking area where we can park our vehicles. In the proposed process, the parking area can be marked with a specific number and a sensor, and with the help of these sensors, the empty space can be identified to park the vehicle. The image processing display consists of seven segments displayed in real time. In addition to the display, an audio system has been interfaced in order to provide oral information about the parking system.

Yusnita et al. (2012) proposed an intelligent system for parking space detection based on image processing techniques that capture and process the brown rounded image drawn at the parking lot and produce information on the empty car parking spaces. It will be displayed on the display unit, which consists of seven segments, in real time. The seven-segment display shows the number of current available parking lots in the parking area. This

proposed system has been developed on a software and hardware platform.

Abel et al. (2021) proposed to make the car parking process more efficient in terms of time and cost. The absence of information on the position of the parking lot makes the car driver takes a long time to find it. In multi-story parking lots, officers cannot constantly monitor the available parking conditions directly, so prospective parking users do not know the position of the open parking space. In addition, many parking lots use automatic door latches but have no parking space information display. Parking system automation can be based on hardware, software, or a combination of hardware and software.

Jeffrey et al. (2012) implemented an energy- and cost-efficient smart parking system for a multi-floor parking facility using WSN. The system monitors the availability of free parking slots and guides the vehicle to the nearest free slot. Additional information, such as the amount of time the vehicle has been parked, is also monitored for billing purposes, along with the status of each meter. Cost is minimized by keeping the number of sensors lower maintaining reliability.

### 3. Problem Definition

It is generally very difficult for people to search for available parking spaces. The process is time-consuming and involves a lot of manual work, resulting in high frustration and stress levels among drivers (Pham et al., 2015). This often leads to traffic congestion and a waste of parking spaces, as some vehicles may be parked in such a way that others cannot access the empty parking slots.

### 4. Methodology

Figure 2 gives a detailed overview of the methodology.

The proposed methodology aims to address this issue by providing drivers with real-time information about the availability of parking spaces. Before entering the parking area, drivers can view the entire parking area on a screen located at the entrance or through a web application that displays the status of parking slots. This makes it more convenient for drivers to avoid traffic congestion at the parking area, as the live parking area status will be

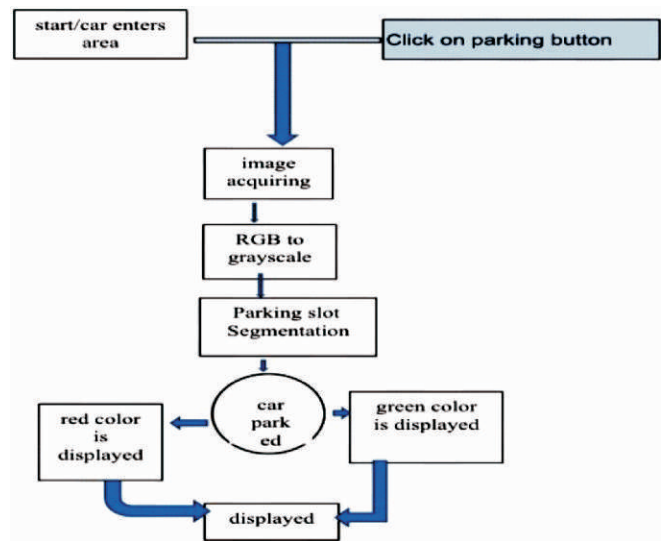


Figure 2. Overview of Existing Methodology

displayed, including the number of free slots available, allowing them to save time. An overview of the proposed system is shown in Figure 3.

The proposed methodology involves converting RGB images into grayscale images and then into binary images to detect free parking slots.

#### 4.1 System Architecture

The system architecture defines the entire flow of the parking system which includes various components. One of the critical components of the system is the image processing module that analyzes the images captured by the cameras installed in the parking slots (Rane et al., 2017). The module uses several image processing techniques such as contour detection, edge detection, and object recognition to detect and identify the vacant

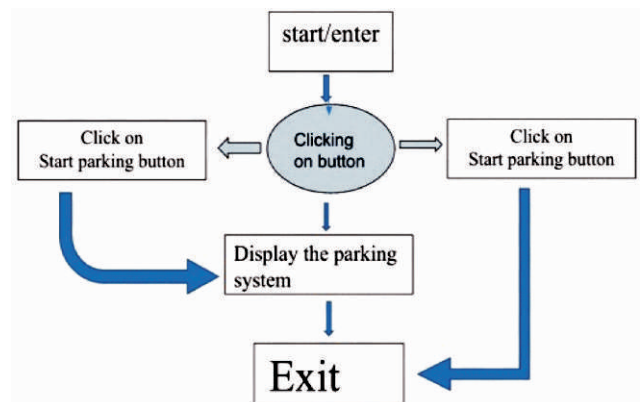


Figure 3. Overview of Proposed System

parking slots. To provide a visual representation of the parking area, the system has a user interface that displays a map of the parking lot with the location and status of each parking slot, i.e., whether it is occupied or empty. The user interface also provides various functionalities to interact with the system, such as selecting a parking slot and viewing its details. Figure 4 provides an overview of the major components of the system.

Moreover, the system includes a classifier module that uses machine learning algorithms to enhance the accuracy of parking slot detection. The classifier module is trained on a dataset of images of parking slots, both empty and occupied. The user interface displays the status of the classifier, indicating whether it is currently active or not. To ensure the system's accuracy, the parking slot status is periodically updated by re-analyzing the parking lot images. The user interface also displays the last update time to ensure system security and user privacy protection.

## 4.2 Identifying the Parking Slot

This research paper focuses on the identification of parking slots using computer vision techniques. The algorithm used involves extracting one frame from a recorded video to label the image where the slot is present, as shown in Figure 5.

The process of identifying a parking slot in a frame involves using various computer vision techniques, such as segmentation, edge detection, and object recognition. The image captured by a sensor, such as a camera, may need to be preprocessed to remove noise or adjust for lighting conditions. Figure 6 shows the process of identifying a parking slot in an image after preprocessing.

Accurate identification of parking slots is essential for the

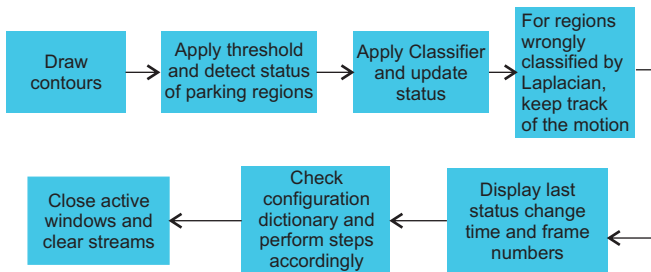


Figure 4. Overview of the Major Component



Figure 5. Frame Extracted from Sensor

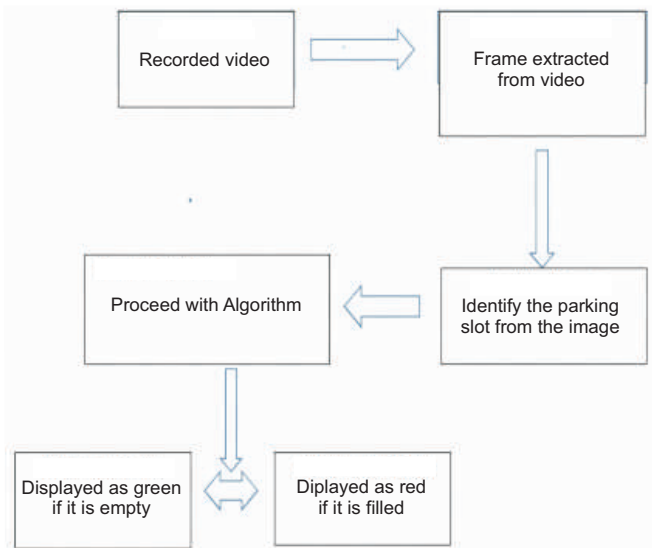


Figure 6. Identifying the Slot

system to provide reliable information about parking availability and occupancy. The system must differentiate between occupied and unoccupied parking slots and accurately track the status of each slot over time. To achieve this, the image processing algorithms used to identify the parking slot must be highly accurate.

## 5. System Module

The proposed image processing module is implemented in the Python language. This module consists of five phases that work together to make parking management easier. A block diagram can be used to represent the processing phases (Yusnita et al., 2012).

### 5.1 Initialization of the System

During the system initialization process, the procedure of drawing is performed (Kaarthik et al., 2017). A picture or

frame can be extracted using the Park slot member function to help identify a vacant parking space. The main purpose of this procedure is to determine a vacant parking slot without requiring any physical intervention. The drawn rectangle should be detectable, clear, and easy to recognize, and contain all the details about the parking area to ensure proper initialization (Kaarthik et al., 2017). The sensors and cameras should be stationary during the system architecture initialization process.

### 5.2 Image Acquisition

Once the system initialization is complete, the second step involves activating another processing module in known as the Image Acquisition module. The Image Acquisition module allows the use of cameras to capture images of the parking lot (Kaarthik et al., 2017). High-resolution cameras that are already present in the parking lot can be used to capture images for processing.

### 5.3 Image Segmentation

Image segmentation is an image processing technique which can be used to recognise and examine the picture at a glance. Each part of the image is obtained by the image segmentation process (Yusnita et al., 2012). By considering the number of pixels that are present in the frame or captured image, visual characteristics are obtained. The obtained image after segmentation will be a higher-quality image as output. Figure 7 shows the flowchart of the system module in image segmentation.

The set of pixels that are in the result can collectively provide the entire image after the result. The outline recognizes the empty parking lot (Yusnita et al., 2012). The process included in the image segmentation and clustering process divides the image into several clusters. The particulars of the image segmentation process are

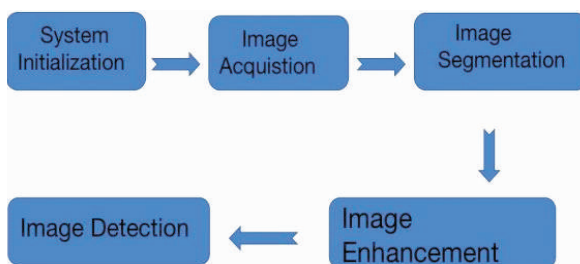


Figure 7. Overview of System Module

shown in Figure 8.

### 5.4 Image Enhancement Module

The image segmentation method provides individual parts of the image, and the specific attributes of each part can be obtained by analyzing the number of pixels in the captured image. The resulting segmented image is of higher quality and can be used to identify vacant parking slots. However, undesirable noise may occur during the image binarization process, and it may be necessary to track down the contours of the recognized image. Additionally, since the cameras take pictures from different locations, the images may contain noise that needs to be removed (Kaarthik et al., 2017).

The resulting disturbance in the picture can be eliminated using a method called morphology. Morphology is a special method used to ignore defects detected during image segmentation. The morphological mechanism goes through the following processes: stretching, tension, and opening and closing processes, and between these four processes, the morphological opening and closing process is the most commonly used denoising process.

### 5.5 Image Defection Module

The Image Recognition Engine accurately identifies edge and contour boundaries in images, which are obtained through the Image Enhancement module. To obtain the exact shape of the image, area and perimeter parameters are taken into consideration (Kaarthik et al., 2017). A correct representation of the image format is necessary to provide information for the driver to park the car in an open parking slot. The shape of an image can be obtained using the following formula, as given in Equation 1,

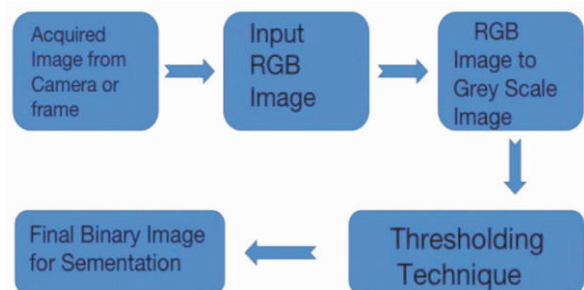


Figure 8. Image Segmentation Process

$$\text{Shape} = (4 \times \pi \times \text{area}) / (\text{perimeter}^2) \quad (1)$$

This formula is often used as a feature descriptor in image detection, which is a mathematical representation of a shape used for object recognition or classification. The "shape factor" described by this formula measures how circular or compact a shape is, based on the ratio of its area to its perimeter. A perfect circle has the maximum possible value for this shape factor, and as a shape becomes more elongated or irregular, the value of the shape factor approaches zero.

## 6. Proposed System

A general design of the proposed system is illustrated in Figure 3. Initially, the image is converted to grayscale to simplify the calculation and determination of its intensity. Next, the grayscale image is converted to a binary image using thresholding. Morphological operations such as closing, opening, dilation, and erosion are then applied to modify the binary pixels of an object in the original image. The final stage of processing involves object detection and the calculation of parked objects. The information viewer displays the availability of parking spaces and their locations. Figure 8 provides details of the image processing algorithm used in the proposed system. Captured images are RGB images that have three channels, including red, green, and blue. The color channels are separated, and each channel is converted to a gray image with varying intensity. Each gray image is then thresholded to obtain three binary images, which are merged to obtain a single binary image. The resulting binary image is processed using dilation to extend the object or structure, followed by erosion to reduce the area. The dilated and eroded image is then post-processed with hole-filling and closing to remove unwanted items. The parking space is marked with a colored rectangle in the final binary image of the empty parking slot for vehicles, as shown in Figure 9.

This binary image has been created by applying various image processing techniques to the original image captured by the camera, such as thresholding, edge detection, and morphological operations. These techniques are used to isolate and extract the parking



Figure 9. Final Binary Image

slots from the background and other objects in the scene, as well as to identify the boundaries and contours of each parking slot. The resulting binary image is a black and white image where the parking slots are shown as white regions, and the background is black. This binary image is used to determine which parking slots are available for use and which ones are currently occupied by analyzing the position, size, and shape of each white region.

When vehicles are parked in the parking space, they cover the sign. Picture processing is then used to detect the yellow tag as an available parking place. Object detection is performed by dividing the image based on each parking space to retrieve data using specific formulas. The obtained data is processed in such a way that information about the availability of parking spaces is gained. This information is displayed as a system output, as shown in Figure 10.

According to Figure 10, the red color indicates a filled parking space, while the green color indicates an available parking space. The available parking space is displayed outside the mall and can also be accessed through a web page that provides real-time information.

