Neurocomputing 617 (2025) 129018

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

An adaptation of hybrid binary optimization algorithms for medical image feature selection in neural network for classification of breast cancer

Olaide N. Oyelade [a,b,*], Enesi Femi Aminu [c], Hui Wang [a], Karen Rafferty [a]

[a] School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, BT9 5BN, United Kingdom
[b] Department of Engineering, Computer and Mathematics, University of Chichester, United Kingdom
[c] Department of Computer Science, Federal University of Technology Minna, Nigeria

ARTICLE INFO

Keywords:
Binary optimizer algorithms
Digital mammography
Convolutional neural network
Image feature selection
Metaheuristic algorithms
Breast cancer
Medical image abnormalities

ABSTRACT

The performance of neural network is largely dependent on their capability to extract very discriminant features supporting the characterization of abnormalities in the medical image. Several benchmark architectures have been proposed and the use of transfer learning has further made these architectures return good performances. Study has shown that the use of optimization algorithms for selection of relevant features has improved classifiers. However continuous optimization algorithms have mostly been used though it allows variables to take value within a range of values. The advantage of binary optimization algorithms is that it allows variables to be assigned only two states, and this have been sparsely applied to medical image feature optimization. This study therefore proposes hybrid binary optimization algorithms to efficiently identify optimal features subset in medical image feature sets. The binary dwarf mongoose optimizer (BDMO) and the particle swarm optimizer (PSO) were hybridized with the binary Ebola optimization search algorithm (BEOSA) on new nested transfer functions. Medical images passed through convolutional neural networks (CNN) returns extracted features into a continuous space which are piped through these new hybrid binary optimizers. Features in continuous space a mapped into binary space for optimization, and then mapped back into the continuous space for classification. Experimentation was conducted on medical image samples using the Curated Breast Imaging Subset of Digital Database for Screening Mammography (DDSM+CBIS). Results obtained from the evaluation of the hybrid binary optimization methods showed that they yielded outstanding classification accuracy, fitness, and cost function values of 0.965, 0.021 and 0.943. To investigate the statistical significance of the hybrid binary methods, the analysis of variance (ANOVA) test was conducted based on the two-factor analysis on the classification accuracy, fitness, and cost metrics. Furthermore, results returned from application of the binary hybrid methods medical image analysis showed classification accuracy of 0.8286, precision of 0.97, recall of 0.83, and F1-score of 0.99, AUC of 0.8291. Findings from the study showed that contrary to the popular approach of using continuous metaheuristic algorithms for feature selection problem, the binary metaheuristic algorithms are well suitable for handling the challenge. Complete source code can be accessed from: https://github.com/NathanielOy/hybridBinaryAlgorithm4FeatureSelection

1. Introduction

In recent times, medical imaging has continuously played a significant role because it has been applied to different clinical purposes. The medical imaging procedures include computed tomography, ultrasound, magnetic resonance imaging, X-ray, and positron emission tomography. These imaging helps to make evaluate or diagnose certain organs. For instance, mammography is useful for breast cancer diagnosis, colonoscopy for bowel entrails checking, and magnetic resonance imaging (MRI) for almost every internal structure in the human body. These helps to mirror variations because of medical situations from the structural, functional, or metabolic levels of the organs [1]. Computer-aided approaches for supporting the processing of these imaging have demonstrated impressive performance. Research development in the use of neural network models is top on this list of computational solutions. These have significantly contributed to the improvement of medical image analysis in terms of diagnosis, treatment planning, or characterization of abnormalities [2]. Typical examples

* Corresponding author at: School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, BT9 5BN, United Kingdom.
E-mail address: o.oyelade@chi.ac.uk (O.N. Oyelade).

https://doi.org/10.1016/j.neucom.2024.129018
Received 28 June 2024; Received in revised form 2 November 2024; Accepted 21 November 2024
Available online 28 November 2024

include the CNNs, which have been employed severally in the aspect of analysis of medical images datasets [3]. This use of CNN have addressed the limitations of other feature maps approaches [4], so that benchmark CNN architectures (such as VGGNet, GoogleNet and ResNet) have now been trained on different medical image modalities (X-Ray and CT-Scan) [5] and are available for knowledge distillation and transfer learning. However, not without pitfalls such as its effectiveness in handling the issue of noise, the volume and complex natured of data input required [6]. To address some of these drawbacks, quite a few neural network architectures have been suggested and experimented with, even the reuse of pre-trained models otherwise known as transfer learning. Similarly, transformer neural network equally assists in diagnosis of disease and other clinical objectives [7] and have been successful in image classification and segmentation tasks [8]. In some studies, researchers developed autonomous learning strategies to carefully analyze medical images [9]. However, these are not still without human intervention owing to the complex nature of medical data [10]. More so, because of inadequate availability of annotated data, training segmentation models for medical images have become difficult [11]. Regardless of all these techniques and methodologies, accuracy of expected results is always reduced attention is not given to feature optimization strategy. The focus of the application optimization strategy is diverse. This might include neural network parameters, and even the image data itself. Optimization strategies help to eliminate bottlenecks which clamp down on classifier. The difficulty of identifying even subtle abnormalities in digital mammography is challenging due to their complex nature and high dimensionality. As a result of this, efforts are consistently aimed at improving analysis of digital mammography images for characterization of abnormalities possibly for diagnosis of breast cancer.

The main goal of most deep learning models applied to medical image analysis is to increase classification accuracy, reduce false positive rates and false negatives rates. Recent efforts have extended this goal to localizing the abnormalities and as well staging the disease. In this instance, the classification of breast imaging into benign, malignant, or normal, largely depends on the existence of abnormalities such as architectural distortions, asymmetries, masses, and micro calcifications [12]. Over time, the use of optimization algorithms has been focused on continuous metaheuristic algorithms with little or no research consideration to ascertain the viability of binary optimization algorithms to addressing this challenge in medical image analysis. Although several works [13,14] have investigated the use of binary optimization algorithms to the task of feature selection in text-based datasets, we found no attempt at applying this to medical image samples. Advances in the use of the hybrid binary optimizer algorithm on text-based datasets has evolved also as the future [15]. This position is affirmed by [16] which stated that the fundamental deficit of the single binary optimization algorithm would limit the optimal performance of coherent feature subset thereby leads to distortion of classification accuracy. Similarly, [17] reaffirmed that the basic inherent deficit of single binary optimizer can potentially affects the best optimal search solution. Therefore, it is important to note that the binary optimization algorithm, and the its corresponding hybrid variants have proven useful for better classification accuracy. This notwithstanding, there is little in literature to demonstrate any attempt for the adaptation of binary or hybrid metaheuristic optimizers for medical image feature selection for improving classification accuracy with CNNs. Metaheuristic algorithms work by normally depending on several instances in a population such as agents in particles for particle swarm optimization (PSO), chromosomes in genetic algorithm (GA), individuals in Ebola optimization search algorithm (EOSA) [18], and red deers for red deers algorithm (RDA), which heavily engage on searching for near-global or global optimum solution [19]. Different variants of hybridized or adaptive metaheuristic have been equally proposed [20]. Notwithstanding, the binary optimizers which leverages a binary solution space has the potential to optimize the weights and structure of netural network models, and as well find descriminant features [21].

The challenge of finding the most suitable and optimal feature subsets useful to help neural classifiers achieve optimal accruacy have motivated the use of continouse metaheuristic algorithms for reduction of search space with neural networks. For instance, the works of [22] demonstrates the possibility of using continuous optimization algorithm namely Q-Learning Embedded Sine Cosine Algorithm (QLESCA) for extraction of optimal feature. In a similar application of the COVID-19 samples, [23] investigated the role of Manta Ray Foraging based Golden Ratio Optimizer (MRFGRO) a continuous metaheuristic algorithm for feature selection. To emphasize the wide application of continuous optimization algorithms to feature selection in medical imaging, [24] and [25] have reported the impressive performance of particle swarm optimization (PSO), cuckoo search optimization (CSO), the firefly algorithm (FFA), the bat algorithm (BA), flower pollination optimization (FPO), whale optimization algorithm (WOA), marine predator algorithm, atom search optimization algorithm (ASOA), Harris hawks optimization algorithm (HHOA), butterfly optimization algorithm (BOA), and grey wolf optimization (GWO). Also, [26] have demonstrated that an immunity based EOSA continuous metaheuristic algorithm can aid in the finding of the best feature subsets of digital mammography. Still on image feature selection, the Fitness-based Memory updated-Crow Search Algorithm (FMCSA) have been used with CNN in [27]. In [28], authors combined the Modified Cosine Similarity (MCS) algorithm with CNN for a Content-Based Medical Image Retrieval (CBMIR) for feature selection. On the other hand, [29] have shown that even the difficult challenge of medical image segmentation can be improved using Kepler optimization algorithm (KOA) and neural network. These works can be extended to even recent continuous optimization algorithms such Coati optimization algorithm (COA) and susceptible-infected-removed optimizer (SIRO) [30]. The use of continuous metaheuristic algorithm has proven useful in feature selection and in formulation optimal combination of hyper-parameter. Furthermore, this is also useful for optimizing weights, number of layers, number of neurons, multimodal image fusion [31–36], image enhancement [37], and learning rate [38]. Reformulation of the optimizer in the binary metaheuristic algorithm nature can effectively support the task of feature selection for enhanced performance leading to high classification rate [24]. This reformulation of the continuous problem space into a binary search space for medical image feature optimization remains unaddressed.

What then is the best approach to formulate the problem of medical image feature selection from continouse space to binary space for improved classfication accuracy? To address this research gap, this study is aimed at adapting a new hybrid binary metaheuristic algortihm namely the EOSA to reforumulate the problem defintion for feature selection from the continouse search space to bianry search space. Continuous optimization algorithms have mostly been used though it allows variables to take value within a range of values. The advantage of binary optimization algorithms is that it allows variables to be assigned only two states, and this have been sparsely applied to medical image feature optimization. This study therefore proposes hybrid binary optimization algorithms to efficiently identify optimal medical features subset from CNN detected features. The methodology is focused on the removal of irrelevant features subsets which have been reported to potentially impede the accuracy of the learning model and prevent information loss as well [14]. To achieve this, a new nested transform function is defined to transform the search space of the feature set into a binary search space for flagging candidate features that requre selection. Secondly, a hybrid of three binary optimzers were proposed with the binary Ebola optimization search algorithm (BEOSA) being the base

method of binary dwarf mongoose optimizer (BDMO), and the binary particle swarm optimizer (BPSO). These resulted in two new binary hybrid methods namely HBEOSA-DMO-NT and HBEOSA-PSO-NT which used the new transform function for optimizing a binary search space. Moreover, binary optimizer using the traditional transform functions were also proposed in the study and named HBEOSA-DMO and HBEOSA-PSO. Thirdly, a novel technique is proposed which sits between the fully connected layer of the neural network and the hybrid binary optimizers to transform the feature set into a binary search space with no loss of real-life representation of the medical image. Experiments are applied to check the potential and limitation of the method proposed [39].

The following highlights the contribution of this study:

a. Designed a novel adaptation of the medical image feature representation in continuous space to a binary search space for the hybrid binary optimization strategies.
b. Proposed two hybrid binary optimization algorithms namely HBEOSA-DMO and HBEOSA-PSO.
c. Investigated the influence of the new nested function on the two proposed hybrid methods so that four variants were derived namely HBEOSA-DMO HBEOSA-DMO-NT HBEOSA-PSO and HBEOSA-PSO-NT.
d. Comparatively investigated the capability of the four binary optimizers with other recent binary methods.
e. Experimentally studied the impact of the new hybrid binary optimizers on improving classification accuracy of applying CNN to digital mammography.

The remainder of this paper is organized as follows. Section 2 accounted for literature related of both continuous and binary optimization algorithms specifically. Section 3 presents the proposed methodology, which includes the novel mathematical model and algorithmic representation, the hybrid binary optimization algorithm, and the mapping mechanism from continuous to binary search space. The discussion for the experimentation setup and datasets was accounted for in Section 4. While Section 5 presents the extensive results and discussion, Section 6 highlights the summary and conclusion of the work.

2. Related works

It has been established that high dimensionality datasets potentially affect the accuracy of the learning algorithms. To proffer solution to this shortcoming, feature selection approaches have been identified as promising. To this end, [40] proposed a wrapper technique premised on the iterative enhancement capacity of the weighted superposition attraction (WSA) algorithm. It was reported that the WSA algorithm claimed superiority over seven other famous approximate algorithms; therefore, concluded to be efficient FS approach. However, the authors concluded that the proposed algorithm could not efficiently handle the problem high dimensional data hence requires improvement.

Furthermore, the research of [15] proposed a new binary coronavirus disease optimization algorithm to tackle the challenge of feature selection in a high dimensional data. The proposed FS algorithm was validated on 26 benchmark datasets; and the results obtained were compared with nine wrapper-based methods. From the experiment, the proposed algorithm shows superiority over the recent nine FS methods. The proposed algorithm used k-nearest neighbor (KNN) classifier although the authors suggested that to further proof the reliability of the algorithm; other classifier(s) is/are encouraged. Moreover, as part of their measure to further improve on the algorithm, they suggested hybridization with other optimizers such as simulated annealing algorithm. Therefore, most literature established that hybridized multi staged binary wrapper-based FS strategy could be a robust pathway to deal with the problem of FS especially among high dimensionality datasets in future. Generally, metaheuristic algorithms (MA) have demonstrated capacity to solve the problem of FS. However, based on the survey carried out by [41]swarm intelligent based MA have consistently received the most substantial attentions by researchers from 2000 to 2022 at the expense of others. In fact, the survey work of [42] unequivocally suggested that interested researchers in this field could advance research to develop new or improve metaheuristic algorithm for evolution and human inspired based algorithms. Thus, the need to advance work on binary FS but using one of the human inspired algorithms that is adaptive Ebola optimization search algorithm. In as much as the existing optimization algorithms have achieved some positive results in high dimensional data space, there is always room for improvement to obtain the best optimal solution within the given search space. In view of this development, the work of [43] proposed an enhancement to a physics-based optimization method called Archimedes optimization algorithm by adding a novel variable depends on the step distance of each individual population. The variable is for the upper and lower limit of a given search space.

The research of [44] employed a two-stage hybridized feature selection algorithm to detect in advance the associated defects of software module. The first stage of the FS was complementing the unique attributes of whale optimization algorithm with annealing simulated algorithm. A reduction in features was achieved however, there are overlapping features among the feature subset; hence the need for the second stage of FS. This stage equally employed the hybrid of kernel extreme learning machine and convolutional neural network. The multi hybrid FS algorithm is indeed promising considering the reported results; however, computational costs are incurred.

The need to improve the existing FS algorithms for high dimensional data is indispensable. To this effect, [45] proposed a two-stage hybrid ant colony FS algorithm capable to deal with high dimensional dataset. Results obtained that the proposed advanced algorithm reported a reduction of the algorithm from falling into a local optimum and execution time considering the 11 high dimensional datasets. However, some redundant and non-redundant features were still found on the training and testing set of the data respectively. Consequently, the performance of some selected feature subsets of the dataset decreases. Therefore, the authors anticipate a solution to this deficiency by considering reviewing the fitness function. Similarly, the research of [46] improved the particle swarm optimization (PSO) algorithm. The improvement leverages on three new approaches considering the new proposed binary PSO variant, sticky binary PSO to increase the evolutionary result. The three approaches include feature ranking information and is proposed for a novel initialization technique. Decreasing the search space by using a dynamic bit's masking scheme is proposed as the second approach. Lastly, to reduce the sudden convergence issue, a modification process showing genetic operations on the optimal location of the improved sticky binary PSO is deployed. The improved PSO significantly decreases the computational time compared with the benchmark PSO based methods considering the results obtained from 12 datasets gotten from University of California Irvine (UCI) repository. However, computation time is still an issue to contend with. This is because the framework assumed by the proposed algorithm demands a huge chunk of time for fitness function estimation.

Furthermore, [16] proposed a novel binary alternate of grasshopper optimization algorithm for selection of features subset. The enhancement lies on sets the locations of grasshoppers with binary values (0 and 1) and uses modest operators to keep abreast the location. Based on twenty datasets of various sizes from UCI repository, the algorithm is tested along with other five popular optimization algorithms for FS problem; and the result is promising. However, the fundamental deficit

of the single binary optimization algorithm would limit the optimal performance of coherent feature subset thereby leading to distortion of classification accuracy. To solve the consistent optimization problem in data mining because of exponential growth of data, interested researchers in the field of FS have proposed several hybrid FS techniques; but mostly deficient to handle high dimensional type of data. Hence, [47] proposed a hybrid sine cosine Harris hawk optimization (HHO) algorithm to solve the deficiency of exploration and exploitation associated with the traditional HHO. The proposed algorithm based on numerical optimization test case, and low and high dimensions of 16 datasets outperformed the individual algorithm of sine-cosine, HHO, and other remarkable FS hybrid algorithms. However, one of the possible areas of interventions as rightly suggested by the literature is the extension of the proposed hybrid algorithm by adding evolutionary operators or other viable strategies for extensive optimization activities. Similarly, [48] equally proposed the hybridization of sine-cosine (SCA) and GA algorithms as another viable approach to solve the challenge of FS among various degree of dimensional datasets. The rationale behind the proposed method is to implant GA into SCA to serve as an internal function to enhance the exploration capability of SCA. The proposed method was evaluated by considering the standard metrics. The result was compared with the traditional sine cosine algorithm, ant lion, and particle swarm optimization algorithms based on the 16-dataset obtained from UCI repository where the proposed method outperformed the other methods. However, the inherent challenge of the fitness function to examine each feature subset in the SCA's search space is still an issue to contend with.

In the research work of [42] a population decreasing strategy is invoked to the new binary gaining-sharing knowledge human based feature selection algorithm that keeps the algorithm from falling into local optima and early convergence. Experiment for the proposed method was carried out on 22 UCI benchmark datasets, and the results were compared with some of the literature verified metaheuristic algorithms. Even though the results obtained were promising, the literature further noted an area of improvement in the FS algorithm by employing chaotic maps. [49] deployed hybrid approach of FS to classified micro-array data. Technique for order preference by similarity to ideal solution as filter-based approach was used, while binary Jaya algorithm with time-changing transfer function was deployed as the wrapper-based method. The hybrid approach was experimented against four other existing popular hybrid FS methods on 10 benchmark datasets of the domain; and the results in terms of classification accuracy and computational speed outweigh the compared four methods. To solve the problem of feature selection for high dimensional data [17] proposed a novel method by transforming the initial dataset through principal component analysis and fast independent component analysis as hybrid data transformation approach. The authors further employed the binary salp swarm algorithm on the transformed dataset to find the optimal features subset. The proposed method was validated on 15 benchmark datasets; and the results obtained were compared with five other binary-based optimizers. From the experiment, an outstanding performance was recorded by the proposed method. However, the basic inherent deficit of single binary optimizer can potentially affect the best optimal search solution. Therefore, as part of the future works, the authors recommended a multi-label feature selection strategy.

The problem associated with most metaheuristic algorithms trapping into local optimal and slow or premature convergence rate when apply for feature selection among high dimensionality data is equally attracting research attentions. This is why the literature of [50] employed a strategy known as chaos scheme added to the search space of vortex search algorithm to attain the global optimal solution and fast convergence rate for feature selection process. The proposed FS method, which is based on chaos (tent) map, was validated on 24 standard datasets; and the results obtained were compared with eight other existing optimizers. From the experiment, the enhanced algorithm claimed superiority over others. However, the literature suggested further improvement on the algorithm by considering integrating it into metaheuristic algorithms. For discrete FS challenges that have discrete search space, most of the conventional metaheuristic algorithms are deficient to handle it; thus, the need for binary based optimizers. This is what led to the motivation for the work of [51] who proposed an enhancement on the binary-based emperor penguin optimizer. Transfer function and position updating strategy are the two-cardinal focus of the improvement. The proposed method was evaluated against seven newly designed optimization algorithms on 25 popular datasets. The results are promising as it showed superiority over the tested seven binary optimizers, however, it recorded poor performances on some of the multimodal functions. The research of [52] argued that most of the genetic based FS algorithms requires tuning parameters for optimal performance; however, tuning the values of the parameters is a tedious task in FS process. To avert this challenge, the authors proposed a new wrapper-based FS method, which only requires population size, and an amount of generation as controlling parameters to obtain a better feature subset from the given dataset. The new method is christened binary teaching learning-based optimization algorithm. The method produced higher accuracy against the initial features based on the given breast cancer dataset. Five classifiers were employed as objective functions. However, the proposed method is only evaluated against itself, that is, feature subset obtained from the proposed method against the original features of only one dataset. Therefore, the robustness of the method is not sufficiently justified against some literature proven optimization algorithms. The authors concluded to improve on the method by developing a novel hybrid algorithm.

The research work of [53] to advanced a new hybrid method that combined the binary modifications of DMO and simulated annealing algorithm christined as binary dwarf mongoose simulated annealing optimizer (BDMSAO). Furthermore, when it comes to coordinated and flexible mechanism to improved local and global exploration capabilities, PSO posses this strength. PSO is also suitable to indicates the fitness function and to define fixed bounds on the optimized parameters. to propose an enhanced version of binary PSO to solve a domain problem. Finally, BEOSA is another new optimization algorithm, which has demonstrated good performance in proffering solution to the problem of binary optimization [54]. Ebola virus disease and its propagation model inspire this algorithm. Thus, this paper aims to take advantage of these three remarkable algorithms by hybridizing them for optimal feature subsets and consequently, good classification accuracy for the domain under consideration.

Several other studies have demonstrated the relevance of using metaheuristic algorithms to finding suboptimal feature space required for finding better classification accuracy. For instance, the work in [55] showed that the problem of imbalanced data samples and feature selection can be combined. The combined problem space was solved using metaheuristic based on Random-SMOTE (RSMOTE) using the self-adaptive Bat algorithm for feature selection on X-ray data samples. Similarly, the use of Gaussian-based Moth-Flame Optimization (MFO) algorithm has been proposed for solving feature selection problems in text-based datasets. The resulting metaheuristic algorithm namely GMSMFO successfully reduced feature space and as well selected suboptimal set necessary for improving classification accuracy [56]. In the same vein, the use of text-based dataset on metaheuristic algorithms in feature selection process have been reported in [57]. Authors demonstrated that the adaptive PSO with leadership learning (APSOLL) can support the performance of classifiers through an effective feature selection mechanism. Their study compared some well-known metaheuristic algorithms with their approach and confirmed that the

Table 1
A summary of review on similar studies discussed in this section.