## 7. Results and Discussion

Vehicle parking can be efficiently and accurately managed using image processing techniques without the need for manual intervention. The proposed system displays the parking lot area with available parking spaces. Empty parking spaces can be easily identified based on a threshold value and the display, which can be as large as an LED screen. Additionally, a web application is available for people to view parking availability from any

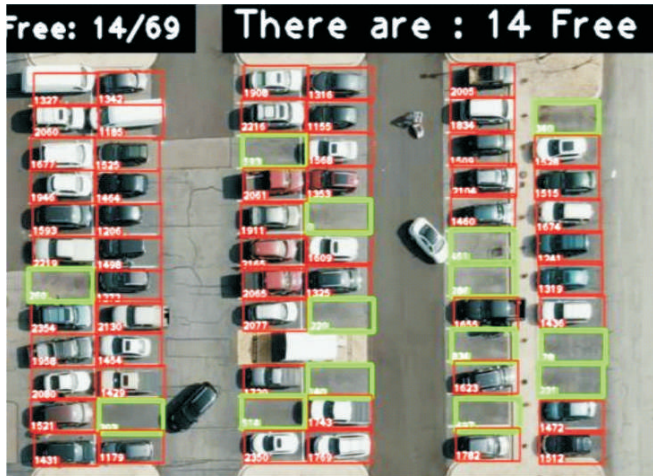


Figure 10. Final Output Image

location using their mobile devices or computers. Figure 11 shows a web page designed for navigating available parking slots. The proposed architecture can greatly benefit parking management by reducing time consumption and maximizing the use of parking spaces, making it ideal for the unbiased parking of vehicles.

### 7.1 Working

When a user enters the web page, they are presented with an interface that offers two options: they can either view the parking cameras or the parking spots, where they can obtain information about available parking slots. This enables them to find a free slot to park their vehicle before entering the parking area, saving them time and energy. It can also help reduce traffic and lower the risk of accidents in the parking area.

### 7.2 Accuracy

Figure 12 demonstrates that the accuracy of the system varies based on the color of the car. The system has higher accuracy for light colored cars, such as white, gray, and silver, as these colors reflect more light and have higher contrast, making it easier for the system to detect the edges and contours of the cars and distinguish them from the background and other objects in the scene. However, the accuracy of the system may be lower for darker colored cars, such as black and dark blue, as these colors absorb more light and have low contrast, particularly in low light or shadowy conditions, making it more difficult for the system to detect the edges and contours of the cars.

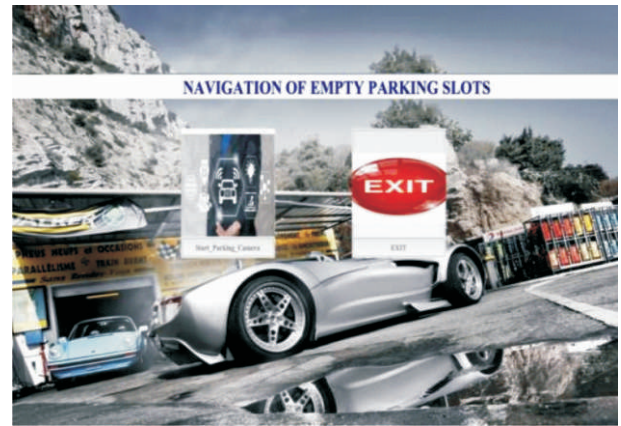


Figure 11. Web Page for the Navigation of Empty Parking Slots

Color	Prediction (%)	Accuracy (%)
Red	99%	98%
Violet	99%	98%
Ash	98%	95%
Black	98%	90%
White	98%	90%

Figure 12. Accuracy

### Conclusion

Once the suggested design is executed in everyday life, parking a vehicle will become effortless. The process of picking out parking spaces and identifying the number of vacant parking spaces can be done using image processing techniques. The parking lot is easily recognizable with no waiting period. The proposal will be put into action for large parking spaces and the parking issue is resolved using image processing, IoT and Wi-Fi. Image processing plays a vital role in calculating the free slots in the parking area of the vehicle to obtain information quickly. By using this proposed system, the availability of free slots are shown on the web application as live or real-time. The main advantage of this system is that it reduces the time spent searching for free slots, and the user can access the website effectively and reduce traffic congestion. The Python language is used, making it easy to compile and rectify any errors or modifications to

the original code. This study proposes a system that enhances the performance of saving users time to locate an appropriate parking space and reduces the general costs of moving to the chosen parking space. The most obvious finding to emerge from this study is that the proposed smart car parking system ensures the reduction of transmitted data through the network and saved energy in the perception layer. On the application layer side, the goal is to save the user time, avoid traffic congestion, find available parking spaces, and reduce car emissions while searching for empty parking spaces.

## References

- [1]. Abel, A., Aulia, S., Ramadhan, D. N., & Hadiyoso, S. (2021). A prototype of parking space information system based on image processing. *Jurnal Rekayasa ElektriKa*, 17(3), 185-190.
- [2]. Alsafery, W., Alturki, B., Reiff-Marganiec, S., & Jambli, K. (2018, April). Smart car parking system solution for the internet of things in smart cities. In 2018 1<sup>st</sup> International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-5). IEEE. <https://doi.org/10.1109/CAIS.2018.8442004>
- [3]. Farooqi, N., Alshehri, S., Nollily, S., Najmi, L., Alqurashi, G., & Alrashedi, A. (2019, November). UParking: Developing a smart parking management system using the internet of things. In 2019 Sixth HCT Information Technology Trends (ITT) (pp. 214-218). IEEE. <https://doi.org/10.1109/ITT48889.2019.9075113>
- [4]. Gupta, R., Budhiraja, N., Mago, S., & Mathur, S. (2020). An IoT-based smart parking framework for smart cities. In *Data Management, Analytics and Innovation: Proceedings of ICDMAI 2020*, 1, 19-32. Springer Singapore. [https://doi.org/10.1007/978-981-15-5616-6\\_2](https://doi.org/10.1007/978-981-15-5616-6_2)
- [5]. Jeffrey, J., Patil, R. G., Narahari, S. K. K., Didagi, Y., Bapat, J., & Das, D. (2012, August). Smart parking system using wireless sensor networks. In *The Sixth International Conference on Sensor Technologies and Applications* (pp. 306-311).
- [6]. Kaarthik, K., Sridevi, A., & Vivek, C. (2017, April). Image processing based intelligent parking system. In 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE) (pp. 1-4). IEEE. <https://doi.org/10.1109/ICEICE.2017.8191876>
- [7]. Pham, T. N., Tsai, M. F., Nguyen, D. B., Dow, C. R., & Deng, D. J. (2015). A cloud-based smart-parking system based on Internet-of-Things technologies. *IEEE Access*, 3, 1581-1591. <https://doi.org/10.1109/ACCESS.2015.2477299>
- [8]. Rane, S., Dubey, A., & Parida, T. (2017, July). Design of IoT based intelligent parking system using image processing algorithms. In 2017 International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1049-1053). IEEE. <https://doi.org/10.1109/ICCMC.2017.8282631>
- [9]. Suhr, J. K., & Jung, H. G. (2016). Automatic parking space detection and tracking for underground and indoor environments. *IEEE Transactions on Industrial Electronics*, 63(9), 5687-5698. <https://doi.org/10.1109/TIE.2016.2558480>
- [10]. Thangam, E. C., Mohan, M., Ganesh, J., Sukesh, C. V., & Prof, A. (2018). Internet of Things (IoT) based smart parking reservation system using raspberry-pi. *International Journal of Applied Engineering Research*, 13(8), 5759-5765.
- [11]. Yusnita, R., Norbaya, F., & Basharuddin, N. (2012). Intelligent parking space detection system based on image processing. *International Journal of Innovation, Management and Technology*, 3(3), 232-235.

## ABOUT THE AUTHORS

*Srujan Panthula is a student in the Department of Information Technology at Guru Ghasidas Vishwavidyalaya (Central University), Bilaspur, India.*



*Yaramala John Wesley is a student in the Department of Information Technology at Guru Ghasidas Vishwavidyalaya (Central University), Bilaspur, India.*



*Alle Saiteja Wesley is a student in the Department of Information Technology at Guru Ghasidas Vishwavidyalaya (Central University), Bilaspur, India.*



*Dr. Sudha Mishra is working in the Department of Information Technology at Guru Ghasidas Vishwavidyalaya (Central University), Bilaspur, India.*

*Agnivesh Pandey is working as an Assistant Professor in the Department of Information Technology at Guru Ghasidas Vishwavidyalaya (Central University), Bilaspur, India.*

## TRACKING MISSING PERSONS USING FACIAL RECOGNITION

By

BHAVIKA GUPTA \*

PALAK AGARWAL \*\*

DISHA DEVALIA \*\*\*

\*..\*\*\* Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, Maharashtra, India.

Date Received: 12/04/2023

Date Revised: 25/04/2023

Date Accepted: 10/05/2023

## ABSTRACT

Finding missing people is a time-critical and labor-intensive task and the longer it takes to locate the person, the lower the likelihood of a successful outcome. To address this challenge, an integrated and centralized database of missing persons using Aadhar card details was developed. The approach incorporates facial recognition technology, specifically the deep face algorithm, which has shown high accuracy in identifying individuals. While facial recognition has been in use for several years, recent advancements have made it easier to identify individuals accurately. By leveraging Artificial Intelligence (AI) powered facial recognition technology, officials can enhance and streamline the process of finding, tracking, and retrieving missing persons. The system matches facial features with the data stored in Aadhar cards, providing a reliable means of identification. This research presents a system that centralizes data, improving the efficiency of locating missing individuals. By utilizing facial recognition and centralizing data, the system offers an efficient approach to find missing people. The integration of technology and data allows quick and more accurate identification, increasing the chances of locating missing persons promptly.

Keywords: Face Recognition, Face Detection, Deep Face, Tracking and Retrieving.

## INTRODUCTION

Human trafficking remains one of the most harrowing issues plaguing the world today. It involves the transportation and transfer of individuals for the purpose of exploitation and financial gain. This heinous crime affects people of all genders, ages, and ethnicities, spanning various historical periods. Unfortunately, India continues to grapple with the prevalence of human trafficking, making it the second-largest organized crime in the country.

To combat this issue, a centralized database within a dedicated portal for missing people was developed. This portal serves as a repository where information about missing people can be uploaded by their families or law enforcement officials. By including the missing person's

Aadhar card details, facial recognition technology can be employed to match their Aadhar card image with the individual's physical appearance, utilizing accurate models. The missing people portal, as depicted in Figure 1, provides a comprehensive solution by centralizing the data, allowing easy access, tracking, and retrieval of information in one place. The integration of facial recognition significantly reduces the need for extensive manpower, expedites the search process, and enhances the overall efficiency.

By leveraging the power of technology and centralizing crucial data, this portal offers a vital tool in the fight against human trafficking. It streamlines the process of locating missing individuals, enabling prompt action and increasing the likelihood of their safe recovery.

## 1. The Issue of Missing Persons in India

In India, the number of reported cases of missing adults, women, and children has reached unprecedented levels. The country grapples with a significant challenge concerning missing persons. Every year, a number of



This paper has objectives related to SDGs



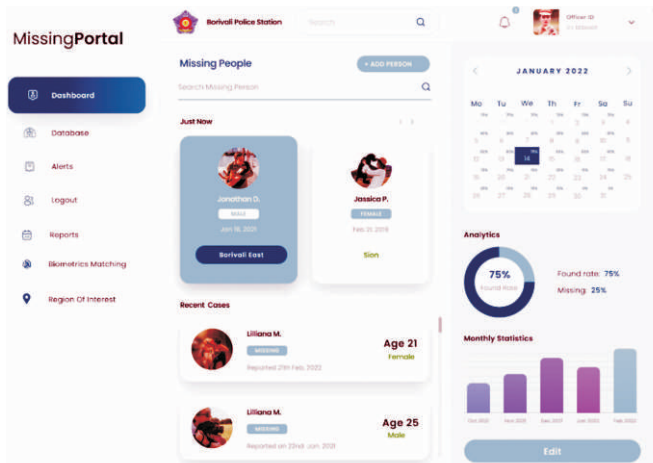


Figure 1. Portal of Missing People

individuals from diverse age groups and backgrounds vanish without a trace. This issue casts a shadow of concern over families, communities, and society as a whole, emphasizing the urgent need for effective solutions.

One of the major factors contributing to the problem of missing persons in India is the absence of a robust and coordinated system for reporting and tracking such cases. Delays often occur in registering missing person reports, impeding crucial initial hours of investigation. The lack of a centralized database hinders efforts to match missing persons with unidentified bodies or locate individuals who have been found but remain unidentified. Another contributing factor is the prevalence of human trafficking networks. Regrettably, India serves as a source, transit, and destination country for human trafficking, with individuals frequently being abducted or deceived with false promises. Many missing persons are believed to have become victims of this abhorrent crime, exacerbating an already grave issue. Addressing the problem of missing persons in India requires comprehensive measures. It involves establishing a robust reporting and tracking system, improving coordination among law enforcement agencies, implementing a centralized database, and intensifying efforts to combat human trafficking. By tackling these challenges head-on, India can take significant strides towards mitigating the problem and ensuring the safety and well-being of its citizens.

The shocking reality is that in 2020, when India was grappling with the Covid-19 outbreak, a staggering 6,701,145 individuals were still listed as missing nationwide. The inefficiency stems from inconsistent databases and poor communication among officials. The traditional methods of finding missing people prove tedious and lack technological advancements. Moreover, rescued victims may suffer from trauma and struggle to identify themselves, further complicating the recognition process. This issue becomes even more challenging when it involves children, who may struggle to provide personal details for identification purposes. To address these challenges, the proposed portal allows the families of victims to file missing complaints online, streamlining the process. After 24 hours, police officials launch First Information Reports (FIRs) for the reported cases. Facial recognition technology is employed to match the concerned person's face with available records, enabling authorities to report the individual's whereabouts to their family members. This approach significantly improves efficiency and expedites the overall process. The research work also highlights the significance of face alignment and feature extraction in the facial recognition process. The proposed technique is evaluated for accuracy and False Acceptance Rate (FAR), demonstrating superior performance compared to other state-of-the-art approaches. The computational analysis conducted on the Face Recognition Grand Challenge (FRGC) dataset showcases an impressive accuracy rate of 96% with a FAR of 0.1. By utilizing advanced facial recognition technology and emphasizing the importance of precise face alignment, the system offers an efficient and reliable solution for recognizing missing individuals.

## 2. Literature Survey

Facial recognition is a popular application of artificial intelligence. It has been found that facial matching can be done to identify a person. But earlier, the accuracy that was observed was less. Sharma et al. (2016) described the processes involved in face recognition, like face alignment and feature extraction. It also emphasizes the importance of the face alignment; thus, the accuracy

and False Acceptance Rate (FAR) are observed by using the proposed technique. The computational analysis shows better performance than other state-of-the-art approaches. The work has been done on the Face Recognition Grand Challenge (FRGC) dataset and gives an accuracy of 96% with a FAR of 0.1.