Ref, Year	Approach	Limitation in the context of this study
[56], 2023	Investigated the use of an improved version of Gaussian-based Moth-flame optimization algorithm for feature selection in images using machine learning algorithms.	The feature selection process is based on machine learning algorithms as against the deep learning method applied in this study.
[58], 2022	Combines Levy Adapted SLnO (LA-SLnO) optimizer with deep machine learning-based data classification which uses Tri-Kernel principal component analysis (TK-PCA) for feature extraction.	The feature extraction method is limiting in its ability to find relevant features for the optimizer to select from to the classifier.
[57], 2022	Feature selection problem was addressed using adaptive PSO with leadership learning (APSOLL).	The approach is based on text-based datasets and not image datasets for feature selection problems. In addition, the study is focused on the use of continuous optimization methods for feature selection.
[59], 2024	The study hybridized PSO, FA, and WOA algorithms to introduce the of deep Q-Learning module to intelligently updates feature importance during feature selection task.	The use of deep Q-learning is novel in the study; however, it was limited to software bug detection dataset when compared with the focus of this study.
[62], 2023	Machine learning based least square support vector machine (LSSVM) classifier was applied to feature selection problem by leveraging on modified quantum-based marine predators algorithm (Mq-MPA) method	In this study, we evaluated the use of machine learning-based classifiers and softmax function in classification tasks. This is and advantage of their work which was focused on LSSVM alone.
[63], 2021	The study investigated the usefulness of Ant Colony Optimization and Tabu search with Fuzzy Rough set for Optimal feature selection (ACTFRO) and Genetic algorithm and Tabu search with Fuzzy Rough set for Optimal feature selection (GATFRO) algorithms for feature selection.	The approach is limited only to continuous optimization (metaheuristic) to image feature selection problem.
[60], 2024	Machine learning method was applied for feature selection while harmony search (HS) and the genetic algorithm (GA) were used for the feature selection task.	The machine learning method used for feature extraction stage undermines the quality of feature learning process
[22], 2023	QLESCA for extraction of opti9mal feature subsets from the high dimensional space resulting from a sha9llow conventional neural network (SCNN) on X-ray samples.	The approach is limited only to continuous optimization (metaheuristic) to image feature selection problem.
[54], 2022	Proposed BEOSA for handling feature selection in text-based dataset.	The limitation of BEOSA with respect to mutation of instances during the exploration and exploitation stages, were inherited.
[23], 2021	Investigated the role of MRFGRO a continuous metaheuristic algorithm for feature selection through reduction of the noisy features	The approach is limited only to continuous optimization (metaheuristic) to image feature selection problem.
[25], 2023	Investigated the use of PSO, CSO, FFA, BA, FPO, WOA, ASOA, HHOA, BOA, and GWO with CNN on feature selection.	The approach is limited only to continuous optimization (metaheuristic) to image feature selection problem.
[27], 2022	FMCSA was used with CNN	The approach is limited only to continuous optimization (metaheuristic) to image feature selection problem.
[28], 2024	Combined the MCS algorithm with CNN for a CBMIR	The approach is limited only to continuous optimization (metaheuristic) to image feature selection problem.
[29], 2024	medical image segmentation can be improved using KOA and neural network on chest X-ray	The approach is limited only to continuous optimization (metaheuristic) to image feature selection problem.

continuous optimization method supersedes others in addressing the problem. On the other hand, [58] showed that a metaheuristic algorithm which is Levy Adapted SLnO (LA-SLnO) which is originally based on SLnO algorithm can improve the feature selection process on features extracted using Tri-Kernel principal component analysis (TK-PCA). Furthermore, the study noted that even the approaches of linear discriminant analysis and linear square regression can also interface with the LA-SLnO to ensure features extracted are optimized for classification such as the neural network. Interestingly, the use of deep Q-learning as a measure for ensuring the integration of a feedback mechanism to feature learning process, has been reported in [59]. The study aimed at addressing the limitation of traditional continuous optimization algorithms in that they fail to intelligently learn the task of feature selection. As such, the use of PSO, FA, and WOA algorithms with the deep Q-learning helped to discover feature importance during feature selection process.

Meanwhile, the performance of examines harmony search (HS) and the genetic algorithm (GA) in feature selection problem as used with support vector machine (SVM) classifier have been investigated in [60]. This was experimented to improve the detection of sarcopenia disease. Using the VOC 2007 dataset, [61] have also shown that Red Fox Optimization Algorithm is suitable for solving the same problem of feature selection. However, using image datasets, [62] have demonstrated that the performance of a least square support vector machine (LSSVM) classifier can be improved in classifying those samples. The study investigated the use of modified quantum-based marine predators' algorithm (Mq-MPA) on the classifier and reported some improvements. Still in the case of image-based classification problem, [63] have combined Ant Colony Optimization and Tabu search with Fuzzy Rough set for Optimal feature selection (ACTFRO) algorithm to support classifiers by aiding the use of only optimal feature sets. Furthermore, the authors also reported that a Genetic algorithm and Tabu search with Fuzzy Rough set for Optimal feature selection (GATFRO) algorithm was likewise investigated for the same task.

In Table 1, a summary review of most related studies considered in this section is provided to highlight their contributions and their limitations.

3. Proposed methodology

The approach described in this section is based on the novel design and use of hybrid binary optimization algorithms to find the optimal feature subsets from a neural network architecture. First, the hybrid binary metaheuristic algorithm is described with focus on the mathematical model and algorithmic representation. Furthermore, integration of the optimizer into the convolutional neural network architecture is presented and discussed.

3.1. The base BEOSA, BPSO, and BDMO algorithms

The base binary optimizer used for the hybrid method described in this subsection, is based on the BEOSA. The aim is to improve the exploration and exploitation phases of the BEOSA using some very outstanding binary optimizer whose mechanisms for achieving these

phases have been reported to be suitable for the problem domain. As result, the BPSO and BDMO are investigated for HBEOSA. The outcome of these hybridizations is two new binary optimizers namely the HBEOSA-PSO and HBEOSA-DMO. Further to this, the transfer functions used for binary optimizers for formalization of the search space are being redesigned to investigate the influence of nested transfer (NT) function on the performance of binary optimizers. These redesigned functions yielded new variants of the hybrids being investigated in this study so that we have the HBEOSA-PSO-NT and HBEOSA-DMO-NT. This subsection therefore is focused on the presentation of the methods and approaches that describe hybrid algorithms.

3.1.1. Binary ebola optimization search algorithm (BEOSA)

The BOESA was first presented in [54] from the original continuous variants EOSA. Using the traditional S-shaped and V-shaped functions, the continuous metaheuristic variant was binarized for solving problems with similar representation. The search space formalism for the binary variants allows for the initialization of solutions in the space with 1's, as seen in Eq. (1), so that the optimization process transforms the dimension (d) of each item or individual (x_i) to values in the range of 0 s to 1 s.

$$x_i = \sum_{j=0}^{d} x_{i,\ j} = 1 \tag{1}$$

The binary optimizer is based on the infection rate of the Ebola disease, so that items in the search space represent human individuals or organisms or animals that can be infected by the disease. So, it is desirable that the search space represents a population X that is susceptible to disease and that can be infected. In Eq. (2), the search space shows that it is possible to have n items or individuals with each having its anatomy represented by $d-size$ features.

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,d} \\ \vdots & \ddots & \vdots \\ x_{1,n} & \cdots & x_{n,d} \end{bmatrix} \tag{2}$$

The infection rate of the disease demonstrates the optimization process with an increased propagation rate capable of yielding different subgroups or subpopulations such as the infected, recovered, and others. However, the infected group yields a new set of individuals whose anatomical feature representation has been mutated. This mutation results in a search space with items or individuals whose configurations are no longer the $d-size1$ s representation. This then allows for computing the best individual in the solution space whose feature representation is optimal and suitable for transformation into the problem space for solving the binary optimization problem. The mutation of the individual x_i is denoted by Eq. (3) so that x_i^{new} is the outcome of the mutation resulting from the infection. Note that the variable $xbest$ in the equation is the computed best individual in the population X, and is obtained by first calculating the fitness of each x_i in X so that $x_1,\ x_2 \ldots x_n$ are sorted to obtain the best.

$$x_i^{new} = \Delta * e^{rnd}\cos(2\pi rnd) * (x_i - xbest) \tag{3}$$

The use of the notation Δ is to represent the mutation factor that differentiates the rate and way in which the individual is mutated by the infection. Meanwhile, the rnd returns a random number generated within the range of $[-1,\ 1]$ of a uniform distribution. However, this mutation was inherited from the continuous variant of the BEOSA with no support for the binarization of the population X. As a result, the transfer function which are S-shaped and V-shaped are applied for this purpose. In Eqs. (4) and (5), two types of the S-shaped namely the S1 and S2 are described. Furthermore, in Eqs. (6) and (7) represents the description of two types of V-shaped transfer functions. Interestingly, the transfer functions are applied to the mutated x_i^{new}.

$$S1 = \frac{1}{1 + e^{(-x_i^{new}/2)}} \tag{4}$$

$$S2 = 1 - \frac{1}{1 + e^{x_i^{new}}} \tag{5}$$

$$V1 = \left|\frac{x_i^{new}}{\sqrt{2 + x^2}}\right| \tag{6}$$

$$V2 = |\tan x_i^{new}| \tag{7}$$

This mutation and binarization functions are at the core of the BOESA and provide for using the method as a base for the hybrid methods. In the following paragraphs, we described the two other methods combined with the BOESA.

3.1.2. Binary particle swarm optimization (BPSO) algorithm

The BPSO was motivated from the natural lifestyle of swarm-like animals such as birds. Their ability to swarm as a population combined with their behavioral interaction presents an interesting optimization process [64]. While the continuous PSO has received a very wide application and acceptance, the binary variant demonstrated another unique algorithm for solving binary optimization problem. Unlike the BOESA which is based on infection of the animals or humans in the population space, the BPSO is focused on the position $p_{i,j}$ of the animals in the swarm so that initialization of the population space is focused on the random positioning. The optimization process is designed around ensuring that this positioning is changed, and the velocity of repositioning is also updated. The aim of the optimization process is to obtain new positions $p_{best,i,j}$ for an item in the swarm as different from its initial position $p_{i,j}$. However, the computation is also influenced by a know best position $g_{best,i,j}$ within the swarm. The velocity of the repositioning is computed using Eq. (8) for a new time $(t+1)$ during the optimization process. Using this computed velocity, the organism in the swarm for that same optimization process time is represented in Eq. (9). Once this repositioning and velocity changing is completed, the transfer function based on the sigmoid function in Eq. (12) is applied to binarize the organism/particle as shown in Eq. (10).

$$v_{i,j}(t+1) = wv_{i,j}(t) + c_1R_1\left(p_{best,\ i,j} - x_{p,\ i,j}(t)\right) + c_2R_2\left(g_{best,\ i,j} - x_{p,\ i,j}(t)\right) \tag{8}$$

$$x_{g,\ i,j}(t+1) = x_{g,\ i,j}(t) + v_{i,j}(t+1) \tag{9}$$

$$x_{p,\ i,j}(t+1) = \begin{cases} 1 rnd < T\left(x_{g,\ i,j}(t+1)\right) \\ 0 rnd \geq T\left(x_{g,\ i,j}(t+1)\right) \end{cases} \tag{10}$$

$$T\left(x_{g,\ i,j}(t+1)\right) = \frac{1}{1 + e^{-x_{g,\ i,\ j}(t+1)}} \tag{11}$$

Note that the use of the i and j notations are to represent the index of the particle in the swarm, position within the particle. In addition, the w, c_1, and c_2 are the inertia weight, acceleration coefficients 1, and acceleration coefficients 2 respectively. Meanwhile, the $R_1$and R_2 are chosen from between 0 and 1.

3.1.3. Binary dwarf mongoose optimizer (BDMO)

The BDMO [53] is based on the continuous variant namely the DMO. The method is motivated by the behavior of the dwarf mongoose animal and their nomadic lifestyle which often require the transit from one location to another in search of food. As a result, the position of the mongoose is the central focus of the mutation or change during the optimization process. Contrary to the BEOSA and BPSO which uses some unique transfer functions to binarize the population of the organisms,

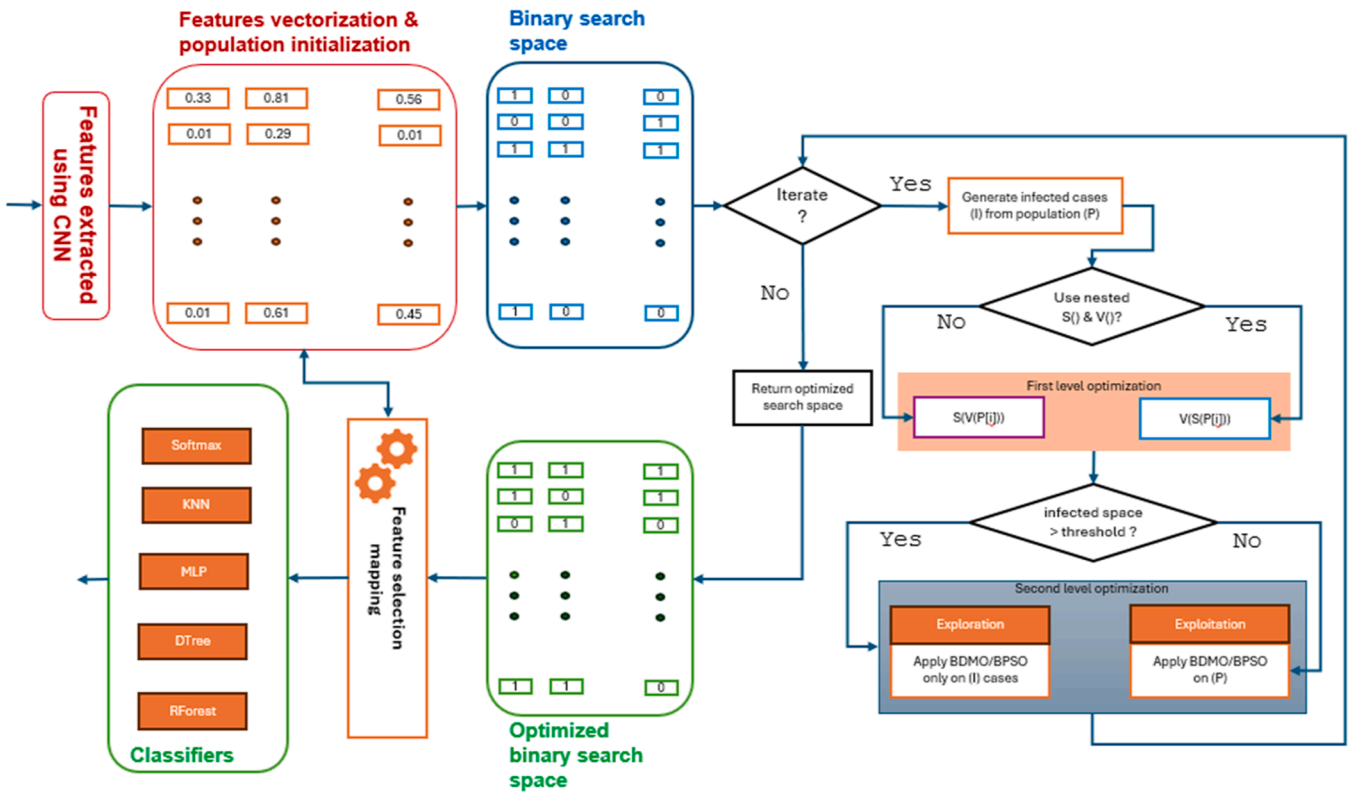


Fig. 1. An illustration of the optimization process of the HBEOSA-DMO, HBEOSA-PSO and their use of the nested transfer functions on the search space.

the BDMO relies on computing the best candidate food position as shown in Eq. (12), and then condition the mutation between 0 s and 1 s using in $0.5 < best^{fp}{}_{i+1} \leq 0.5$ Eq. (13). Note that the $x^{fp}{}_{i+1}$ is the current best computed food location, and the *peep* notation is a sound made during the search for such food location. The computation for the position of each dwarf mongoose in the population is obtained using Eq. (14).

$$best^{fp}{}_{i+1} = fp_i + phi * peep \tag{12}$$

$$x^d = \begin{cases} selected\ feature x^{fp}{}_{i+1} > 0.5 \\ feature\ not\ selected x^{fp}{}_{i+1} \leq 0.5 \end{cases} \tag{13}$$

$$x^{fp}{}_{i+1} = \begin{cases} best^{fp}{}_{i} - CF * rand * \left[x^{fp}{}_{i} - \vec{M}\right] if \varphi_{i+1} > \varphi_i exploration \\ best^{fp}{}_{i} + CF * rand * \left[x^{fp}{}_{i} - \vec{M}\right] else\ exploitation \end{cases} \tag{14}$$

Note that the use of the notation *CF* which is the parameter for directing the collective velocity of the movement of the mongoose group in the equations can be interpreted as $CF = (1 - \frac{iter}{Max_{iter}})^{(2\frac{iter}{Max_{iter}})}$, and $\vec{M}$ which is the vector controlling the group movement to sleep in another sleeping mound (sm) is represented by $\vec{M} = \sum_{i=1}^{n} \frac{x_i * sm_i}{x_i}$. Here Max_{iter} and *iter* denotes the maximum number of iterations for the optimization process and the current iteration point in the optimization process. Meanwhile, the φ notation helps to compute the average sleeping mound as given by $\varphi = \sum_{i=1}^{n} \frac{sm_i}{n}$.

These basic modelling for the BEOSA, BPSO, and BDMO provides a platform for generating two new algorithms with the focus on addressing the limitation of the BOESA. In the following paragraphs, we described how the combination of the methods is achieved.

3.2. Hybrid BEOSA-PSO and hybrid BEOSA-DMO

The hybrid methods are designed to improve the exploration and exploitation of the BEOSA so that where the algorithm seeks to explore new location in the search space, the BPSO or BDMO can be adapted to do the search. Similarly, where the exploitation process is desired to achieve a better local search for the optimal solution, the BPSO and BDMO are adapted for this. The outcome of this adaptation of the two optimizers produced the hybrid BEOSA with BPSO (HBEOSA-PSO) and hybrid BOESA with BDMO (HBEOSA-DMO). And algorithmic representation is designed to illustrate how this integration is achieved.

3.2.1. Overview of the hybrid binary optimizer

Fig. 1 summarizes the optimization procedure of the proposed HBEOSA-PSO and HBEOSA-DMO as applied to feature selection and optimization problem in medical image classification. Meanwhile, the feature extraction process using the CNN technique is discussed in subsection 3.2 and the use of the nested transfer functions are detailed in this subsection. The proposed method applies a forward mapping function to correlate the vectorized feature sets into a binary search space. This is to allow for the optimization of the binary search space using a two-level optimization strategy. Once the optimization of the binary search space is completed with the fitness function confirming that candidate solutions have been ranked, then the backward mapping function is applied. The selected features are then extracted to the classifier for classification operation.

3.2.2. The nested transfer functions

Further to this, the traditional transfer function applied for the

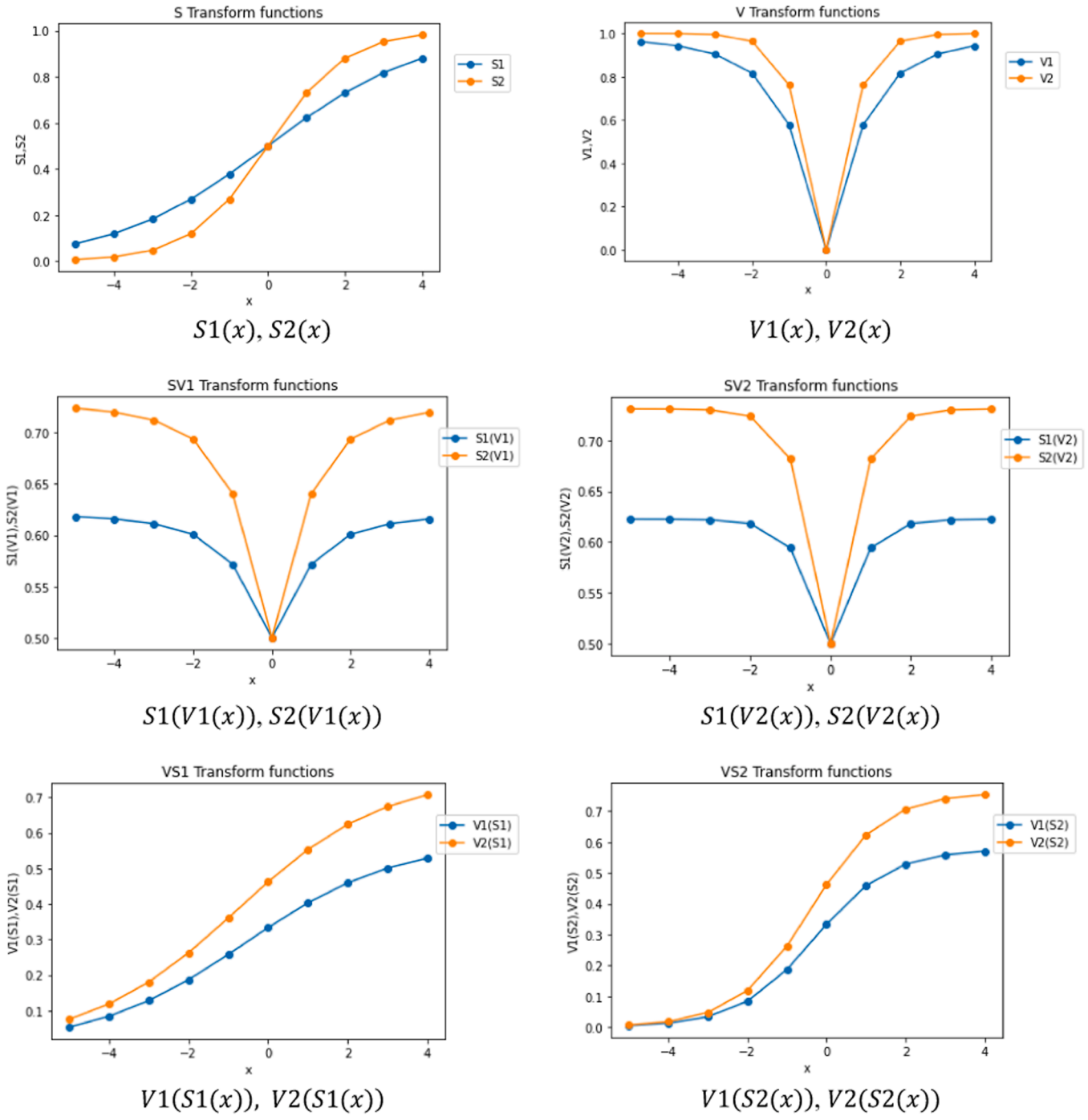


Fig. 2. Transfer function applied for BEOSA, BPSO, HBEOSA-DM and HBEOSA-PSO which are $S1$, $S2$, $V1$, $V2$, $S1(V1)$, $S2(V1)$, $S1(V2)$, $S2(V2)$, $V1(S1)$, $V2(S1)$, $V1(S2)$, and $V2(S2)$.