Ren et al. (2017) explains how a random forest model is applied to each feature point by pixel difference and the regression estimation of forest model to estimate the shape of training samples for linear least squares fitting. The model is then used to test the sample feature point location regression estimation and shape optimization, so as to realize the automatic localization of facial feature points. Though the Random Forest algorithm is 95% accurate, the problem with this method is that it is quite a time-consuming process.

The Cam Shift algorithm is quite expensive and also time-consuming (Goyal et al., 2017). The intention of the paper is a deep study of face detection using open CV. A tabular comparison is performed in order to better understand the algorithms. It talks about various algorithms like Adaboost and Haar cascades. This paper aims to help understand the best prerequisites for face detection.

The concept of multi-view face detection using Convolutional Neural Networks (CNN) is extended to provide a tagging system for the detected faces (Mehta et al., 2018). For face detection, we use the Deep Dense Face Detector, which uses a single model based on deep convolutional neural networks. All the detected faces are recognized using the Local Binary Pattern Histograms (LBPH) method. Precision, recall, and F-measure are the parameters used to measure the performance of the algorithm. An accuracy of 85% is achieved for tagging the faces that are successfully detected. It will be used with CCTV footage to identify people in cases of theft. But the accuracy observed was 85%.

Zhao et al. (2020) proposed a joint Bayesian framework in order to evaluate the similarity of feature vectors and achieve highly competitive face classification accuracy. Comprehensive experiments were conducted on our compiled Chinese Academy of Sciences' Pose,

Expression, Accessory and Lighting (CAS-PEAL) dataset and achieved a 98.52% face recognition performance. Moreover, our proposed face recognition system can robustly handle various face recognition attacks in various contexts.

There are more such papers describing a lot of problems that are observed related to the accuracy or detection of faces and the algorithm used. So there is a need to solve the problems using efficient algorithms to speed up the process.

### 3. Methodology

To combat human trafficking, especially child trafficking, an app or government portal linked with the Aadhaar information system may be conceptualized for the integration of information. All state and district-level police authorities have access to the missing person's information, including their photo and age, through the dedicated portal. This information may be used by the designated authorities for the search of missing persons in the locality under their jurisdiction. This would help the police authorities find missing persons in their localities with authenticated and integrated information systems (Gholape et al., 2021). In a similar vein, the Aadhaar-enabled integrated information system can be used to locate the address and contact information of parents or other relatives. So created for the person found by the police authorities or civil society, especially for the children who do not have proper or adequate information for their address or parents.

On the user side, individuals can submit photos of their missing loved ones through a user-friendly interface. The application will then use facial recognition technology to compare these images with a database of known faces, aiding in the search and recovery processes. Figure 2 shows the user-side application flow.

For users, family members need to log in to the application and provide all the necessary details about the missing individuals.

On the police side, the application flow is equipped with a specialized interface tailored for law enforcement officials. They have access to the facial recognition

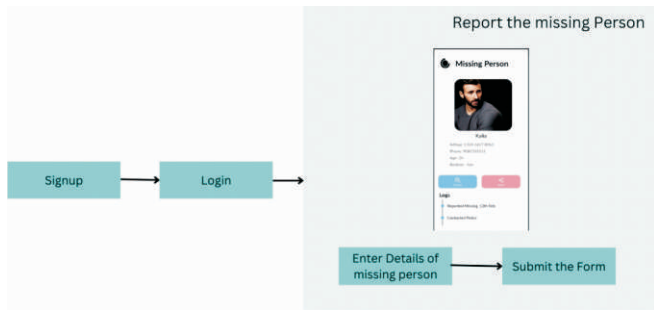


Figure 2. User Side Application Flow

system and database, where they can input missing person reports along with associated images. The application utilizes advanced algorithms to analyze the facial features of the missing person and compare them with the database. When a potential match is identified, the application generates alerts, enabling law enforcement officers to conduct further investigations and take appropriate action. This streamlined process empowers the police to efficiently track and locate missing persons using facial recognition technology, augmenting their capabilities in solving such cases. The police side application flow is illustrated in Figure 3.

The data filled out by the family members of missing persons is reflected on the dashboard of the respective police station.

After 24 hours, when the family members register a First Information Report (FIR), the filled-in data is verified alongside the FIR. Upon verification, the information is shared with the central database. Consequently, the specific details of the missing person become accessible to the police stations nationwide. If a similar or lost person is found, their identity can be verified using this portal, leveraging facial recognition algorithms for verification purposes.

### 3.1 Proposed Methodology

Figure 4 provides an overview of the proposed system and visualizes its various components. The proposed system aims to address the issue of missing persons by incorporating several features and functionalities to streamline the process. The "Person Missing Notification" component allows individuals or families to report a missing person through a dedicated platform. They can

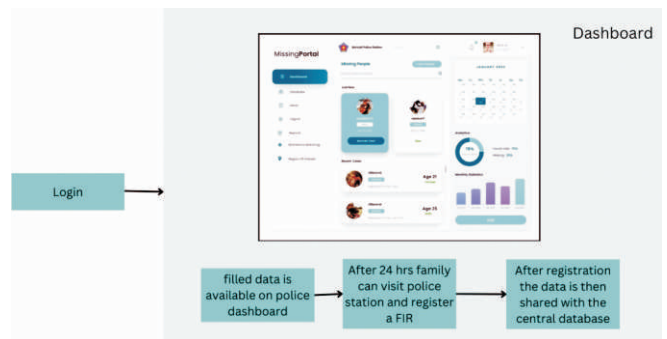


Figure 3. Police Side Application Flow

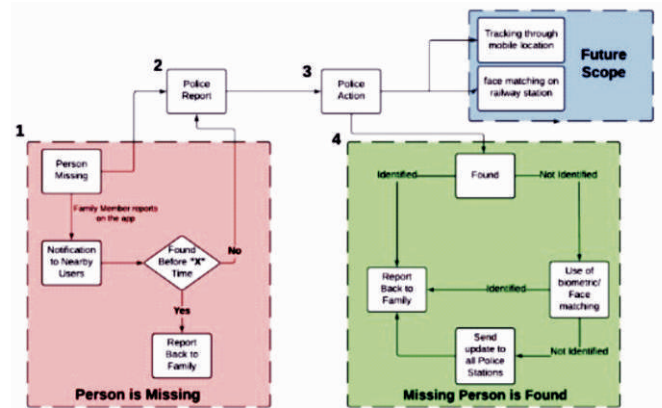


Figure 4. Proposed System

provide crucial details such as the person's name, photograph, physical attributes, and last known location. Once a missing person report is filed, the system generates a comprehensive police report. This report encompasses all relevant information provided by the reporting party, ensuring that the law enforcement agencies have accurate and up-to-date data. The system enables law enforcement agencies to access and manage the reported cases. They can assign officers, track progress, and communicate with the reporting party through the platform. This fosters effective coordination and facilitates timely action on missing person cases.

The system incorporates mobile tracking capabilities, enabling authorized personnel to track the location of the missing person through their mobile devices. This feature proves invaluable in narrowing down search areas and expediting the search process. Furthermore, the system integrates facial recognition technology with surveillance

systems at railway stations. When a missing person's photograph is uploaded, the system compares it with live or recorded video feeds from railway stations. If a potential match is detected, authorities are alerted, enabling them to take prompt action.

In cases where a missing person is found or identified, the system facilitates communication between law enforcement agencies and the concerned family. The family is promptly notified of the person's recovery, and the relevant details are shared to ensure a smooth reunification process. Overall, the proposed system leverages various components and functionalities to enhance the efficiency and effectiveness of locating missing persons, improve coordination among law enforcement agencies, and helps in timely reunification with their families.

#### 4. Existing Methodologies in Deep Face

Facial Recognition (FR) has been widely adopted as the most popular biometric method for verifying an individual's identity. It has found extensive applications in various domains, including the military, business, government, and daily life. The emergence of the Eigen face technique in the early 1990s marked a significant milestone in FR research (Wang & Deng, 2021).

Deep Face is a simple face recognition and facial attribute analysis framework for Python that considers age, gender, emotion, and race (Patil et al., 2021). Several processing layers are used in deep learning to learn data representations using various levels of feature extraction. State-of-the-art performance has been significantly enhanced by the deep learning technique, which is distinguished by using a hierarchical structure to combine pixels into an invariant face representation (Wang & Deng, 2021).

It is a hybrid face recognition framework that encases cutting-edge models like VGG-Face, Google FaceNet and OpenFace. On the Labeled Faces in the Wild (LFW) benchmark, OpenFace achieves near-human accuracy and introduces a brand-new categorization benchmark for mobile contexts (Zhao et al., 2020). DeepFace can outperform existing systems with very little adaptation

because it was trained on a sizable collection of face images from a population that is quite different from the one used to create the assessment benchmarks. Also, the system creates a very compact representation of the face, in contrast to other previous systems and move towards tens of thousands of appearance features. Taigman et al. (2014) states that DeepID is the benchmark in the field of Deep Feature Regression (DFR). This model's purpose is to recognize frontal faces (Shahsavari et al., 2020). ArcFace has a clear geometric interpretation to geodesic distance on a hypersphere. It routinely exceeds the current state of the art and is simple to deploy with little additional computation (Deng et al., 2019). Dlib is a C++-written open-source software library that runs on all platforms. Its layout is significantly influenced by concepts from component-based software engineering and contract design. This indicates a collection of separate software components, each of which is supported by excellent documentation (King, 2009) and SFace.

According to experiments, humans can recognize faces with an accuracy of 97.53%, whereas those models have already attained and surpassed that mark. Deep Face trained a 9-layer model on 4 million facial photos in 2014 to attain SOTA accuracy on the renowned LFW benchmark was close to human performance under unrestricted conditions (Deep Face: 97.35% vs. Human: 97.53%) (Wang & Deng, 2021). Comparison table for different models with respect to LFW score and YTF score is given in Figure 5.

They have access to several features if deep face-to-face recognition is used.

#### 4.1 Face Verification

Comparing two faces to see if they match or not is the task of face verification. Consequently, face verification is frequently utilized to assess a candidate's face to another. This can be used to verify whether a face on a physical document matches the one on an ID. Finding a face in an image database is the goal of the task known as "face recognition". Running face verification repeatedly is necessary for face recognition (Dass et al., 2012).

## 4.2 Facial Attribute Analysis

Describing the visual characteristics of face images is a key aspect of face recognition. Facial attribute analysis can be employed to determine various attributes such as age, gender, emotions, and potentially race or ethnicity.

## 4.3 Real-Time Face Analysis

By utilizing the live video feed from a webcam, users can perform face identification and facial feature analysis using Deep Face. This feature allows for practical testing

Model	LFW Score	YTF Score
Facenet512	99.65%	-
SFace	99.60%	-
ArcFace	99.41%	-
Dlib	99.38 %	-
Facenet	99.20%	-
VGG-Face	98.78%	97.40%
Human-beings	97.53%	-
OpenFace	93.80%	-
DeepID	-	97.05%

Figure 5. Comparison Table for Different Models with Respect To LFW Score and YTF Score

of deep face recognition tasks.

The evaluation results demonstrate that the SFace model, which employs deep face techniques, achieves an impressive score of 99.60%, as stated in the official documentation. The accuracy of different techniques assessed on the LFW dataset is depicted in Figure 6.

Regular convolutional neural networks are used in face recognition models, and they are in charge of representing faces as vectors. A face pair of the same person should be more similar than a face pair of two different people, according to our expectations. Different metrics, including cosine similarity, Euclidean distance, and L2 form, can be used to calculate similarity. In the default setting, cosine similarity is used. Based on experiments, the Euclidean and L2 form appears to be more stable than the cosine and regular Euclidean distances.

## 4.4 Triplet Loss

The embedding is shown as in a d-dimensional Euclidean space, it embeds an image  $x$  i.e.,  $d f(x) \in \mathbb{R}$ . It also constrain this embedding to exist on the d-dimensional

THE ACCURACY OF DIFFERENT METHODS EVALUATED ON THE LFW DATASET.						
Method	Public. Time	Loss	Architecture	Number of Networks	Training Set	Accuracy $\pm$ Std(%)
DeepFace [20]	2014	softmax	Alexnet	3	Facebook (4.4M,4K)	97.35 $\pm$ 0.25
DeepID2 [21]	2014	contrastive loss	Alexnet	25	CelebFaces+ (0.2M,10K)	99.15 $\pm$ 0.13
DeepID3 [36]	2015	contrastive loss	VGGNet-10	50	CelebFaces+ (0.2M,10K)	99.53 $\pm$ 0.10
FaceNet [38]	2015	triplet loss	GoogleNet-24	1	Google (500M,10M)	99.63 $\pm$ 0.09
Baidu [58]	2015	triplet loss	CNN-9	10	Baidu (1.2M,18K)	99.77
VGGface [37]	2015	triplet loss	VGGNet-16	1	VGGface (2.6M,2.6K)	98.95
light-CNN [85]	2015	softmax	light CNN	1	MS-Celeb-1M (8.4M,100K)	98.8
Center Loss [101]	2016	center loss	Lenet+-7	1	CASIA-WebFace, CACD2000, Celebrity+ (0.7M,17K)	99.28
L-softmax [104]	2016	L-softmax	VGGNet-18	1	CASIA-WebFace (0.49M,10K)	98.71
Range Loss [82]	2016	range loss	VGGNet-16	1	MS-Celeb-1M, CASIA-WebFace (5M,100K)	99.52
L2-softmax [109]	2017	L2-softmax	ResNet-101	1	MS-Celeb-1M (3.7M,58K)	99.78
Normface [110]	2017	contrastive loss	ResNet-28	1	CASIA-WebFace (0.49M,10K)	99.19
CoCo loss [112]	2017	CoCo loss	-	1	MS-Celeb-1M (3M,80K)	99.86
vMF loss [115]	2017	vMF loss	ResNet-27	1	MS-Celeb-1M (4.6M,60K)	99.58
Marginal Loss [116]	2017	marginal loss	ResNet-27	1	MS-Celeb-1M (4M,80K)	99.48
SphereFace [84]	2017	A-softmax	ResNet-64	1	CASIA-WebFace (0.49M,10K)	99.42
CCL [113]	2018	center invariant loss	ResNet-27	1	CASIA-WebFace (0.49M,10K)	99.12
AMS loss [105]	2018	AMS loss	ResNet-20	1	CASIA-WebFace (0.49M,10K)	99.12
Cosface [107]	2018	cosface	ResNet-64	1	CASIA-WebFace (0.49M,10K)	99.33
Arcface [106]	2018	arcface	ResNet-100	1	MS-Celeb-1M (3.8M,85K)	99.83
Ring loss [117]	2018	Ring loss	ResNet-64	1	MS-Celeb-1M (3.5M,31K)	99.50

Figure 6. Accuracy of Various Techniques Evaluated on the LFW dataset (Wang & Deng, 2021)

hypersphere,  $\|f(x)\| = 1$ .

Using the live video feed from the webcam, this feature allows test face identification and facial feature analysis. An image of a particular person,  $a$ (anchor), is closer to all other photos of that same person,  $x^p$  (positive), than it is to any image of another person,  $x^n$ (negative) and it is given by Equations 1 and 2,

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 \quad (1)$$

$$\forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in T \quad (2)$$

where, the margin between pairs, both positive and negative, is denoted by  $\alpha$ .  $T$  has a cardinality of  $N$  and is the collection of all feasible triplets in the training set. The loss is then minimized and it is represented in Equation 3,

$$\sum_i^N \left[ \|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right] \quad (3)$$

There would be a lot of triplets that are simple to satisfy if one generated every feasible triplet (Schroff et al., 2015).

## 5. Datasets and Evaluation

Various models like VGG-Fae, Facenet, FaceNet512, ArcFace and SFace are evaluated and compared. Figure 7 features the dataset used for the analysis and evaluation.

The dataset shown above is utilized for evaluating the algorithm. This dataset comprises multiple faces, serving as a database of citizens. Our objective is to compare the user input with the images in this dataset and determine whether the input face matches any of the pictures in the database. The user input image is displayed in Figure 8.

The image presented above represents the user's input. It is compared with each picture in the dataset to verify if this face exists in the dataset or not. The outcomes for various models are shown in Figure 9.

According to the results obtained for the VGG-Face, Facenet, FaceNet512, ArcFace, and SFace models, all of them were successful in comparing and identifying the faces in the dataset. However, there are variations in parameters such as distance and processing time among these models.

The first model, VGG-Face, yielded a positive response

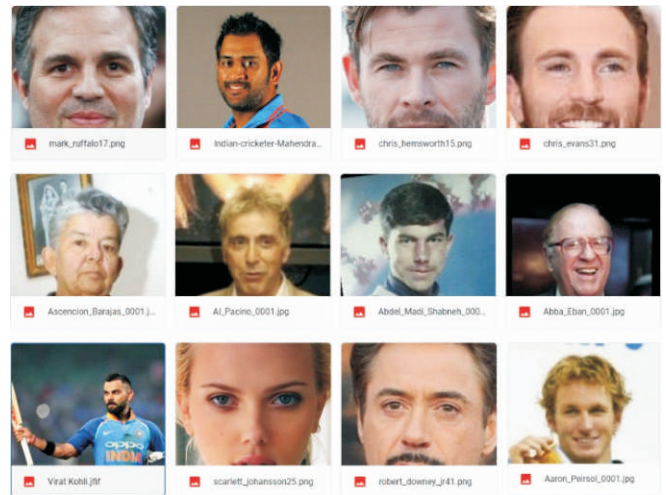


Figure 7. Dataset Used



Figure 8. User Input Image

with a distance of 0.149036, which is lower compared to the results obtained from other models. The time taken for processing was 1.965082. The Facenet model provided a positive response with a distance of 0.185329 and a processing time of 1.927456. The Facenet512 model produced a positive response with a distance of 0.265612, which is the highest among all the models, and a processing time of 2.190437. The ArcFace model resulted in a positive response with a distance of 0.222220 and a processing time of 1.405846. Lastly, the SFace model yielded a positive response with a distance of 0.415969, the highest among all the models, and the processing time was the lowest at 1.271688.

	model	found result	distance	time taken
0	VGG-Face	/content/drive/MyDrive/Images_DEMO/Copy of chr...	0.149036	1.965082
1	Facenet	/content/drive/MyDrive/Images_DEMO/Copy of chr...	0.185329	1.927456
2	Facenet512	/content/drive/MyDrive/Images_DEMO/Copy of chr...	0.265619	2.190437
3	ArcFace	/content/drive/MyDrive/Images_DEMO/Copy of chr...	0.222220	1.405846
4	SFace	/content/drive/MyDrive/Images_DEMO/Copy of chr...	0.415969	1.271688

Figure 9. Results using Various Models

## Conclusion

By efficiently comparing the images of missing persons with databases of known faces, including surveillance footage and unidentified bodies, facial recognition technology significantly expedites the identification process, increasing the likelihood of successful reunification with their families. Additionally, the integration of biometric face matching and real-time tracking capabilities empowers authorities to locate missing persons more effectively, even at railway stations or other high-traffic areas. However, it is crucial to address concerns related to privacy, data security, and the responsible use of facial recognition technology. Striking the right balance between public safety and individual rights is essential, ensuring that the system is implemented ethically and transparently. Embracing this technology-driven solution allows us to strive for a future where the anguish and uncertainty faced by families of missing persons are significantly reduced, fostering a society where no one is left behind. The proposed portal addresses the issues of centralization, tracking, and retrieving missing persons. By sharing information across the country, the portal increases the possibility of locating individuals and improves the speed of the process while reducing manual work. The utilization of an accurate model enhances the overall efficiency of the facial recognition system.

## References

[1]. Dass, R., Rani, R., & Kumar, D. (2012). Face recognition techniques: A review. *International Journal of Engineering Research and Development*, 4(7), 70-78.

[2]. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019).

Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4690-4699). <https://doi.org/10.1109/CVPR.2019.00482>

[3]. Gholape, N., Gour, A., & Mourya, S. (2021). Finding missing person using ML, AI. *International Research Journal of Modernization in Engineering Technology and Science*, 3, 1517-1520.

[4]. Goyal, K., Agarwal, K., & Kumar, R. (2017, April). Face detection and tracking: Using OpenCV. In 2017 *International Conference of Electronics, Communication and Aerospace Technology (ICECA)*, 1, 474-478. IEEE. <https://doi.org/10.1109/ICECA.2017.8203730>

[5]. King, D. E. (2009). Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10, 1755-1758.

[6]. Mehta, J., Ramnani, E., & Singh, S. (2018, February). Face detection and tagging using deep learning. In 2018 *International Conference on Computer, Communication, and Signal Processing (ICCCSP)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ICCCSP.2018.8452853>

[7]. Patil, S., Gaikar, P., Kare, D., & Pawar, S. (2021). Finding missing person using AI. *International Journal of Progressive Research in Science and Engineering*, 2(6), 101-104.

[8]. Ren, X., Ding, J., Sun, J., & Sui, Q. (2017, October). Face modeling process based on Dlib. In 2017 *Chinese Automation Congress (CAC)* (pp. 1969-1972). IEEE. <https://doi.org/10.1109/CAC.2017.8243093>

[9]. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and

clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 815-823).

[10]. Shahsavarani, S., Analoui, M., & Ghiass, R. S. (2020). M<sup>2</sup> Deep-ID: A novel model for multi-view face identification using convolutional deep neural networks. *arXiv preprint arXiv:2001.07871*. <https://doi.org/10.48550/arXiv.2001.07871>

[11]. Sharma, S., Shanmugasundaram, K., & Ramasamy, S. K. (2016, May). FAREC-CNN based efficient face recognition technique using Dlib. In *2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)* (pp. 192-195). IEEE. <https://doi.org/10.1109/ICACCCT.2016.7831628>

[12]. Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1701-1708).

[13]. Wang, M., & Deng, W. (2021). Deep face recognition: A survey. *Neurocomputing*, 429, 215-244. <https://doi.org/10.1016/j.neucom.2020.10.081>

[14]. Zhao, F., Li, J., Zhang, L., Li, Z., & Na, S. G. (2020). Multi-view face recognition using deep neural networks. *Future Generation Computer Systems*, 111, 375-380. <https://doi.org/10.1016/j.future.2020.05.002>

---

### ABOUT THE AUTHORS

Bhavika Gupta, Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, Maharashtra, India.

Palak Agarwal, Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, Maharashtra, India.

Disha Devalla, Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, Maharashtra, India.

# PERFORMANCE ANALYSIS OF PHOTONICS-BASED RF TRANSCEIVER FOR HIGH-SPEED DATA TRANSMISSION

By

MAYANK KAUSHIK \*

J. S. SIDANA \*\*

SHIVENDRA KUMAR \*\*\*

P. M. MENGHAL \*\*\*\*

\*-\*\*\* Military College of Electronics and Mechanical Engineering, Secunderabad, India.

Date Received: 27/05/2023

Date Revised: 02/06/2023

Date Accepted: 12/06/2023

## ABSTRACT

Current digital radar systems have limitations at higher frequencies, and a new approach is needed to be able to operate in different environments and at higher frequencies. Photonics offers a solution to these limitations. It has Ultra-Wide Bandwidth (UWB) and high precision, which allows for the flexible generation of highly consistent Radio Frequencies (RF) signals and the accurate direct digitization of signals without down-conversion. This research proposes a novel Wavelength-Division Multiplexing (WDM)-based photonics link for a radar demonstrator. The proposed system uses a single pulsed laser as a source to design a transceiver for high-speed data transmission and reception. The system can generate tunable radar signals and their echoes, avoiding the up/down conversion of radio frequency and ensuring both high resolution and simulation-based operation. The proposed system has the potential to revolutionize radar technology by enabling high resolution that can operate at higher frequencies. The system is also scalable and can be easily adapted to different radar applications.

Keywords: Radar, Photonics, OptiSystem, Laser, Resolution, Bandwidth, Carrier Frequencies, WDM.

## INTRODUCTION

In the last few decades, photonic techniques have become increasingly popular for the processing and development of Radio Frequency (RF) signals in wireless communications and radar applications. This is driven by the need for more accurate and efficient electronic models. Such systems have the potential to directly generate, process, and digitize signals with large bandwidths and high central frequencies while maintaining signal fidelity (low noise levels and minimal spurious signal content). This is likely to remain a challenge as the focus on higher signal agility and higher center frequencies for simulation-based frameworks increases.

In this system, the desired signal can be rapidly changed under software control. Signal agility can provide improved radar resolution (Ragheb et al., 2020). Coherent radar models use both the frequency (phase) and amplitude of the radar signal reflected from a given target to provide velocity/range data (Brandão et al., 2019). This system offers a novel design for a radar waveform with a low center frequency. The selected bandwidth provides the desired range resolution. The process involves shifting the waveform to a higher center frequency by multiplying it with a continuous-wave and stable RF carrier signal, then amplifying and transmitting it to the target. When the waveform reaches the radar receiver, it is shifted to the intermediate frequency by again multiplying it with a stable RF signal, and the target's velocity/range is extracted.

The Electronics and Radar Development Establishment (LRDE) has developed a state-of-the-art surveillance



This paper has objectives related to SDG



radar that is portable and battery-powered. The radar can scan a designated area to detect and track a variety of targets, including low-flying helicopters, combat, and light vehicles, groups of walking men, running or walking men, and crawling men. The radar can also detect targets with Doppler tone.

Conventional radars use microwave or Radio Frequency (RF) components for both the receiver and transmitter modules. However, these components have a limitation in only generating frequencies of a few GHz due to inherent high levels of noise and low phase stability.

The LRDE radar overcomes these limitations by using a new type of component called a terahertz (THz) sensor. THz sensors can generate frequencies of up to 100 GHz, which allows the radar to detect smaller and more distant targets than conventional radars. The LRDE radar is a new technology that can detect small and distant targets. This is used to help the Indian armed forces gather intelligence and target enemies.

## 1. Motivation of the Study

To investigate the potential of photonics-based RF transceivers for high-speed data transmission, it offers several advantages over traditional electronic RF transceivers, including,

- *Higher Data Rates:* Photonics-based RF transceivers can support much higher data rates than electronic RF transceivers. This is because light can travel much faster than electrons, and photonic components can be made much smaller than electronic components.
- *Longer Range:* Photonics-based RF transceivers can transmit data over much longer distances than electronic RF transceivers. This is because light is not affected by attenuation as much as radio waves.
- *Lower Power Consumption:* Photonics-based RF transceivers consume much less power than electronic RF transceivers.

## 2. Objectives

- To investigate the performance of photonics-based RF transceivers for high-speed data transmission.
- To identify the key factors that affect the performance

of photonics-based RF transceivers.

- To develop a design methodology for photonics-based RF transceivers that can achieve high performance and low cost.
- To evaluate the performance of the proposed design methodology using simulations and experiments.

## 3. Photonics-Based Radar

Radars based on completely electronic techniques suffer from a shortage of bandwidth, slow transmission speed, and low resolution. These limitations make it difficult to detect and identify small targets with low speed and at low altitudes in a complex electromagnetic environment. Laghezza et al. (2016) and Skolnik (2017) proposed photonics-based technologies that can efficiently deal with such issues. These systems have extraordinary features such as multidimensional multiplexing, high phase stability, very low transmission losses, high-speed signal processing, and broad bandwidth. Microwave photonics has emerged and proposed the opportunity of using unlimited bandwidth with flexibility. It can also realize the effectiveness of RF devices by covering the complete frequency ranges from a few MHz to tens of GHz (Liu et al., 2022).

De and Raj (2023) focused on photonics for the generation of RF signals and their detection for radars. The RF signals can be generated photonically, evading the need for microwave oscillators. This can be achieved by heterodyning (mixing) two lasers in a photodiode or optoelectronic oscillator. The optoelectronic oscillators exhibit the best stability, up to 40 GHz, but their adaptability is restricted to a few gigahertz (GHz). This can simply be achieved by choosing any mode pair out of the spectrum of Mode-Locked Lasers (MLLs). It is due to the condition of intrinsic phase locking. Furthermore, choosing laser modes with adjustable wavelength detuning permits adaptable RF carrier production with tunable frequency. The heterodyne approach can integrate an arbitrary modulation of the produced RF signal by modulating one of the beating modes of the laser (Cenkeramaddi et al., 2020).

The radar receivers depend on simulation-based radio

that requires Analog-to-Digital Converters (ADCs) that receive input signals, with high-speed sampling rates, a high signal-to-noise ratio, and are appropriate for broadband signals (Cenkeramaddi et al., 2021). Aperture jitter can restrict the electronic ADCs of the sampling process, i.e., the interval fluctuations period amongst two consecutive samples that destructs the precision of the digitized signal by integrating the noise floor. Such an issue can be resolved by optical sampling involving a sequence of short optical pulses. The RF signal can modulate these short optical pulses to be digitized. It can be further detected and digitized by each optical pulse at its peak amplitude (Keiser, 1983). Therefore, sampling rates of numerous GHz are attainable, where the sampling pulses have a timing jitter lesser than 100 fs. MLLs are specifically appropriate pulse sources, with low temporal jitter and narrow pulse widths (Gupta et al., 2021).