BOESA function is overridden to accommodate a new approach described in this study. In this improvement, the S1, S2, V1 and V2 transfer functions are nested to investigate what performance improvement a nested binary-mutation function can achieve in the optimization process. In Eq. (15), the nesting of the two transfer functions are illustrated and conditioned on the derivation of a $rand(0|1)$ randomly generated number of either 1 or 0. The S1 and V1 transfer functions are applied when the $rand(0|1)$ returns a 0, while the S2 and V2 are nested when $rand(0|1)$ returns 1. Once the $S(x_i^k), V(x_i^k)$ have been computed, the $d-size$ dimension of x_i is mutated within the context of the binary representation by deriving a random number r which satisfying $S(x_i^d) < r > T(x_i^d)$ as represented in Eq. (16).

$$S(x_i^d), \quad V(x_i^d) = \begin{cases} S2(x_i^d), V2(x_i^d) \& rand(0|1) == 1 \\ S1(x_i^d), V1(x_i^d) \& rand(0|1) == 0 \end{cases} \tag{15}$$

$$x_i^d = \begin{cases} 1 \& r > \quad S(x_i^d) | r > \quad T(x_i^d) \\ 0 \& otherwise \end{cases} \tag{16}$$

The nesting of the S-shaped transfer functions is based on $S1(V1)$, $S2(V1)$, $S1(V2)$, and $S2(V2)$ as represented in Eqs. (17), (18), (19), and (20) respectively.

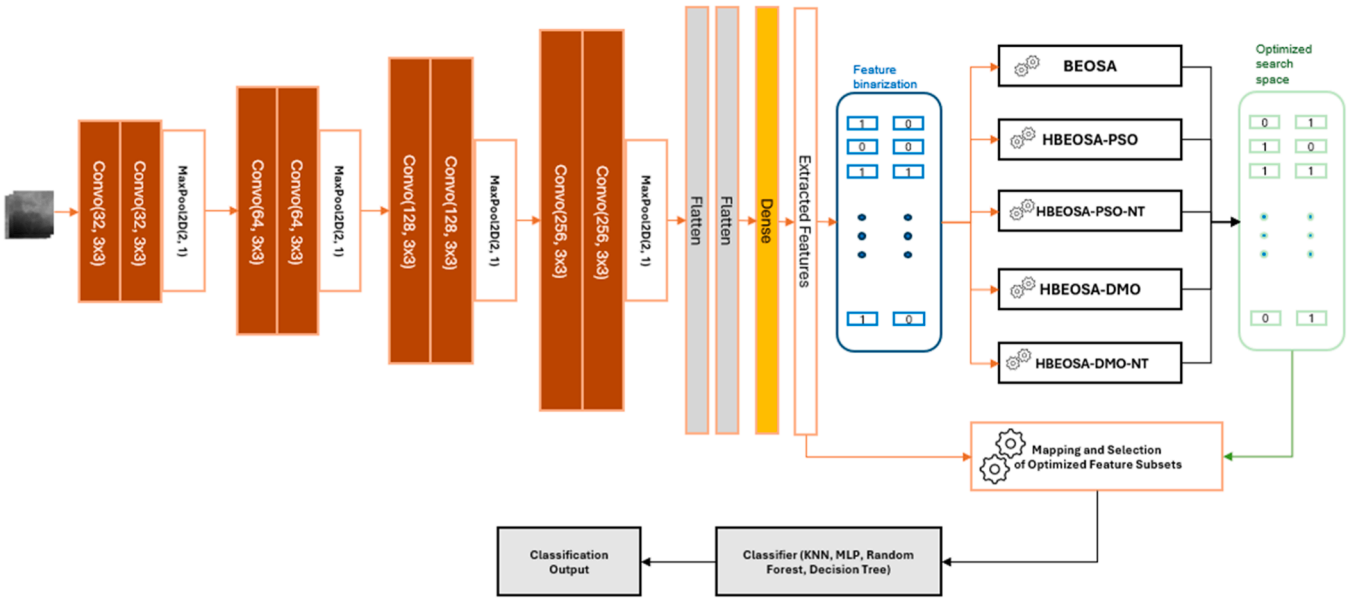


Fig. 3. A pipeline of feature extraction, feature optimization, and feature classification composing of a neural network architecture and binary optimization algorithms.

$$S1(V1) = \frac{1}{1 + e^{\left(-\left(\left|\frac{x}{\sqrt{2+x^2}}\right|\right)\Big/2\right)}} \tag{17}$$

$$S2(V1) = 1 - \frac{1}{1 + e^{\left(\left|\frac{x}{\sqrt{2+x^2}}\right|\right)}} \tag{18}$$

$$S1\left(V2\right) = \frac{1}{1 + e^{(-(|\tan x|)/2)}} \tag{19}$$

$$S2\left(V2\right) = 1 - \frac{1}{1 + e^{(|\tan x|)}} \tag{20}$$

Similarly, the nesting of the V-shaped transfer functions is based on the $V1(S1)$, $V2(S1)$, $V1(S2)$, and $V2(S2)$, functions in Eqs. (21), (22), (23), and (24) respectively.

$$V1\left(S1\right) = \left|\frac{\left(\frac{1}{1+e^{(-x/2)}}\right)}{\sqrt{2 + \left(\frac{1}{1+e^{(-x/2)}}\right)^2}}\right| \tag{21}$$

$$V2\left(S1\right) = \left|\tan\left(\frac{1}{1+e^{(-x/2)}}\right)\right| \tag{22}$$

$$V1\left(S2\right) = \left|\frac{\left(1 - \frac{1}{1+e^{x}}\right)}{\sqrt{2 + \left(1 - \frac{1}{1+e^{x}}\right)^2}}\right| \tag{23}$$

$$V2\left(S2\right) = \left|\tan\left(1 - \frac{1}{1+e^{x}}\right)\right| \tag{24}$$

The investigation on the shape distortion resulting from the application of the nesting transfer function revealed that the basic S-shape and V-shape were strongly represented. Interestingly, we observed some slight reshaping appearing within figures – demonstrating subtle points of influencing the binarization process. In Fig. 2, an illustration of $S1$, $S2$, $V1$, $V2$, $S1(V1)$, $S2(V1)$, $S1(V2)$, $S2(V2)$, $V1(S1)$, $V2(S1)$, $V1(S2)$, and $V2(S2)$ transfer functions are shown. The use of this function on the HBEOSA-PSO method resulted in the HBEOSA-PSO-NT algorithm. Also, the HBEOSA-DMO was investigated on these nested functions to obtain the HBEOSA-DMO-NT algorithm.

The curves indicates that all the $S1$, $S2$, $V1$, $V2$, $S1(V1)$, $S2(V1)$, $S1(V2)$, $S2(V2)$, $V1(S1)$, $V2(S1)$, $V1(S2)$, and $V2(S2)$ are distinct in their representation and interpretation. For instance, the curves of $S1$ and $S2$ shows that the pattern of the latter when applied to the dimension of x_i is different from the former. $S1$will return a new x_i based on it slightly linear S-shaped pattern while $S2$ will return a new x_i with a straighter S-shaped pattern. This same effect is rippled into $V1(S1)$, $V2(S1)$, $V1(S2)$, and $V2(S2)$. On the other hand, the transform pattern of $V1$and $V2$ are different as observed in the curve pattern because $V2$ demonstrates a narrow V-shape while $V1$ pattern indicates a shallow V-shaped. These differences in their transform approach reveal an interesting influence on the entire optimization process which this study aims to investigate and draw on the comparative analysis.

3.2.3. *Algorithm of the hybrid binary optimizers*

The pseudocode for the HBEOSA-DM, HBEOSA-PSO, HBEOSA-PSO-NT, and HBEOSA-DMO-NT is listed in Algorithm 1. The inputs to the algorithm are the number of iterations for the optimization process, the population of the base BEOSA method, the short and long change factor rate, and the dimension of each organism or individual in the population. The desired output from the optimization process is the best solution representing the optimal solution in the search space. Before the optimization process begins, the population is generated, and an index infected case is selected based on the first individual in the population. The use of the S-to-V-shaped transfer functions, that is the $S1(V1)$,

$S2(V1)$, $S1(V2)$, and $S2(V2)$, is conditioned to the exploration phase as seen on lines 16–22. On the other hand, the application of the V-to-S-shaped transfer functions, that is the the $V1(S1)$, $V2(S1)$, $V1(S2)$, and $V2(S2)$, is conditioned to the exploitation phase as listed on lines 23–29.

Algorithm 1. HBEOSA-DMO and HBEOSA-PSO method

```
1    Input: maxIter, psize, srate, lrate, dim
2    Output: gbest
3  begin
4    X =generate population based on psize
5    I, gbest ← X[0], X[0]
6   isThreshold = rand(1|0)
7    while e  < maxIter and  size (I) > 0 do:
8      generate a subpopulation to from the infected for quarantine
9      I=difference of current infected cases (I) from quarantine cases
10     for i in 1 to size(I) do:
11       generate new infected (nI) case from X
12       nI_i = (nI_i − gbest) ∗ e^rand ∗ cos(2π ∗ rand)
13       if ! isThreshold
14         for j in 1 to dim do:
15           randomly generate d between 1|0
16           if displacement(nI_i) > 0.5 do:
17             update size of nI using srate
18             s= { S2(V1(nI_{i,j}))| S2(V2(nI_{i,j}))   d = 1
                  { S1(V1(nI_{i,j}))| S1(V2(nI_{i,j}))   d = 0
19             if s >= rand do:
20               nI_{i,j} = 1
21             else:
22               nI_{i,j} = 0
23           else:
24             update size of nI using lrate
25             v = { V2(S1(nI_{i,j}))| V2(S2(nI_{i,j}))   d = 1
                   { V1(S1(nI_{i,j}))| V1(S2(nI_{i,j}))   d = 0
26             if t >= rand do:
27               nI_{i,j} = 1
28             else:
29               nI_{i,j} = 0
30       if displacement(nI_i) < 0.5:
31         nI = BDMO(nI)
32       else:
33         X = BPSO(X)
34       Evaluate new fitness of nI_i
35     I ← nI
36     Update gbest, I and quarantine variables
37   return gbest
38 end
```

Once the binary mutation of every x_i has been achieved across the $d-size$ dimension of that individual or organism, the BDMO method is applied for generating a new set of infected subgroups during the exploration phase on lines 30–31. Also, when the algorithm shifts to the exploitation phase, the BPSO is applied to regenerate the complete susceptible population on lines 32–33. This is informed by the progressive change in the immunity of the susceptible population and the increasing mutation of the virus among the infected.

The evaluation of the fitness of each x_i is based on the Eq. (25) which applies a classifier *clf* to the x_i to compute all indexes of 1 in x_i. Furthermore, a control variable ω which represents a random number obtained in the range [0,1] for weight assignment to the error classification, is used to keep the fitness value within range. The parameter d

Table 2
Parameter settings for CNN, HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA binary metaheuristic algorithms.

Method	Parameter	Description
BEOSA	β_1=0.1, β_2=0.1, β_3=0.1, β_4=0.1, and π=0.1,	Contact rate of infected (β_1), of host (β_2), with the dead (β_3), and with the recovered (β_4), and recruitment rate (π)
BDMO	Nb=3, peep=1, τ = rand(0, 1), L= round(0.6*D*nb)	Number of baby-seaters (Nb), Peep sound, tau operator (τ), babysitter exchange parameter (L)
BPSO	c_1=2, c_2=2, W=0.9, V_{max}=6	Positive learning factors constant 1 (c_1) and constant 2 (c_2), initial weight (W), maxfor imum velocity vector (V_{max})
CNN	∂ = 1e − 06, α = *Adam*, β_1 = 0.5 and β_2 = 0.999, ε = 1e − 08, τ = 0.0002, φ = 32, W = 299,H = 299, epoch=40, stride=1, layers = 8, padding=false, batch size=64, dropout=0.5 filter size=3, filter counts=32, 64, 128, 256	learning rate (∂), optimizer algorithm (α), beta1 (β_1), beta2 (β_2), epsilon (ε), L2 regularizer rate (τ), batch size, image width (W), image height (H) and number of training epochs respectively

denotes the number of features represented in the datasets. $\left|\dot{F}\right|$ is a set of features selected as represented in x_i and $|F|$ is the number of those features which are total features sets represented in the dataset. On the contrary, a cost function, represented in Eq. (26), is also computed by the hybrid method to compute how far the x_i is from an optimal position. The parameter ω we applied to control the quality of the classification performance, and to as well to regulate the relevance of the feature subsets selected. Experimentally, we choose the value of 0.99 for the ω parameter because conventionally, $\omega \in [0, 1]$.

$$fit = \omega * \left(1 - clf(x_i)\right) + \left((1-\omega) * \frac{\left|\dot{F}\right|}{|F|}\right) \tag{25}$$

$$cost = 1 - fit \tag{26}$$

Where $1 - clf(x_i)$ is the classification error rate computed as: $1 - \frac{number\ of\ correctly\ classifed}{number\ of\ classified\ in\ data}$. In the following subsection, we describe the application of these hybrid binary optimizations algorithms to finding a subset of optimal features extracted from image samples using neural network architectures.

3.3. Formulation of the HBEOSA methods for neural network optimization

The challenge of feature optimization and selection has often been addressed using the continuous metaheuristic algorithm approach. In this section, we demonstrate the approach to formulation of the application binary metaheuristic algorithm to the same problem of feature optimization. In Fig. 3, the pipeline shows the feature extraction process, feature optimization, classification, and sample labeling or prediction. The digital mammography samples are passed as input to the pipeline. This is followed by a neural architecture which extracts the image features using its convolutional-pooling layers. The dimension of the extracted complete features is passed for creation of search space and initialization. The search space becomes input to the hybrid binary optimization methods. Output from the binary optimizers is the optimized search space which is then mapped to the feature sets for selection of the matching spaces. The selected feature subsets are supplied to the classifier for the classification process which returns the prediction.

The CNN architecture applied for the feature selection task is composed of four blocks of convolutional-pooling layers. Each block has two-padded convolutional layers which is then followed by a max-pooling layer. The convolutional layers were designed to use a 3×3 filter size while the filter counts were kept varied across blocks. For instance, the first two-padded convolutional layers in the first block use 32 filter counts, while those similar layers in the second block were architected to use 64 filter counts. The filter counts of 128 and 256 were applied to convolutional layers of the third and fourth blocks of the convolutional-pooling blocks respectively. Meanwhile, the filter count used for the max-pooling layers in each block is 2, while the strides is 1. The four blocks of convolutional-pooling layers are followed by two flattened layers and a dense layer. These last three layers are applied for vectorizing the matrix representations of the extracted features. The motivation for application of the CNN architecture is based on experimental evaluation and benchmarking of the performance of the different architecture for feature extraction task on the datasets considered in this study. Further to this, the study applied some architectural pruning and parameter tuning to arrive at the current configuration of the architecture and parameter combination. Meanwhile, the choice of a CNN architecture over the transformer-based vision transformers (ViTs) is informed on the aim to control the model optimization process which is easier with CNN compared with ViTs. Moreover, the moderate model size combined with the efficient hierarchical and spatially invariant feature extraction approach of the architecture and proven impressive performance over ViTs in some task [65], inspired its use for the investigation reported in this study.

The search space is built based on the dimension of the feature sets extracted. For instance, given that a 100-sample batch (f_{100}) of digital mammography is supplied as input to the neural architecture, and a width and height of 299×299 represents the image size, then $f_i = \begin{bmatrix} f_{1,1} & \cdots & f_{1,299} \\ \vdots & \ddots & \vdots \\ f_{299,1}, & \cdots & f_{299,299} \end{bmatrix}$ where i = 1–100. However, for the optimization purpose, we are not interested in the feature sets but rather the dimension of the feature set. It is expected that the outcome from the vectorization of the f_i will yield transformed feature representation as $f_i = \begin{bmatrix} v_1 \\ \vdots \\ v_D \end{bmatrix}$. The dimension D of f_i is used for building the search space so that only 100 individuals or organism are added to the search space with each having a vector representation in the Ddimension initialized to 1 s, that is $x_i = \begin{bmatrix} 1_1 \\ \vdots \\ 1_D \end{bmatrix}$. The x_i in the search space now represents the input to the binary optimization algorithms. These algorithms pass their inputs through the optimization process to yield outputs such as $x_i = \begin{bmatrix} 0_1 \\ \vdots \\ 1_D \end{bmatrix}$ as the optimized feature space. The task of the mapping and selection compartment is to ensure that $f_i = \begin{bmatrix} v_1 \\ \vdots \\ v_D \end{bmatrix}$ is mapped to $x_i = \begin{bmatrix} 0_1 \\ \vdots \\ 1_D \end{bmatrix}$ so that only corresponding slots with 1 s are composed as the feature subset. The feature subset for every f_i is supplied to the classifier for classification operations which usually return probability distributions that can be mapped into the label set.

The discussion in this section has described the design of the binary optimization methods and the corresponding variants which are based on the nested transfer functions. Furthermore, the architecture of CNN used for feature extraction, and the formalism for the binarization of the feature space through optimization process have been discussed. In the following sections, the experimentation demonstrating the methodology described in this section are presented and discussed.

Table 3
datasets listing and their corresponding details such as number of features, classes, and instances, a description on the dataset.

Dataset and References	Number of Instances	Number of features	Number of classes	Dimensionality of dataset
BreastEW	569	30	2	High-dimensional dataset
WaveformEW	5000	40	3	
Sonar	208	60	2	
PenglungEW	325	73	7	
KrVsKpEW	3196	36	2	
Leukemia	7070	72	2	
Ionosphere	351	34	2	
Colon	2000	62	2	
CongressEW	435	16	2	Medium-scale dimensional dataset
Lymphography	148	18	4	
SpectEW	267	22	2	
Vote	300	16	2	
Zoo	101	16	7	
Exactly	1000	13	2	Low-dimensional datasets
Iris	150	4	2	
Exactly2	1000	13	2	
M-of-n	1000	13	2	
Tic-tac-toe	958	9	2	
Wine	178	13	3	

4. Experimentation setup and datasets

The experimentation of the methodology described in the previous section was conducted using some selected datasets and within some computational environments. In this section, a description of the configurations and settings for the algorithms and methods proposed are presented. Furthermore, the datasets applied for exhaustive experimentation are presented and discussed to provide for reproducibility of the results obtained in this study.

4.1. Computation environment configuration and settings

Two categories of experimentation are conducted in this study involving hybrid binary optimization methods, and experiment on the application of the hybrid algorithms to medical image analysis. The training of the binary metaheuristic algorithms was conducted using a computer running Windows 10 operating system, with Intel processor Core i5-4200 with speed of 1.70 GHz and 2.40 GHz. The memory size for this computer is 16 GB, with sufficient secondary drive space. On the other hand, the computational environment for the application of the hybrid binary optimizer was conducted in Intel Silver Xeon 4210 computer system with about 10 processors or central processing unit (CPU). The memory of the system is 256 GB and external drive space as much as 4TB-960GB of hard disk drive (HDD) – solid-sate drive (SSD). In addition to the primary memory, the system provided 2 extra graphical processing units (GPU) with each having 32 GB space.

In Table 2, the summary is presented of the algorithmic configurations and settings used for experimenting on the computational environment described in the previous paragraph. The binary optimization algorithms BEOSA, BDMO and BPSO are described based on their parameter settings. We note that these single algorithm-based settings were applied to the hybrid binary methods experimented with in this study. This implies that HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, and HBEOSA-PSO-NT were experimented based on the settings and configurations of their based methods.

The values for the optimization algorithms were experimentally investigated and proven suitable for their application to feature selection problem. For instance, the choice of the recruitment and infection rates of BEOSA have been investigated in our previous studies [54], and also the choice of the values for initial weight and babysitter exchange parameter for BPSO and BDMO respectively were inspired from a previous study [14]. In Section 5, a thorough sensitivity and computation cost analysis on these benchmark datasets, upon desirable performance, we proceeded to apply to address the main research problem of the study. These becnhmark datasets used for this purpose are namely BreastEW, Colon, Ionosphere, KrVsKpEW, Leukemia, PenglungEW, Sonar, WaveformEW, CongressEW, Lymphography, SpectEW, Vote, Zoo, Exactly, Exactly2, Iris, M-of-n, Tic-tac-toe, and Wine. Furthermore, the neural network applied for the feature extraction was experimented using the parameters and some hyperparameters described and listed in the table. Here, we explicitly mentioned the learning rates, the learning optimizer, beta one and two values, epsilon parameter, L1 regularizer value, the batch size for training, and lastly the image size. All these

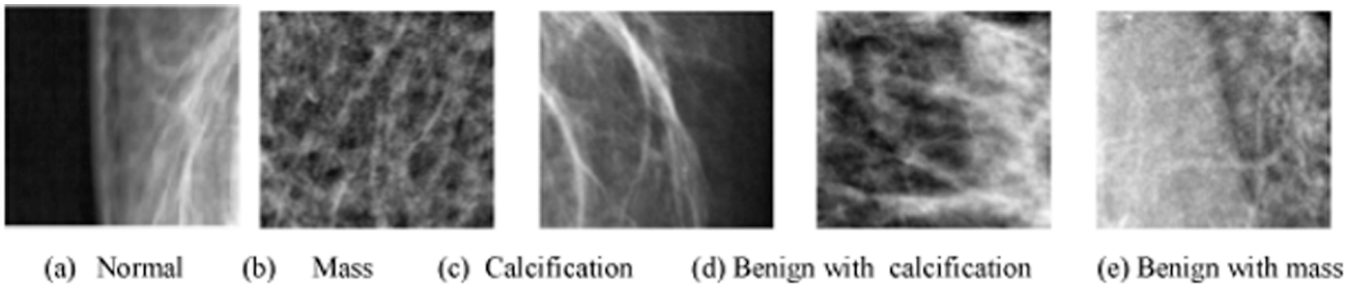


Fig. 4. Samples of images to be extracted from a combined datasets sourced from DDSM+CBIS and MIAS databases. Image labels follows: (a) Normal sample (N), (b) Mass abnormality (M), (c) calcification abnormality (CALC), (d) benign calcification (BC), and (e) benign with mass (BM).