Photonics-based radar uses photonic components to generate, transmit, receive, and process radar signals (Pan & Zhang, 2020; Urick et al., 2015). This can lead to several advantages over traditional radar systems, such as increased bandwidth, improved performance, and reduced cost (Rai et al., 2021).

To implement a photonics-based radar system it uses a Mach-Zehnder Modulator (MZM) to modulate optical pulses from a Mode-Locked Laser (MLL). The modulated pulses are then amplified and transmitted. After reflecting off a target, the pulses are received by a photodiode and converted back to an electrical signal. The electrical signal is then processed by a Digital Signal Processor (DSP) to extract information about the target, such as its range, velocity, and size.

The use of photonics in radar systems is still in its early stages, but it has the potential to revolutionize the way radar is used in a variety of applications, such as surveillance, navigation, and weather forecasting (Brandão et al., 2019; Serafino et al., 2019).

### 3.1 Steps Involved in Photonics-Based Radar

- *RF Production:* The MLL generates optical pulses with a very high repetition rate (e.g., 1 GHz). These pulses

are then passed through a tunable optical filter, which selects two spectrum modes of the MLL. One mode is used to modulate the radar pulse, while the other is used for frequency shifting.

- *Sampling:* The modulated pulses are then amplified and transmitted. After reflecting off a target, the pulses are received by a photodiode and converted back to an electrical signal. The electrical signal is then passed through a serial-to-parallel converter, which parallelizes the signal into several lower-rate sample streams.
- *Digitization:* The parallel sample streams are then digitized using several electronic ADCs with high precision and low sampling frequency. The ADCs digitize the pulse peak power, which is then used to reconstruct the original signal.
- *Signal Processing:* The parallel digital samples are then interleaved in the DSP. The DSP then uses a variety of algorithms to process the signal and extract information about the target, such as its range, velocity, and size.

### 3.2 Transceiver of Photonic Radar System

This research discusses a photonic-based transceiver for radar systems. It typically consists of a pulsed baseband signal generator that is integrated with a photonic source, such as a laser (for long-distance transmission) or a Light-Emitting Diode (LED) (for short-distance data transmission), using a Mach-Zehnder modulator (MZM). The laser source acts as a high-frequency carrier for the message signal. It can then be combined with a filter that eliminates unwanted signals and noise, and the signal is passed to an amplifier for amplification. Wavelength Division Multiplexing (WDM) is used for optical transportation. Researchers have integrated various laser wavelengths, and the transmitter sends data for communication purposes with high-speed transmission. Signals from different wavelengths are combined at the transmitter side using a multiplexer, while the demultiplexer splits them into different signals at the receiver end. These merging and splitting steps are performed by a prism-like optical device. The antenna then transmits the signal

through an optical fiber to the receiving end, or via Free-Space Optics (FSO) (Ghatak, 2010). The receiver has another filter that removes channel noise and passes the signal to a low-noise amplifier. The photodiode then collects the obtained signal and feeds it to a digital signal processor for further processing.

Figure 1 shows the architecture of WDM-based photonic radar for optical transmission systems that involves the efficient design of a transceiver. A pseudorandom bit sequence generator is integrated with an optical laser source. The output of both is fed into a Mach-Zehnder interferometer with a gain of 20 dB and a noise figure of 4 dB. The novel interface works on wavelength division multiplexing by using three channels for data transmission. The data is sent in an encrypted form, making it difficult for attackers to understand it. The optical fiber has a length of 10 km and can operate up to 50 km, connecting both ends. At the receiver end, a photodiode is attached to receive the optical signal in its maximum amount to minimize data loss. Afterward, a filter is attached to remove unwanted signals.

Figure 2 shows the architecture of an optical data transmission system that is susceptible to noise disruptions. The same set of inputs is used in both Figure 1 and Figure 2. In Free Space Optics (FSO), the data is transmitted up to a range of 100 meters without any losses at the same power levels. The range can be increased by increasing the power. The final results are shown with the help of a Bit Error Rate (BER) analyzer and an RF spectrum analyzer.

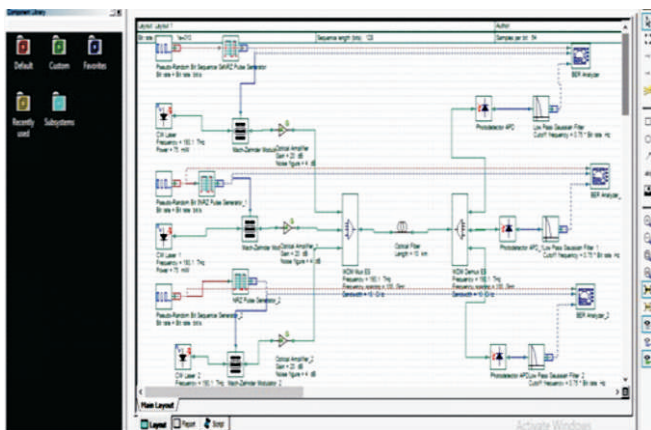


Figure 1. Optical Transmission Architecture for WDM Photonic Radar

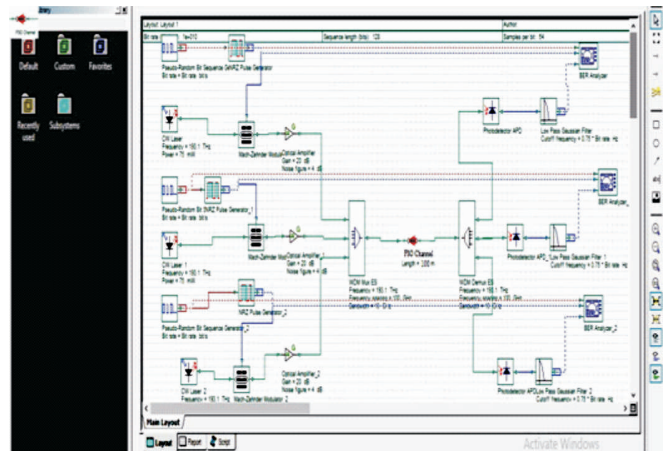


Figure 2. WDM Photonic Radar with Free Space Optics (FSO) Interface

## 4. Results and Discussions

OptiSystem software was used to develop an experimental setup and to obtain analytical results. The main components of the transceiver are shown in Figures 1 and 2. The operating band is in the THz range due to the use of a laser light source operating at 1550 nm. A pulse generator is integrated with the laser source to transmit encrypted data. The final results were obtained using a Bit Error Rate (BER) analyzer. The analyzer showed clear results without any noise or distortion, indicating that the obtained signal was minimally corrupted by the environment or other factors. BER is one of the key metrics to assess the quality of a data interface. Different data transmission models can be compared based on their BER. It is represented by the following mathematical expression,

$$BER = \frac{N_e}{N_t}$$

The Bit Error Rate (BER) is defined as the number of bits received in error divided by the total number of bits received in a defined interval of time. For contemporary data transmission systems, data is transmitted in larger blocks called packets. Each packet contains a fixed number of bits, which is determined by the type of network. The presence of a single bit error in a packet causes the entire packet to be corrupted. This can result in the loss of a large amount of data.

## 4.1 Q-factor

Q-factor is a measure of the quality of a digital signal from an analog perspective. It is calculated as the Signal-to-Noise Ratio (SNR). The SNR can be estimated experimentally by measuring the difference between the signal level and the noise level. The Q-factor and the error rate can then be used to assess the quality of the signal.

Q-factor is a key metric for evaluating the performance of Wavelength Division Multiplexing (WDM) systems. WDM systems transmit multiple data streams over a single optical fiber by using different wavelengths of light. The Q-factor is used to ensure that the data streams are transmitted reliably with a low error rate.

The optical hierarchy is a set of standards for optical networks. The goal of the optical hierarchy is to create a common platform for different network types to support advanced services.

BER and Q-factor are closely related. BER is a measure of the number of errors in a data stream, while the Q-factor is a measure of the quality of the signal. A high Q-factor corresponds to a low BER.

Q-factor is a good choice for monitoring WDM systems because it is easy to measure, independent of the digital signal format, and can be used to measure a wide range of transmitted speeds.

Figure 3 shows that the Q factor of a BER analyzer is 4.478, which indicates that the communication channel is of

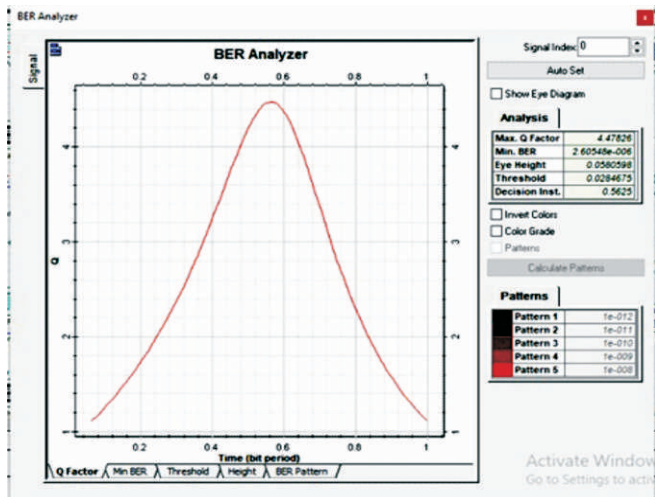


Figure 3 Output of Q factor on BER Analyzer with Q=4.478

good quality and is very reliable.

## 4.2 Minimum BER

The quality of transmission in telecommunication models can be directly analyzed by computing how well the output signal matches the input. Bit Error Rate (BER) is the signal quality metric used for digital communication models. Signal-to-Noise Ratio (SNR) is also used for analog communication systems.

Figure 4 represents the minimum Bit Error Rate (BER) of the BER Analyzer. The BER is a measure of the number of bits that are incorrectly received in data transmission. A lower BER indicates a higher quality of data transmission. In this case, the BER of 2.60548e-006 is very low, which means that the BER Analyzer is performing well and is capable of transmitting data with a high degree of accuracy.

## 4.3 Threshold

The BER threshold value is calculated as  $1.0E-n$ , where  $n$  is a positive integer from 1 to 10. The BER-threshold-signal-degrade value should be set below the Forward Error Correction (FEC) limit, which is the maximum BER that can be tolerated before the FEC mechanism can no longer correct all errors. This will ensure that proactive protection is initiated before the FEC mechanism is overwhelmed and packet loss occurs.

A high BER threshold value or a long interval between BER measurements may saturate the internal bit error counter register. This is because the bit error counter register can

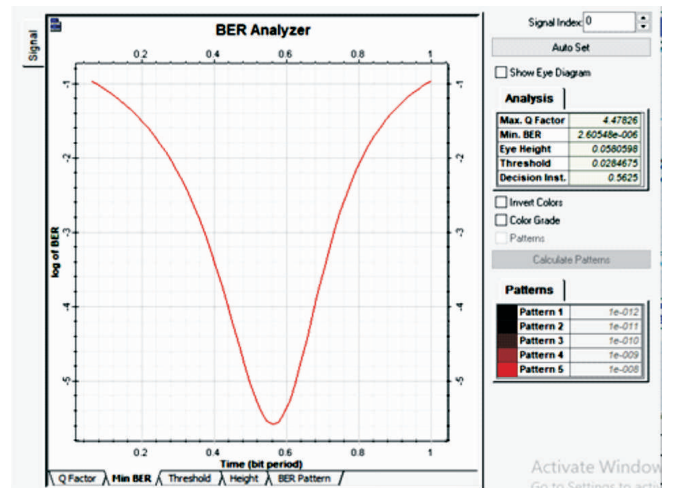


Figure 4. Output of Min BER on BER Analyzer having the Value 2.60548e-006

only store a limited number of errors before it overflows. If the BER threshold value is too high or the interval between BER measurements is too long, the bit error counter register may overflow before it can record all of the errors that have occurred. This can lead to inaccurate BER measurements and may prevent the system from detecting and responding to signal degradation. Figure 5 represents the output of threshold on BER analyzer having the value 0.0284675.

To analyze the performance of the designed WDM photonics-based transceiver, several parameters were considered and compared with the existing X-band surveillance radar. The operating band of photonics is THz,

which means license-free operation with unlimited band access. While the surveillance radar operates on C-band or X-band, which has a narrow bandwidth and is more susceptible to attack. WDM-based photonic systems can transmit data at much faster rates than X-band radar. The transmission range of photonic-based systems is up to 100 meters in free-space optics (up to 50 kilometers using optical fiber), while the transmission range of surveillance radar is 15-20 kilometers. In both cases, the RF carrier center frequency and RF bandwidth are the same, and the polarization is equivalent. Table 1 shows the comparative analysis between the technical specifications of X-band surveillance radar with designed WDM-based Photonic radar.

To further improve performance, semiconductor lasers, which are widely used in direct-modulation analog interfaces, can be replaced with solid-state lasers in microwave photonics. Solid-state lasers have lower gain dynamics than semiconductor lasers, which leads to enhanced high-frequency Relative Intensity Noise (RIN) performance and narrower line widths. By narrowing the line width, phase noise in the optical domain can be reduced.

## Conclusion

This research entails the successful simulation of a photonics-based transceiver. The architecture discussed in this research is combined with the complete radar

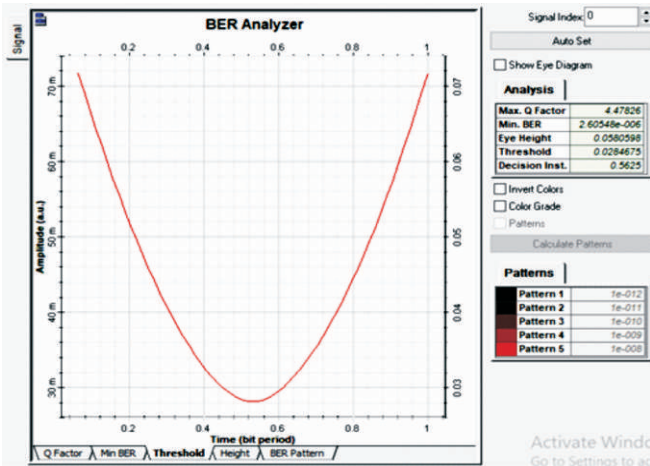


Figure. 5 Output of Threshold on BER Analyzer having the Value 0.0284675

Characteristic	WDM-based Photonic Radar	X-band Surveillance Radar
Operating Frequency	Photonic (193.1 THz or 1550 nm)	X-Band or C-band
Number of channels	3 (can be increased based on utilization)	21
Channel Spacing	20 MHz	20 MHz
Transmitter power	50 mW (Used in setup)	5 W (peak power)
IF	120 MHz (used the same as BFSR)	120 MHz
Instrumental Range	Up to 50 km (optical fiber)	15-20 km
Angle Resolution	> 2 degrees	> 4 degrees
RF carrier Center frequency	10.3 GHz (used the same as BFSR)	10.3 GHz
RF Bandwidth	400 MHz	400 MHz
Polarization	Linear, Vertical	Linear, Vertical
Display	Simulation results on Laptop	Plan Position Indicator (PPI) with zoom, B-scope
Processor	17 core processor	Pentium MMX Processor CPU @266 MHz 64 MB RAM
Memory	4 GB RAM	64 MB RAM
Display Monitor	14-inch laptop display	10.4-inch color Liquid Crystal Display (LCD) Thin-Film Transistor (TFT)
Pointing Device	Cursor	Mini Joystick

Table 1. Specifications of X-band Surveillance Radar and WDM-based Photonic Radar

system to check the final output and ascertain the characteristics of existing radars. The size of the radar will considerably reduce due to the utilization of photonics, as well as improved resolution and fewer distortions.