Table 4
Digital mammography dataset listing according to their class-based distribution.

Dataset and References		Total image samples	Normal samples (N)	Samples with Mass (M)	Samples with calcification (CALC)	Samples with benign calcification (BC)	Samples with benign with mass (BM)
DDSM+CBIS [67]		21844	19077	728	660	708	671
	Training samples	16382	14307	546	495	531	503
	Evaluation samples	2182	1907	72	66	70	67
	Testing samples	3274	2861	109	99	106	100
MIAS [66]		3075	2718	45	159	36	117

Table 5
Evaluation of the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA methods using high-dimensional datasets on accuracy, fitness, cost function and feature counts.

Dataset	Algorithm	Acc-50	Acc-100	Fit-50	Fit-100	Cost-50	Cost-100	FC	CT
BreastEW	HBEOSA-DMO	0.942	0.940	0.056	0.054	0.944	**0.946**	15.000	2540.629
	HBEOSA-DMO-NT	0.949	**0.965**	0.057	**0.021**	**0.943**	0.979	8.600	2766.993
	HBEOSA-PSO	0.963	0.953	0.072	0.035	0.928	0.965	7.000	2260.324
	HBEOSA-PSO-NT	**0.965**	0.946	**0.016**	0.053	0.984	0.947	6.381	3086.175
	BEOSA	0.947	0.947	0.054	0.054	0.946	**0.946**	**5.000**	**0.021**
Colon	HBEOSA-DMO	**1.000**	**1.000**	0.085	0.085	0.915	0.915	**2.000**	2413.129
	HBEOSA-DMO-NT	**1.000**	**1.000**	**0.002**	**0.000**	0.998	1.000	134.000	2088.655
	HBEOSA-PSO	**1.000**	**1.000**	0.089	0.157	**0.911**	**0.843**	76.077	2469.002
	HBEOSA-PSO-NT	**1.000**	**1.000**	**0.002**	**0.000**	0.998	1.000	45.315	2392.427
	BEOSA	**1.000**	**1.000**	0.0005	0.002	0.9995	0.9980	396.000	**0.022**
Ionosphere	HBEOSA-DMO	0.886	0.879	0.0849	0.085	0.9151	0.9146	**2.000**	708.539
	HBEOSA-DMO-NT	0.900	0.936	0.100	0.087	0.900	**0.913**	8.500	723.254
	HBEOSA-PSO	0.886	0.857	**0.055**	0.086	0.945	0.914	5.000	793.166
	HBEOSA-PSO-NT	0.914	0.936	0.114	**0.073**	**0.886**	0.927	4.500	786.935
	BEOSA	**0.943**	**0.943**	0.057	**0.057**	0.943	0.943	**2.000**	**0.013**
KrVsKpEW	HBEOSA-DMO	0.937	0.938	0.641	0.058	**0.359**	0.942	16.000	2943.689
	HBEOSA-DMO-NT	0.931	0.947	0.057	0.072	0.943	**0.928**	15.000	2894.754
	HBEOSA-PSO	0.900	0.934	0.148	0.057	0.852	0.943	1057.872	2846.351
	HBEOSA-PSO-NT	**0.966**	**0.961**	**0.037**	**0.028**	0.963	0.972	**10.257**	3115.554
	BEOSA	0.948	0.948	0.055	0.055	0.945	0.945	12.899	**0.012**
Leukemia	HBEOSA-DMO	**1.000**	**1.000**	0.0003	**0.000**	0.9997	1.000	145.000	3047.149
	HBEOSA-DMO-NT	0.987	**1.000**	**0.00047**	**0.000358**	0.9995	0.9996	213.000	2693.147
	HBEOSA-PSO	0.947	**1.000**	0.0001	0.0005	0.9999	0.9995	165.000	3301.407
	HBEOSA-PSO-NT	**1.000**	0.973	0.0004	0.0003	0.9996	0.9997	**91**.000	2799.496
	BEOSA	0.933	0.933	0.066	0.066	**0.934**	**0.934**	**74.000**	**0.016**
PenglungEW	HBEOSA-DMO	0.767	0.667	0.044	0.002	0.956	0.998	40.000	248.420
	HBEOSA-DMO-NT	0.867	0.733	0.068	0.067	0.932	**0.933**	17.500	250.196
	HBEOSA-PSO	0.800	0.767	0.069	0.023	**0.931**	0.977	49.000	256.531
	HBEOSA-PSO-NT	0.767	0.867	0.067	0.059	0.933	0.941	59.000	256.038
	BEOSA	**0.933**	**0.933**	**0.000**	**0.000**	1.000	1.000	**9.732**	**0.018**
Sonar	HBEOSA-DMO	**0.929**	0.810	0.096	0.120	0.904	**0.880**	17.000	646.479
	HBEOSA-DMO-NT	0.869	**0.917**	0.143	0.119	**0.857**	0.881	**15.000**	672.351
	HBEOSA-PSO	0.857	0.821	**0.095**	**0.072**	0.905	0.928	9.407	513.774
	HBEOSA-PSO-NT	0.881	0.905	0.142	0.095	0.858	0.905	15.112	634.425
	BEOSA	0.905	0.905	0.099	0.099	0.901	0.901	25.000	**0.013**
WaveformEW	HBEOSA-DMO	0.811	0.780	0.200	0.197	0.800	0.803	20.000	4554.501
	HBEOSA-DMO-NT	0.813	**0.820**	0.177	0.189	0.823	0.811	24.500	4146.997
	HBEOSA-PSO	0.781	0.798	**0.114**	**0.086**	0.886	0.914	16.000	4549.970
	HBEOSA-PSO-NT	**0.820**	0.810	0.194	0.169	0.806	0.831	**17.034**	4685.436
	BEOSA	0.801	0.800	0.202	0.202	**0.798**	**0.798**	20.000	**0.012**

parameter configurations and settings for the algorithms and models provided means for conducting the experiment in the computational environments described in this subsection. In the following subsection, the datasets applied for the experimentation are discussed.

4.2. Datasets

The nature of the two-fold experimentation conducted in this study necessitated the use of two categories of datasets namely the text-based datasets and the data sources from medical image samples. The first category of datasets was used for investigating the viability of the hybrid binary optimization methods to check their performance with respect to classification accuracy, fitness, and cost function values. Once the outcome of this first experiment proved useful, we advanced to the second experiment which applied the medical image datasets to the hybrid binary optimizers to improved feature extraction of the neural network's architecture.

4.2.1. Text-based datasets

There are three types of datasets used for the investigation of the performance of the hybrid binary optimizers. These are the high-dimensional datasets, medium-dimensional datasets, and low-dimensional datasets. The high-dimensional datasets include the BreastEW, Colon, Ionosphere, KrVsKpEW, Leukemia, PenglungEW, Sonar, and WaveformEW. The CongressEW, Lymphography, SpectEW, Vote, and Zoo were listed as the medium scale sized datasets. The low-dimensional datasets include Exactly, Exactly2, Iris, M-of-n, Tic-tac-toe, and Wine. This dimensionality classification is based on the number of instances and the number of features listed in each dataset in Table 3.

Interestingly, the datasets have were drawn from different domains, demonstrating the richness and rigour of the experimentation testing the viability of the hybrid binary optimizers. For instance, the BreastEW is biology-based and medical oriented dataset, the WaveformEW is generates three classes of waves with each class sampled at 21 intervals, the Sonar represents a dataset of sonar signals, the PenglungEW is a biology-based and medical dataset, the KrVsKpEW and Tic-tac-toe are game-based, the Leukemia dataset is a biology-based and medical-based dataset, the Ionosphere is based on electromagnetic records, the colon dataset is also a biology-based and medical-based as similar to Lymphography, SpectEW, and M-of-n. on the other hand, the CongressEW dataset is a congressional voting, as well as the Vote dataset. The Zoo and Iris are biology-based datasets. The Exactly and Exactly2 are artificial binary classification datasets, while the Wine dataset is based on information on chemicals in wines.

4.2.2. Image-based datasets

The second experimentation applying the hybrid binary optimizer to medical images uses the MIAS [66] which has been preprocessed into the CBIS of the DDSM+CBIS [67]. In Fig. 4 some samples of these datasets ate shown and some ground truth labeling listed with each sample. Th samples have samples with normal mammography images, and those with different abnormalities such as the calcification, and micro mass. Furthermore, the dataset allows for using samples with benign mass and benign calcification features.

Table 6
Evaluation of the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA methods using medium-dimensional datasets on accuracy, fitness, cost function and feature counts.

Dataset	Algorithm	Acc-50	Acc-100	Fit-50	Fit-100	Cost-50	Cost-100	FC	CT
CongressEW	HBEOSA-DMO	0.954	0.931	**0.036**	0.046	0.964	**0.954**	**1.000**	477.454
	HBEOSA-DMO-NT	0.954	**0.971**	0.058	**0.026**	0.942	0.974	6.500	466.500
	HBEOSA-PSO	**0.989**	0.966	0.047	**0.026**	**0.953**	0.974	1.996	515.259
	HBEOSA-PSO-NT	0.966	0.954	0.037	0.036	0.963	0.964	4.000	531.094
	BEOSA	0.943	0.943	0.060	0.060	0.940	0.940	5.000	**0.013**
Lymphography	HBEOSA-DMO	0.833	0.000	0.136	0.165	0.864	**0.835**	6.000	718.970
	HBEOSA-DMO-NT	**0.917**	0.900	**0.074**	0.037	0.926	0.963	7.500	599.726
	HBEOSA-PSO	0.783	0.900	0.233	**0.036**	**0.767**	0.964	9.000	650.063
	HBEOSA-PSO-NT	0.900	**0.917**	0.103	0.067	0.897	0.933	**4.124**	636.246
	BEOSA	0.900	0.000	0.103	0.103	0.897	0.897	6.524	**0.011**
SpectEW	HBEOSA-DMO	0.843	**0.870**	0.170	0.150	0.830	0.850	**7.000**	617.395
	HBEOSA-DMO-NT	**0.889**	0.824	**0.096**	0.187	0.904	**0.813**	10.000	596.261
	HBEOSA-PSO	0.815	0.796	0.186	**0.077**	**0.814**	0.923	7.360	601.872
	HBEOSA-PSO-NT	0.870	**0.870**	0.130	0.132	0.870	0.868	7.632	608.789
	BEOSA	0.870	**0.870**	0.132	0.132	0.868	0.868	**7.000**	**0.016**
Vote	HBEOSA-DMO	**0.975**	0.942	**0.004**	0.051	0.996	**0.949**	4.500	465.388
	HBEOSA-DMO-NT	0.958	**0.983**	0.051	**0.020**	**0.949**	0.980	5.500	491.004
	HBEOSA-PSO	0.967	0.958	0.019	0.050	0.981	0.950	1.498	459.809
	HBEOSA-PSO-NT	0.967	0.967	0.019	0.035	0.981	0.965	**3.410**	459.633
	BEOSA	0.967	0.967	0.035	0.035	0.965	0.965	3.416	**0.011**
Zoo	HBEOSA-DMO	0.950	0.920	**0.002**	0.147	0.998	**0.853**	3.000	494.763
	HBEOSA-DMO-NT	0.925	0.900	0.104	0.053	**0.896**	0.947	6.000	477.552
	HBEOSA-PSO	**1.000**	0.900	0.092	0.053	0.908	0.947	**1.641**	454.429
	HBEOSA-PSO-NT	**1.000**	**1.000**	0.004	**0.006**	0.996	0.994	8.000	440.724
	BEOSA	1.000	1.000	0.006	**0.006**	0.994	0.994	8.000	**0.011**