Photonics-based radars are the future of radars, as they have inherent characteristics of large bandwidth, reduced transmission loss, suitable and flat response, very high signal processing speed, and immunity to Electromagnetic Interference (EMI) environments. Photonics techniques have a broad range of new and upcoming applications, such as quantum information processing, water purification, Visible Light Communication (VLC), Light-Emitting Diodes (LEDs)/next-generation lasers, optical computers, sensors, and more.

In the defense/military, lasers are used for navigation purposes, providing information on the range of targets and missiles/guns. Lasers are also used in several treatment/diagnostic processes in the healthcare and medical field. For example, they can perform delicate surgical processes on the eye and other body parts of a human being.

## References

- [1]. Brandão, T. H., Scotti, F., Filgueiras, H. R. D., Alves, A. A. D. C., Onori, D., Melo, S., ... & Sodr e Jr, A. C. (2019). Coherent dual band radar system based on a unique antenna and a photonics based transceiver. *IET Radar, Sonar & Navigation*, 13(4), 505-511. <https://doi.org/10.1049/iet-rsn.2018.5046>
- [2]. Cenkeramaddi, L. R., Bhatia, J., Jha, A., Vishkarma, S. K., & Soumya, J. (2020, November). A survey on sensors for autonomous systems. In *2020 15<sup>th</sup> IEEE Conference on Industrial Electronics and Applications (ICIEA)* (pp. 1182-1187). IEEE. <https://doi.org/10.1109/ICIEA48937.2020.9248282>
- [3]. Cenkeramaddi, L. R., Rai, P. K., Dayal, A., Bhatia, J., Pandya, A., Soumya, J., ... & Jha, A. (2021). A novel angle estimation for mmWave FMCW radars using machine learning. *IEEE Sensors Journal*, 21(8), 9833-9843. <https://doi.org/10.1109/JSEN.2021.3058268>
- [4]. De, S., & Raj, A. A. B. (2023). A survey on photonics technologies for radar applications. *Journal of Optics*, 52(1), 90-119. <https://doi.org/10.1007/s12596-022-00897-x>
- [5]. Ghatak, A. (2010). *Optics*. McGraw-Hill, New Delhi.
- [6]. Gupta, S., Rai, P. K., Kumar, A., Yalavarthy, P. K., & Cenkeramaddi, L. R. (2021). Target classification by mmWave FMCW radars using machine learning on range-angle images. *IEEE Sensors Journal*, 21(18), 19993-20001. <https://doi.org/10.1109/JSEN.2021.3092583>
- [7]. Keiser, G. (1983). *Optical Fiber Communications*. McGraw-Hill, New York.
- [8]. Laghezza, F., Scotti, F., Onori, D., & Bogoni, A. (2016, May). ISAR imaging of non-cooperative targets via dual band photonics-based radar system. In *2016 17<sup>th</sup> International Radar Symposium (IRS)* (pp. 1-4). IEEE. <https://doi.org/10.1109/IRS.2016.7497319>
- [9]. Liu, Y., Zhang, Z., Burla, M., & Eggleton, B. J. (2022). 11 GHz Bandwidth photonic radar using MHz electronics. *Laser & Photonics Reviews*, 16(4), 2100549. <https://doi.org/10.1002/lpor.202100549>
- [10]. Pan, S., & Zhang, Y. (2020). Microwave photonic radars. *Journal of Lightwave Technology*, 38(19), 5450-5484. <https://doi.org/10.1109/JLT.2020.2993166>
- [11]. Ragheb, A., Esmail, M. A., Seleem, H., Sethi, W. T., Ashraf, M. A., Fathallah, H., & Alshebeili, S. A. (2020). Photonics-based multi-band/multi-mode radar signal generation. *Photonic Network Communications*, 39, 91-101. <https://doi.org/10.1007/s11107-019-00859-7>
- [12]. Rai, P. K., Ids e, H., Yakkati, R. R., Kumar, A., Khan, M. Z. A., Yalavarthy, P. K., & Cenkeramaddi, L. R. (2021). Localization and activity classification of unmanned aerial vehicle using mmWave FMCW radars. *IEEE Sensors Journal*, 21(14), 16043-16053. <https://doi.org/10.1109/JSEN.2021.3075909>
- [13]. Serafino, G., Scotti, F., Lembo, L., Hussain, B., Porzi, C., Malacarne, A., ... & Bogoni, A. (2019). Toward a new generation of radar systems based on microwave photonic technologies. *Journal of Lightwave Technology*, 37(2), 643-650. <https://doi.org/10.1109/JLT.2019.2894224>
- [14]. Skolnik, M. I. (2017). *Introduction to Radar System*.

McGraw-Hill, New York.

[15]. Tang, Z., Li, Y., Yao, J., & Pan, S. (2020). Photonics based microwave frequency mixing: Methodology and applications. *Laser & Photonics Reviews*, 14(1), 1800350.

<https://doi.org/10.1002/lpor.201800350>

[16]. Urick, V. J., Williams, K. J., & McKinney, J. D. (2015). *Fundamentals of Microwave Photonics*. John Wiley & Sons.

---

## ABOUT THE AUTHORS

*Mayank Kaushik, Military College of Electronics and Mechanical Engineering, Secunderabad, India.*

*J. S. Sidana, Military College of Electronics and Mechanical Engineering, Secunderabad, India.*

*Shivendra Kumar, Military College of Electronics and Mechanical Engineering, Secunderabad, India.*

*Dr. P. M. Menghal, is currently working as an Associate Professor at Military College of Electronics and Mechanical Engineering, Secunderabad, India.*

## MUSIC RECOMMENDER SYSTEM USING FACIAL FEATURE RECOGNITION

By

HARSH MISHRA \*

DEEP KOTHARI \*\*

PRAJWAL JAISWAL \*\*\*

SHEETAL RATHI \*\*\*\*

\*-\*\*\*\* Department of Computer Engineering, Thakur College of Engineering & Technology, Mumbai, India.

Date Received: 30/04/2023

Date Revised: 25/05/2023

Date Accepted: 05/06/2023

### ABSTRACT

Facial expressions convey a lot of information about the emotional state of a person and even a small change in it can help to detect changes in an individual's mood. Researchers have conducted numerous studies in the fields of machine learning and computer vision, training ML models to detect various human emotions or moods based on captured facial features. In this research, a system was proposed which is capable of identifying the user's emotions and moods and suggesting a list of appropriate songs to improve their mood. Incorporating mood detection offers significant benefits to customers' mental health and personal satisfaction. The objective of this system is to capture the user's image, detect their mood and create a music player system that recommends music based on the user's real-time mood, utilizing a web camera and deep learning algorithms.

Keywords: Face Recognition, Emotion Detection, Music, Mood Detection, Recommendation System.

### INTRODUCTION

The easiest way to understand a person in a conversation is by observing their facial expressions and body language. Facial expressions are fundamental to human communication and can be considered a form of non-verbal communication (Raut, 2018). In recent research areas such as digital imaging, pattern recognition, and computer vision, face recognition and emotion recognition have garnered significant attention (Raut, 2018). Face recognition involves a two-step process of locating faces in photos or videos, regardless of whether the face is personal or not. Emotion recognition is widely used in applications like smart card systems, surveillance, image database research, crime prevention, video indexing, civil applications, security, adaptive human-

computer interfaces, and multimedia contexts.

Hence, the primary objective of this research is to develop a system capable of identifying various user emotions, ranging from joy to sadness and from calmness to shock. Once the user's emotion is evaluated, the proposed system presents them with a music playlist containing music clips from different genres that can enhance their mood. For instance, if the user's facial expression is categorized as happy, the most suitable playlist, such as a classical music playlist, is activated, offering a variety of music clips for the user to choose from. The main aim of this research is to address user needs and create a music player system that recommends music based on the user's real-time mood, utilizing a web camera and deep learning algorithms.

### 1. Bridging the Gap: Personalized Emotion-Based Music Recommendation System

Nowadays, music platforms have made it incredibly easy to access a vast library of music. These platforms continuously work on enhancing music organization and



This paper has objectives related to SDG



search management, addressing the issue of choice, and facilitating the discovery of new music pieces. Recommendation systems have gained popularity, assisting users in selecting suitable music for various occasions. However, there is still a gap between personalization and emotion-based recommendations. Music holds a significant influence over humans and is widely used for relaxation, mood regulation, stress reduction, and maintaining mental and physical well-being (Donald, 2014). Music therapy encompasses a broad range of clinical settings and methodologies to support overall well-being (Hemanth et al., 2018). This research aims to present a personalized music recommendation system that takes into account the listener's emotions. By combining artificial intelligence technologies and generalized music therapy approaches, a recommendation system is being developed to help individuals select music suitable for different life situations while promoting their mental and physical health.

Traditional music players require users to manually browse through their playlists and select songs that align with their mood and emotional state. Although these features fulfill basic user needs, they still necessitate manual searching and song selection based on the user's current mood and behavior. Users often feel the need and desire to browse their playlists sporadically, depending on their mood and emotions.

## 2. Literature Review

Mahadik et al. (2021) mentioned that the implementation was done in two parts. First was mood detection, and second was the song recommendation system. Mood detection was further divided into two modules, the first being face detection and the second being mood detection. For face detection, the Java face detector class was chosen over the OpenCV library of Python since OpenCV is computationally expensive and challenging to integrate with the Android app. For mood detection, Mobile Net and Keras were used, and again, Keras is computationally heavy and slow on Android, so Mobile Net was also used. For the facial expression dataset, two datasets from Kaggle were used: the FER 2013 dataset

and the MMA facial expression recognition dataset. For the second part of the implementation, i.e., the song recommendation system, the dataset is downloaded from Kaggle, which is labeled with different moods, and Firebase is chosen to be the dataset.

Li et al. (2020) suggested copying a few data points and extracting only useful features from images, proposed a new face cropping and rotation strategies and a simplification of the convolutional neural network (CNN) to make data more abundant so that only useful facial features can be extracted. Experiments to evaluate the proposed method were performed on the CK and JAFFE databases. The experiments recorded high accuracy of 97.38% and 97.18% for 7-class experiments on the CK+ and JAFFE databases, respectively. The results show that the proposed FER approach achieved competitive results in terms of training time, testing time, and recognition accuracy. Furthermore, the proposed method can be implemented on an ordinary computer without GPU acceleration.

Gilda et al. (2017) divided the overall problem into three sections. The first one is emotion detection. CNN was used, and the dataset used was FER2013. The final network was trained on 20K images and tested against 5K images. This emotion detection model was recorded with an accuracy of 90%. The second section was to apply deep learning and classify the songs in the song database into different emotions. Again, CNN was used for that. The model achieved a classification accuracy of 97.6%. The third section is to recommend songs. Once the mood is detected, the user was provided with a playlist of all the songs classified as being in that mood.

James et al. (2019) proposed an intelligent agent that sorts a music collection based on the emotions conveyed by each song and then suggests an appropriate playlist to the user based on his or her current mood. The user's local music collection is initially clustered and supported by the emotion of the song, i.e., the mood of the song. This is often calculated by taking into consideration the lyrics of the song. Whenever the user wishes to get a mood-based playlist, the user takes an image of themselves at that instant. This image is subjected to facial detection

and emotion recognition techniques were used to recognize the emotion of the user. The best-matched music for this emotion is then recommended to the user.

### 3. Proposed Work

Our proposed system is an automatic emotion-based music controller. As shown in Figure 1, a camera is used to capture photos, which serve as input for the system. These photos are then passed to the facial emotion detection model, which classifies them into categories such as happy, neutral, sad, or surprised.

Based on the identified expression or emotion, the user is presented with a music playlist containing a selection of music clips. For instance, if the user's expression is classified as happy, they can choose music clips from a playlist curated with joyful songs. A freely available



Figure 1. Proposed Work

dataset is utilized from the internet consisting of 182 photos representing four different emotions: joyful, neutral, surprised, and sad. The system architecture, as depicted in Figure 2, encompasses face detection algorithms integrated with mood detection, leading to the selection of a playlist based on the detected mood.

The first component of the architecture is the face detection module, which utilizes advanced computer vision algorithms to accurately detect and track the user's face in real-time. By analyzing facial features and expressions, the system gathers valuable information regarding the user's emotional cues (Rusia & Singh, 2021). Once the user's face is detected, mood detection algorithms come into play, interpreting the user's emotional state. These algorithms analyze facial expressions and assign a specific mood label, such as happy, sad, angry, or calm, accurately reflecting the user's current emotional state.

Based on the detected mood, the system proceeds to the next step, which involves selecting a suitable playlist of songs from a website or music platform. The system architecture considers a vast collection of songs along

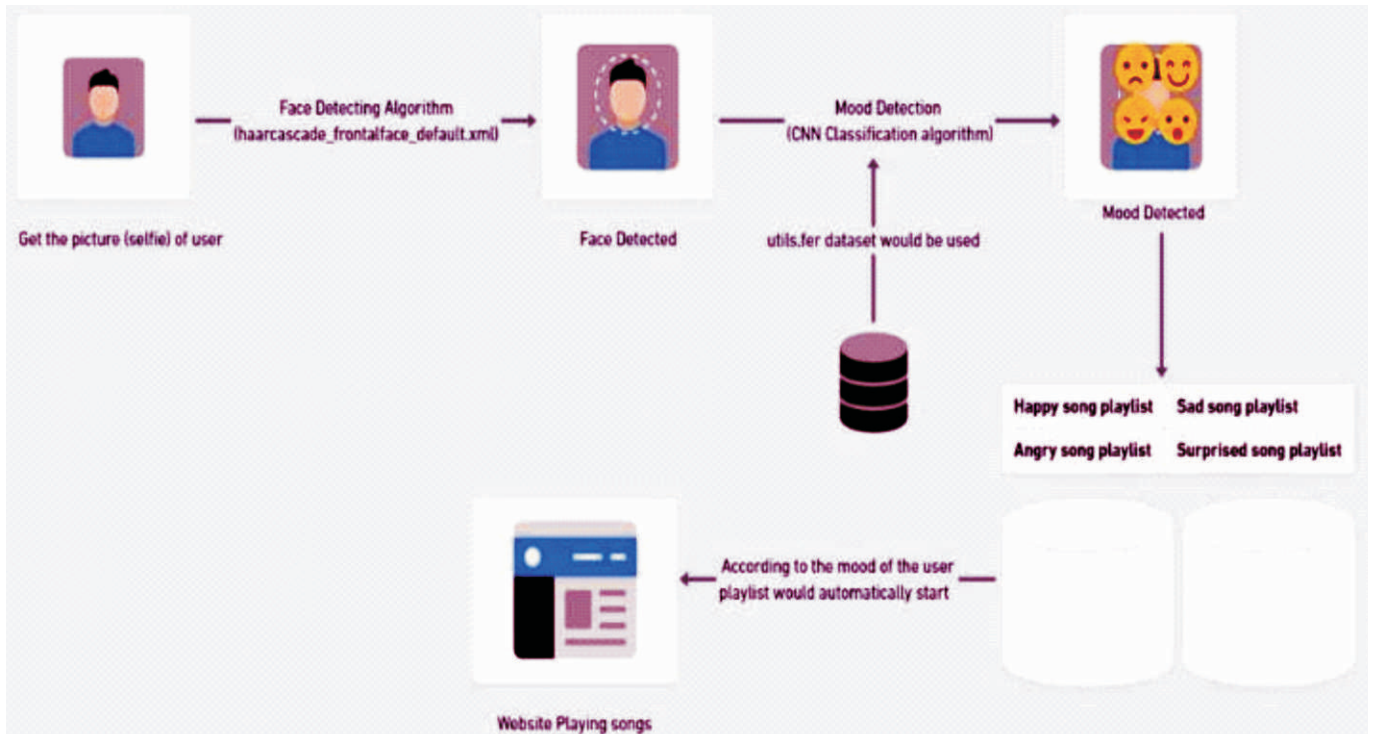


Figure 2. Proposed System Architecture

with their associated mood tags or metadata. By employing machine learning and artificial intelligence techniques, the system matches the user's detected mood with appropriate songs known to evoke similar emotions. This personalized song selection ensures that the user is presented with music that resonates with their current emotional state. The recommended playlist of songs is then played through the integrated website or music platform, allowing the user to enjoy carefully curated music that enhances their mood, relaxation, and overall well-being.

Figure 3 illustrates the proposed methodology and the flow of options used to obtain the optimal approach, resulting in the desired outcomes. The methodology begins with implementing a face detection algorithm using Python. This can be achieved by utilizing libraries such as OpenCV or Dlib, which provide pre-trained models and functions for accurate face detection. These algorithms analyze images or video streams, accurately detecting and locating human faces within them. Once the face is detected, the methodology incorporates a recommendation engine for mood detection. This engine utilizes machine learning techniques to analyze the facial expressions captured by the face detection algorithm. To support mood-based music recommendations, a comprehensive song database is integrated into the methodology. This database comprises a wide range of songs, each associated with specific mood tags or metadata.

The recommendation engine leverages this database to select suitable songs that align with the user's detected mood. This ensures that the recommended songs resonate with the user's emotional experience, providing an immersive and personalized music listening journey. By combining face detection algorithms, a recommendation engine for mood detection, and a diverse song database, this proposed methodology enables a music recommendation system that can understand and respond to the user's emotional state. It offers a unique and tailored music selection, enhancing relaxation, mood regulation, and overall well-being.

The stages involved in this research include image acquisition, face acquisition, emotion detection, enabling the corresponding playlist, dataset creation, feature description, playlist generation, playlist adaptation based on mood, and end-of-day analysis.

### 3.1 Image Acquisition

In facial feature recognition, image acquisition refers to the process of capturing images or video frames that contain human faces. It involves using cameras or other imaging devices to obtain high-quality and well-posed facial images for subsequent analysis and detection of specific facial features or expressions (Fung-Lung et al., 2019). Nowadays, music platforms have made it simple to access a vast library of music. They continuously work on improving music organization and search management to address the issue of choice and make it easier for users to discover new music pieces.

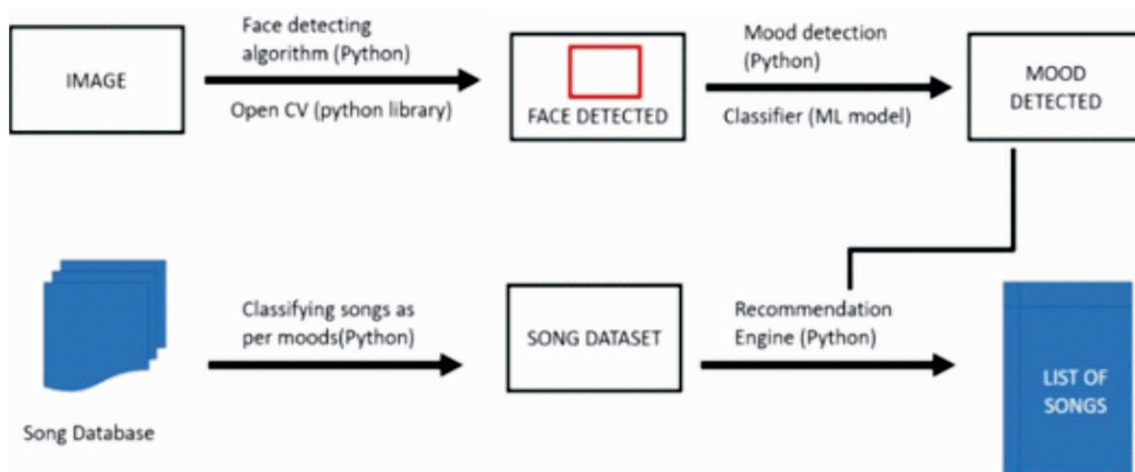


Figure 3. Proposed Methodology

### 3.2 Face Detection

Once the image is obtained, the system initiates the face detection process. The algorithm employs a sub-window approach to scan the image and identify facial features. By systematically examining different regions of the photo, the algorithm searches for patterns and characteristics indicative of a face. To optimize the system's performance, the image is cropped or trimmed to solely encompass the detected face (Meena & Sharan, 2016). This cropping step allows for improved accuracy and efficiency in subsequent facial feature recognition tasks. By isolating the face within the image, the algorithm can solely focus on analyzing the relevant facial features, thereby enhancing the overall performance and effectiveness of the proposed system.

### 3.3 Emotion Detection

Next, to determine the user's sentiment, the methodology utilizes the Principal Component Analysis (PCA) method, which is widely recognized for its effectiveness in facial expression detection. Firstly, a face space is constructed using PCA, where the eigenvectors with the highest eigenvalues are selected to represent the most significant facial features. The acquired image is then projected onto this face space. To detect the user's emotion, scores for each emotion are computed based on the projected image (Giri et al., 2022). These scores reflect the likelihood or intensity of different emotions being expressed in the user's facial expression. The emotion corresponding to the highest score among the calculated emotion scores is chosen as the detected emotion for the image. By employing the PCA method and constructing a face space, the methodology effectively captures and analyzes the user's facial expressions, enabling the detection of their underlying emotions. This approach provides a reliable means to determine the user's sentiment and facilitates personalized music recommendations based on their emotional state.

### 3.4 Enabling the Corresponding Playlist

The proposed system incorporates the display of an appropriate music playlist based on the identified

emotion of the user. Considering the four emotions—pleasant, natural, surprise, and sadness—the system prepares four distinct playlists, each carefully curated with music samples that align with the corresponding emotion. For pleasant emotions, the system activates a classical music playlist, offering soothing and melodic compositions that promote relaxation and enjoyment. In the case of natural emotions, the system triggers a new-age music playlist featuring calming and serene tracks that evoke a sense of tranquility and peacefulness.

To address unpleasant emotions such as surprise and sadness, the system enables a designer music playlist. This playlist is specifically designed to uplift the user's mood by providing emotionally uplifting and motivational tracks that help alleviate negative feelings and restore a more positive state. By tailoring the music playlists to specific emotions, the proposed system ensures that users are presented with music that complements and enhances their emotional experiences. This personalized approach to music selection contributes to a more engaging and therapeutic music listening experience.

### 3.5 Dataset

The dataset contains 48x48 grayscale images of faces displaying different emotions. The faces have been automatically registered so that they are centered and occupy a consistent amount of space in each image. The task is to categorize each face based on the expressions into the following four categories: angry, disgusted, afraid, happy, sad, surprised, and neutral. The training set consists of 28,709 examples, and the public test set consists of 3,589 examples. The song data consists of 686 different English songs with features such as name, album, artist, length, popularity, and other key attributes.

### 3.6 Feature Description

The user begins by opening the app and initiating the music recommendation process. As the app opens, the user's image is captured through the webcam or camera. The app utilizes face detection algorithms to identify and locate the user's face within the image. Once the face is detected, the app proceeds to identify the user's mood based on facial expressions. Using techniques like PCA or

other facial recognition methods, the app analyzes the captured image to determine the user's emotional state.

Based on the identified mood, the app suggests a selection of songs that align with the user's emotional state. These song suggestions are tailored to match the user's mood and provide an immersive music experience. However, if the user does not prefer the suggested song, the app provides a button for the user to explore more options. Clicking the button will display additional songs from which the user can choose a song of their preference, ensuring a personalized music selection that suits their mood and taste.

### **3.7 Generating the Playlist**

The first task in the music recommendation process is to generate a playlist based on the user's preferences. To begin, a random song is suggested to the user from the available music library. A threshold duration is set, specifying the minimum amount of time the user should listen to the suggested song before making a decision. If the user listens to the first song for a duration exceeding the threshold, it indicates a positive response. In this case, 20 similar songs are selected based on the characteristics of the initially suggested song, and a playlist is generated.

However, if the user chooses to skip the suggested song before the threshold duration is reached, the system proceeds to suggest a new random song. This process continues until the user listens to a song for a duration longer than the threshold, indicating their preference. By employing this iterative approach, the system aims to understand the user's music preferences based on their engagement with the suggested songs. It allows for the creation of a playlist that aligns more closely with the user's taste, enhancing their music experience and increasing the likelihood of finding songs they enjoy.

### **3.8 Altering the Playlist Based on the Mood**

Now the playlist is generated, for the user to enjoy music. However, there could be two reasons for the user to skip a song. The first reason could be that they no longer want to listen to the song, either because they dislike it or because their mood has changed. The second reason could be that the user's mood has changed, and the playlist needs

to be adjusted accordingly. To address this, the user can click their picture again when they skip or change the song, and their mood can be identified once more.

If the newly identified mood matches the previous one, it can be concluded that the user's mood has not changed, but they simply no longer enjoy the current song. In this case, the song can be removed from the playlist. However, if the newly detected mood is different from the previous one, the playlist needs to be altered in a slightly different manner. A random song of the newly classified mood can be added to the playlist, and it will be recommended as the next song after the user skips the previous one. Again, the threshold can be checked, and if it is cleared, around 60%–70% of the songs in the current playlist can be replaced with songs that match the newly identified mood.

### **3.9 Analysis**

During the user's sleeping hours, valuable analysis can be conducted to gain insights into their music listening habits and emotional patterns. One analysis involves calculating the average duration (in seconds) of the songs that effectively changed the user's mood. By assessing the time spent listening to songs that had a noticeable impact on the user's emotional state, it becomes possible to understand the duration required for music to significantly influence their mood. Another analysis focuses on observing specific moods and determining which artists' songs the user listens the most. By analyzing the frequency of songs from particular artists associated with particular moods, patterns emerge, providing valuable information about the user's musical preferences and the connection between certain artists and specific emotional states. Additionally, calculating the average duration of the songs listened by the user can provide insights into their preference for shorter or longer music tracks, influencing the recommendations by understanding their preferred song duration.

### **3.10 Deployment**

The deployment phase focuses on making the system available and accessible to users. This involves setting up the necessary infrastructure, including servers,

databases, and hosting environments, to ensure efficient handling of user requests and delivery of recommendations. Rigorous testing and optimization are also performed to ensure reliable and secure system performance. Front-end development is responsible for designing and creating the user interface through which users interact with the system (Ma et al., 2018). This includes designing intuitive and visually appealing interfaces that allow users to input preferences, view recommended playlists, and control music playback. It also involves integrating user feedback mechanisms and features to enhance the user experience and engagement. Together, deployment and front-end development ensure the effective delivery of the personalized music recommendation system to users, providing a seamless and enjoyable music discovery and listening experience.

#### 4. Results and Discussions

As the first model created, `fer_1.h5`, encountered the issue of overfitting. This occurred due to an excessive number of consecutive conv2D layers without max-pooling. Max-pooling layers were removed, because the image size was already small, and applying max-pooling reduced it further to a dimension where training the model became impossible (20 epochs)

An improved model has been created. Max-pooling layers were added again and using padding in the conv2D layers to address the aforementioned issue. Additionally, a batch normalizer has been included. The model was trained on this improved architecture for 20 epochs, but the problem of overfitting still persists. The results of the model, including the highest accuracy achieved and the loss history, are depicted in Figure 4.

The third training model exhibited slight improvements compared to the second model. The issue of overfitting was addressed by eliminating half of the convolutional layers and making adjustments to the number of filters and filter sizes. Despite these modifications, overfitting still persisted. In the fourth model, transfer learning was employed using the pre-trained ResNet50 model, which was initially trained on ImageNet. After 55 epochs of training, the model achieved an accuracy of 37% on the

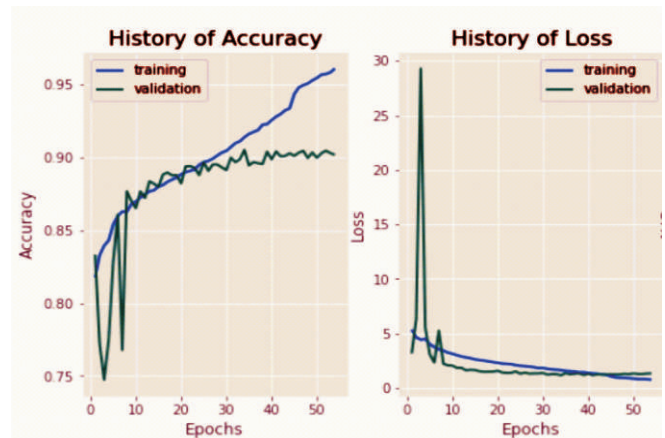


Figure 4. Results of the Model with the Highest Accuracy (92% Accuracy on Test Data)

test data. However, signs of overfitting emerged after the 40th epoch, as indicated by a widening gap between the training and validation lines. Even after training over 50 epochs, the accuracy of the test data remained stagnant at 37%, with an exacerbation of the overfitting issue. The fifth model adopted transfer learning by utilizing the pre-trained VGG16 model from ImageNet for the first 55 epochs. This model achieved an accuracy of 46.6% on the test data; however, overfitting persisted as a problem. In the sixth model, improvements were made over the third model. Data augmentation was introduced to mitigate overfitting, and a dropout layer was added just before the flattened layer. This approach effectively addressed the overfitting problem. After 60 epochs, the accuracy of the test data reached approximately 51%, with similar accuracies for the training and validation data. The seventh model involved transfer learning using the pre-trained VGG16 model, along with data augmentation, dropout layers, and regularization techniques to combat overfitting. However, this model ended up suffering from underfitting. In an effort to reduce underfitting, the dropout layer and regularization were removed in the eighth model, aiming to lean towards overfitting in search of the optimal line. Nevertheless, the model continued to underfit, suggesting that the combination of data augmentation and transfer learning contributed to the underfitting. Nonetheless, the accuracy of this model was slightly better than the previous one. For the ninth model, the

# RESEARCH PAPERS

data augmentation employed in the seventh model was removed, resulting in a performance similar to that of the eighth model.

The tenth model was unsuccessful as the sole utilization of the VGG16 architecture without transfer learning did not yield favorable results. In the 11<sup>th</sup> model, transfer learning was employed in conjunction with data augmentation. Additionally, Gaussian noise, batch normalization, and L2 regularization were added to specific layers after the pre-trained model. This configuration resulted in a training accuracy of 36% and a validation accuracy of 35%.

The 12<sup>th</sup> model was an improvement over the sixth model. It involved doubling the number of Conv2D and max-pooling layers and incorporating regularization techniques. The training accuracy achieved was 46%, while the validation accuracy reached 39%. In the 13<sup>th</sup> model, a conventional CNN architecture was employed, consisting of a Conv2D layer followed by a batch normalization layer. Data augmentation was utilized without relying on transfer learning. This model achieved a training accuracy of 60% and a validation accuracy of 52%.

The 14<sup>th</sup> model combined transfer learning, data augmentation, regularization, and batch normalization. With this approach, a training accuracy of 60% and a validation accuracy of 59% were attained. For the 15<sup>th</sup> model, the top 10 layers of the pre-trained VGG16 model was frozen, and several dense layers with different regularizers were added. This model performed the best, achieving a training accuracy of 96.6% and a validation accuracy of 90%. The predictions from this model, which demonstrated the highest accuracy, can be observed in Figure 5.

Figure 6 presents a summary and performance evaluation of all the models used in the personalized music recommendation system. The evaluation is based on several metrics, including the model description, number of epochs, training accuracy, and validation accuracy. Each model is characterized by its specific architecture and parameters. The number of epochs represent the number of iterations the model undergoes



Figure 5. Predictions by the Model with the Most Accuracy

Model No.	Model Description	Epochs	Training Accuracy	Validation Accuracy	Problem
1	Classic CNN structure with two many convolutional layers	20	0.892	0.5475	Overfitting
2	Added more pooling layers in the model 1 and also added batch normalizer	20	0.9523	0.5695	Overfitting
3	Removed half of the convolutional lauers from model to overcome overfitting	20	0.9256	0.5302	Overfitting
4	Transfer learning of the Resnet50 model trained on imagenet	105	0.4356	0.3742	Low Accuracy
5	Transfer learning of the VGG16 model trained on imagenet	55	0.9691	0.4514	Overfitting
6	The 6th model is the upgradation of 3rd model. We are adding a layer of augmentation and dropout layer.	60	0.5234	0.5138	Low Accuracy
7	The 7th model is transfer learning (pre-trained VGG16 model) with data augmentation, dropout layers and regularization.	55	0.33	0.344	Underfitting
8	To reduce underfitting, we decided to remove the dropout layer and regularization in the 8th model and move to overfit to find the best line	55	0.4319	0.3998	Underfitting
9	We removed the data augmentation from the 7th model and got the same result as the 8th model	55	0.3791	0.4085	Underfitting
10	This model was a failed attempt where we used only the VGG16 architecture without transfer learning.	-	-	-	-
11	We used transfer learning with data augmentation and added gaussian noise, batch normalization and L2 regularizer to some layers after the pre-trained model	55	0.3538	0.353	Underfitting
12	This model is the updation of 6th model by doubling the number of conv2D and max-pooling layers and also adding regularization	55	0.4253	0.3912	Underfitting
13	This model is a typical CNN architecture with a conv2D layer followed by a batch normalization layer, data augmentation and no transfer learning	100	0.5875	0.5463	Low Accuracy

Figure 6. Summary and Performance of All the Models

during training. The training accuracy reflects the model's performance on the training data, while the validation accuracy assesses its performance on unseen validation data.

Throughout the evaluation process, potential challenges and problems encountered by each model are identified. This may include issues such as overfitting,

underfitting, convergence problems, or difficulties in handling large datasets. By analyzing the model descriptions and performance metrics, adjustments and optimizations can be implemented to enhance the overall accuracy and effectiveness of the personalized music recommendation system.

## Conclusion

The method used in this research work has approaches for emotion detection and successfully distinguishes between happy, sad, natural, and astonished emotions. Based on the assessed emotion, the system generates a playlist for the user containing music that aligns with their mood. Extensive research has demonstrated the ability to leverage facial expressions for emotion detection and provide personalized music recommendations. The research highlights the potential of facial feature recognition technology for creating more immersive and tailored music recommendation systems. With further advancements and optimizations, this technology can revolutionize the way people discover and engage with music, opening up new avenues for emotional connection and enjoyment through personalized music selections.

## Acknowledgment

Our utmost gratitude to Dr. Sheetal Rathi, our guide, for her invaluable support and guidance at every step and our heartfelt appreciation goes to all the mentors and faculties of Thakur College of Engineering and Technology, who played a pivotal role in helping us accomplish this work.

## References

- [1]. Donald, B. (2014). *Music Recommendation System: "Sound Tree"*. Retrieved from <https://www.slideserve.com/birch/music-recommendation-system-sound-tree>
- [2]. Fung-Lung, L., Nycander-Barúa, M., & Shiguihara-Juárez, P. (2019, August). An image acquisition method for face recognition and implementation of an automatic attendance system for events. In 2019 *IEEE XXVI International Conference on Electronics, Electrical Engineering and Computing (INTERCON)* (pp. 1-4). IEEE. <https://doi.org/10.1109/INTERCON.2019.8853603>
- [3]. Gilda, S., Zafar, H., Soni, C., & Waghurdekar, K. (2017, March). Smart music player integrating facial emotion recognition and music mood recommendation. In 2017 *International Conference on Wireless Communications, Signal Processing and Networking (WISPNET)* (pp. 154-158). IEEE. <https://doi.org/10.1109/WISPNET.2017.8299738>
- [4]. Giri, S., Singh, G., Kumar, B., Singh, M., Vashisht, D., Sharma, S., & Jain, P. (2022, April). Emotion detection with facial feature recognition using CNN & OpenCV. In 2022 *2<sup>nd</sup> International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 230-232). IEEE. <https://doi.org/10.1109/ICACITE53722.2022.9823786>
- [5]. Hemanth, P., Adarsh, Aswani, C. B., Ajith, P., & Kumar. V. A. (2018). EMO PLAYER: Emotion based music player. *International Research Journal of Engineering and Technology*, 5(4), 4822-4827.
- [6]. James, H. I., Arnold, J. J. A., Ruban, J. M. M., Tamilarasan, M., & Saranya, R. (2019). Emotion based music recommendation system. *Emotion*, 6(3), 2096-2101.
- [7]. Li, K., Jin, Y., Akram, M. W., Han, R., & Chen, J. (2020). Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy. *The Visual Computer*, 36, 391-404. <https://doi.org/10.1007/s00371-019-01627-4>
- [8]. Ma, S., Zhang, X., Wang, S., Zhang, X., Jia, C., & Wang, S. (2018). Joint feature and texture coding: Toward smart video representation via front-end intelligence. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(10), 3095-3105. <https://doi.org/10.1109/TCSVT.2018.2873102>
- [9]. Mahadik, A., Milgir, S., Patel, J., Jagan, V. B., & Kavathekar, V. (2021). Mood based music recommendation system. *International Journal of Engineering Research & Technology*, 10(6), 553-559.
- [10]. Meena, D., & Sharan, R. (2016, December). An approach to face detection and recognition. In 2016 *International Conference on Recent Advances and Innovations in Engineering (ICRAIE)* (pp. 1-6). IEEE.

<https://doi.org/10.1109/ICRAIE.2016.7939462>

[11]. Raut, N. (2018). *Facial Emotion Recognition using Machine Learning* (Master's thesis, San José State University). <https://doi.org/10.31979/etd.w5fs-s8wd>

[12]. Rusia, M. K., & Singh, D. K. (2021). An efficient CNN

approach for facial expression recognition with some measures of overfitting. *International Journal of Information Technology*, 13, 2419-2430. <https://doi.org/10.1007/s41870-021-00803-x>

---

## ABOUT THE AUTHORS

Harsh Mishra, Department of Computer Engineering, Thakur College of Engineering & Technology, Mumbai, India.

Deep Kothari, Department of Computer Engineering, Thakur College of Engineering & Technology, Mumbai, India.

Prajwal Jaiswal, Department of Computer Engineering, Thakur College of Engineering & Technology, Mumbai, India.

Sheetal Rathi, Department of Computer Engineering, Thakur College of Engineering & Technology, Mumbai, India.

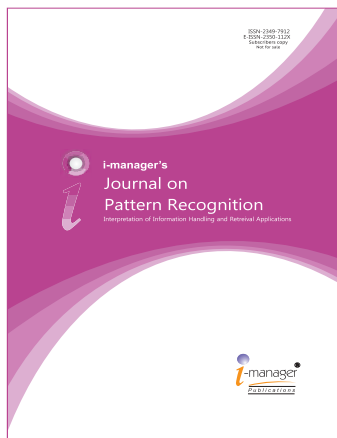
# i-manager's Journal on Pattern Recognition (JPR)

Published by i-manager Publications, India.

<http://www.imanagerpublications.com>

**Editor-in-Chief: Dr. G.R. Sinha**  
*Professor*  
*Myanmar Institute of Information Technology*  
*Mandalay, Myanmar.*

## Call For Papers for 2023



i-manager's Journal on Pattern Recognition provides innovative ideas for the researchers, academics, professionals and high level engineering students to bring together the current developments in the applications of medicine, robotics, and remote sensing by satellites. Pattern recognition is characterized by automated searches over a large number of observations and huge combinatorial spaces and is concerned with the automatic discovery of regularities in data through the use of computer algorithms. i-manager's Journal on Pattern Recognition will help to update the researchers in the upcoming Pattern Recognition schemes.

i-manager Publications is a leading publishing house specialized in publishing Scientific, Technology, Education and Management Journals. The credibility and impact of these publications in the Indian economic and academic environment reinforced the founding vision of i-manager Publications: promoting technology education and present international research worldwide.

### Why Publish with us?

- No Publishing Fee
- Abstracting & Indexing in leading databases
- Double-blind Peer Review
- Highly qualified Editorial Board
- Publishing Journals since 2004
- Maximum Publicity in Social media
- Rapid Publication: 2-3 months
- Full archive available from Volume 1 onwards

### Abstracting / Indexing



Submission email : [submissions@imanagerpublications.com](mailto:submissions@imanagerpublications.com)

## Overall Topics Covered:

- ◆ Pattern Recognition in Time Dependent Data
- ◆ Signal Processing for Pattern Recognition
- ◆ Fingerprint and Character Recognition
- ◆ Face and Object Detection and Recognition
- ◆ Pattern Recognition in Health Informatics Applications
- ◆ Character, Text and Language Recognition
- ◆ Classification and Clustering Methods
- ◆ Statistical Techniques and Pattern Analysis
- ◆ Target Recognition Systems
- ◆ Signal, Video and Speech Recognition
- ◆ Object Recognition and Tracking
- ◆ Pattern Recognition in Biosciences Applications
- ◆ Pattern Recognition in Computational Science Applications
- ◆ Syntactical and Structural Pattern Recognition
- ◆ Biometric Recognition
- ◆ Predictive Modelling Applications
- ◆ Large Data Forecasting Applications
- ◆ Satellite Image Processing
- ◆ Hypergraphy Imaging
- ◆ Asynchronous Pattern Matching
- ◆ Pattern Matching Techniques
- ◆ Feature Selection and Extraction
- ◆ Neuromorphic Systems
- ◆ Space Science Applications using Pattern Recognition
- ◆ Domain Adaptation
- ◆ Man and Machine Diagnostics
- ◆ Person Identification and Industrial Inspection
- ◆ Document Image Enhancement
- ◆ Remote Sensing Applications
- ◆ Brain Signal Pattern Recognition
- ◆ Face Animation
- ◆ Forgery Detection System
- ◆ Cyber Deception using Artificial Intelligence
- ◆ Data Mining and Machine Learning Methods
- ◆ Advanced Reinforcement Learning Techniques
- ◆ Pattern Recognition in Machine Learning
- ◆ Deep Learning -Control Self-Driving Cars (and Other Types of Robots)
- ◆ Image Segmentation
- ◆ Sentiment Analysis
- ◆ Applications of pattern recognition in Genetic Algorithm
- ◆ Deep Neural Networks
- ◆ Support Vector Machine
- ◆ Computer Vision and Big Data Learning
- ◆ AI-based Pattern Recognition Methods
- ◆ Human Behavior Recognition
- ◆ Urban 3D Reconstruction
- ◆ Visual Odometry

### Features

Articles, Research Papers, Review Papers.

### Periodicity

January-June, July-December

### Target Audience

Researchers, consultants, progressive circle of readers from premier education bodies, Academicians and leading corporate etc.



<https://www.facebook.com/imanPatternRecognition/>

<https://scholar.google.co.in/citations?hl=en&user=1ZLQEZcAAAAJ>

<https://twitter.com/imanagerpub>



3/343, Hill view, Town Railway Nager, Nagercoil  
Kanyakumari Dist. Pin-629 001.  
Tel: +91-4652-231675, 232675, 276675

e-mail: [info@imanagerpublications.com](mailto:info@imanagerpublications.com)  
[contact@imanagerpublications.com](mailto:contact@imanagerpublications.com)  
[www.imanagerpublications.com](http://www.imanagerpublications.com)