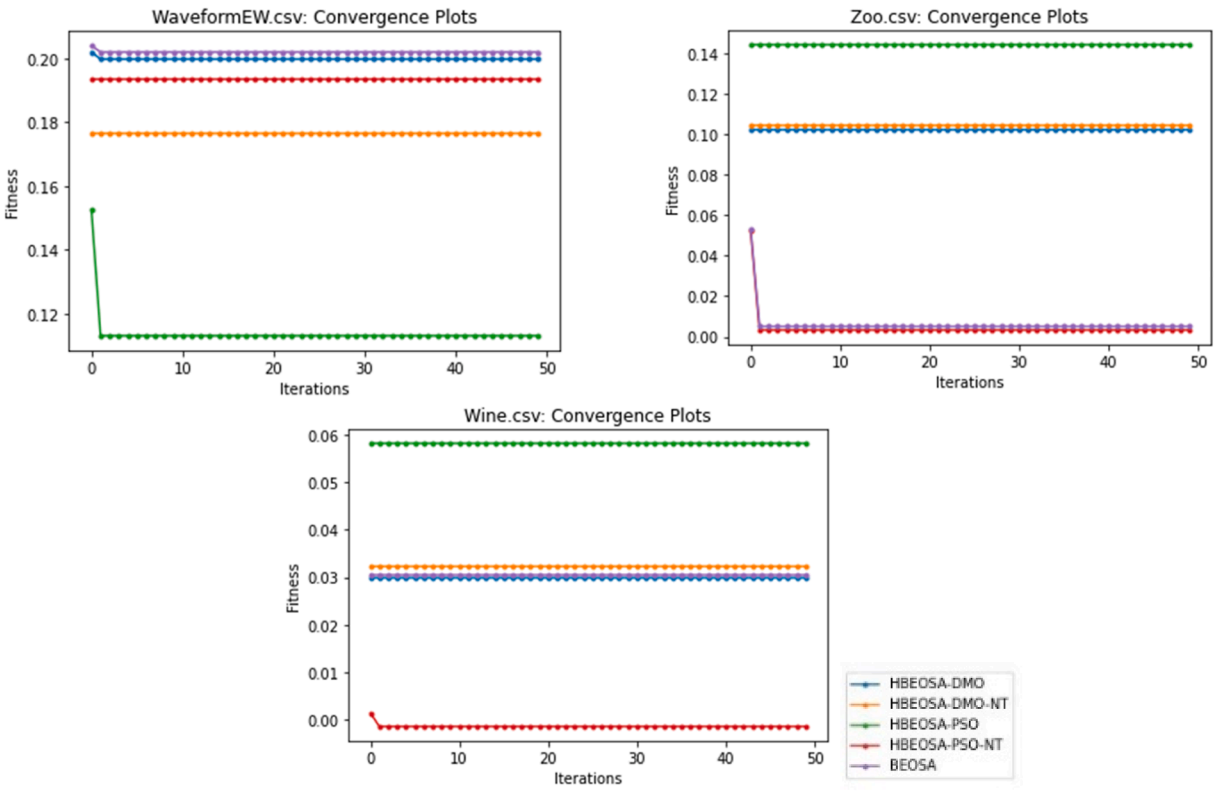
To eliminate noise from the MIAS and DDSM+CBIS datasets, this study used image enhancement technique, namely, contrast-limited adaptive histogram equalization (CLAHE). Outcome of the images from the CLAHE operation, a wavelet decomposition packet function is applied to extract high resolution and rich feature representation of each image passed through seam carving procedure.

The breakdown of the image samples applied for the investigation of the proposed hybrid binary solution is provided in Table 4. The table lists the statistics of data applied from DDSM+CBIS and those from the MIAS sources. Sample size of 21,844 were resulting from DDSM+CBIS having 16,382 allocated for training, 2182 for evaluation during training, and 3274 for testing or prediction. On the other hand, the MIAS samples were set aside for a second-round evaluation/testing of the trained model. This is necessary to mitigate against overfitting the model.

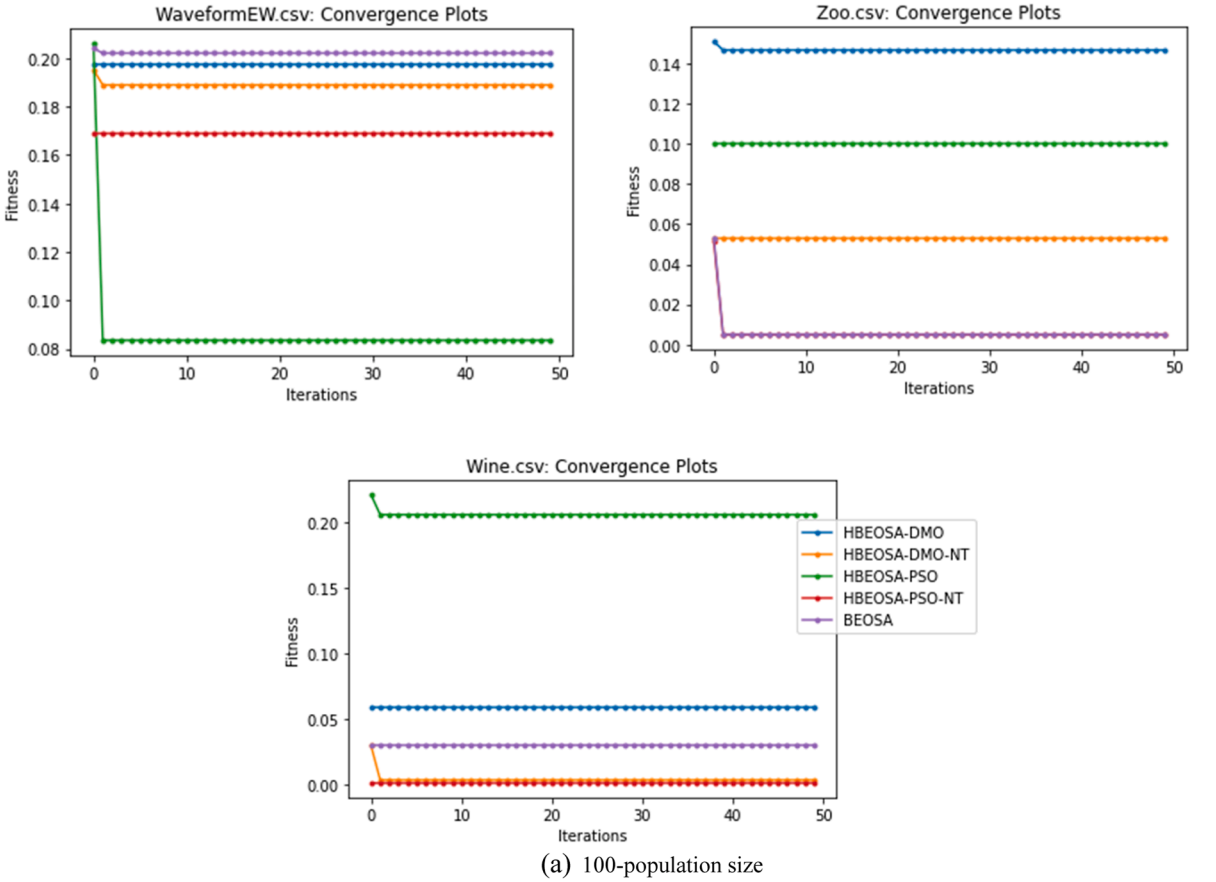
These datasets with the five different categories of sample labels only

Table 7
Evaluation of the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA methods using low-dimensional datasets on accuracy, fitness, cost function and feature counts.

Dataset	Algorithm	Acc-50	Acc-100	Fit-50	Fit-100	Cost-50	Cost-100	FC	CT
Exactly	HBEOSA-DMO	0.760	0.652	0.168	0.301	0.832	0.699	6.000	934.955
	HBEOSA-DMO-NT	0.695	0.792	0.308	0.154	**0.692**	0.846	7.000	946.215
	HBEOSA-PSO	0.890	0.758	**0.009**	0.308	0.991	**0.692**	**1.000**	1027.426
	HBEOSA-PSO-NT	0.760	0.877	0.306	0.232	0.694	0.768	6.668	1151.243
	BEOSA	**0.980**	**0.900**	0.022	**0.022**	0.978	0.978	**2.336**	**0.011**
Exactly2	HBEOSA-DMO	0.700	0.722	0.237	0.238	0.763	0.762	**1.000**	1153.631
	HBEOSA-DMO-NT	0.770	0.760	0.238	0.238	**0.762**	0.762	3.500	820.203
	HBEOSA-PSO	0.740	0.650	**0.025**	**0.235**	0.975	**0.765**	2.000	855.708
	HBEOSA-PSO-NT	**0.778**	**0.762**	0.222	0.238	0.778	0.762	3.000	1091.790
	BEOSA	0.765	0.760	0.236	0.236	0.764	0.764	5.000	**0.011**
Iris	HBEOSA-DMO	0.953	0.967	0.071	0.054	**0.929**	0.946	2.000	2146.082
	HBEOSA-DMO-NT	**0.973**	**0.973**	0.036	0.070	0.965	**0.920**	1.600	2019.983
	HBEOSA-PSO	0.940	0.933	**0.001**	0.066	0.999	0.934	**1.000**	2187.545
	HBEOSA-PSO-NT	0.967	0.953	0.003	**0.033**	0.998	0.967	1.014	1884.631
	BEOSA	0.967	0.967	0.036	0.035	0.965	0.964	**1.000**	**0.016**
M-of-n	HBEOSA-DMO	0.810	0.797	0.326	0.217	**0.674**	**0.783**	6.000	874.501
	HBEOSA-DMO-NT	0.843	0.820	0.183	0.208	0.817	0.792	96.000	854.640
	HBEOSA-PSO	**0.918**	0.900	0.166	0.127	0.834	0.873	1.276	708.316
	HBEOSA-PSO-NT	0.837	**0.910**	0.152	**0.027**	0.848	0.973	**4.550**	757.834
	BEOSA	0.855	0.850	**0.148**	0.148	0.852	0.852	6.000	**0.010**
Tic-tac-toe	HBEOSA-DMO	0.742	0.719	0.179	0.252	0.921	**0.748**	**1.000**	1816.082
	HBEOSA-DMO-NT	0.776	0.773	0.209	0.224	0.791	0.776	6.000	1686.338
	HBEOSA-PSO	0.758	**0.779**	0.244	**0.018**	**0.756**	0.982	**1.000**	1804.136
	HBEOSA-PSO-NT	**0.820**	0.773	**0.151**	0.238	0.849	0.762	5.000	1778.422
	BEOSA	0.781	0.781	0.222	0.222	0.778	0.778	5.000	**0.014**
Wine	HBEOSA-DMO	0.903	0.903	0.030	0.059	0.970	**0.941**	5.000	1014.824
	HBEOSA-DMO-NT	0.972	**1.000**	0.032	0.004	**0.968**	0.996	4.500	762.768
	HBEOSA-PSO	**0.986**	0.889	**0.003**	0.025	0.997	0.975	**2.137**	840.244
	HBEOSA-PSO-NT	**0.986**	0.986	0.056	**0.002**	0.944	0.998	3.065	1090.850
	BEOSA	0.972	0.972	0.031	0.031	0.969	0.969	4.000	**0.011**

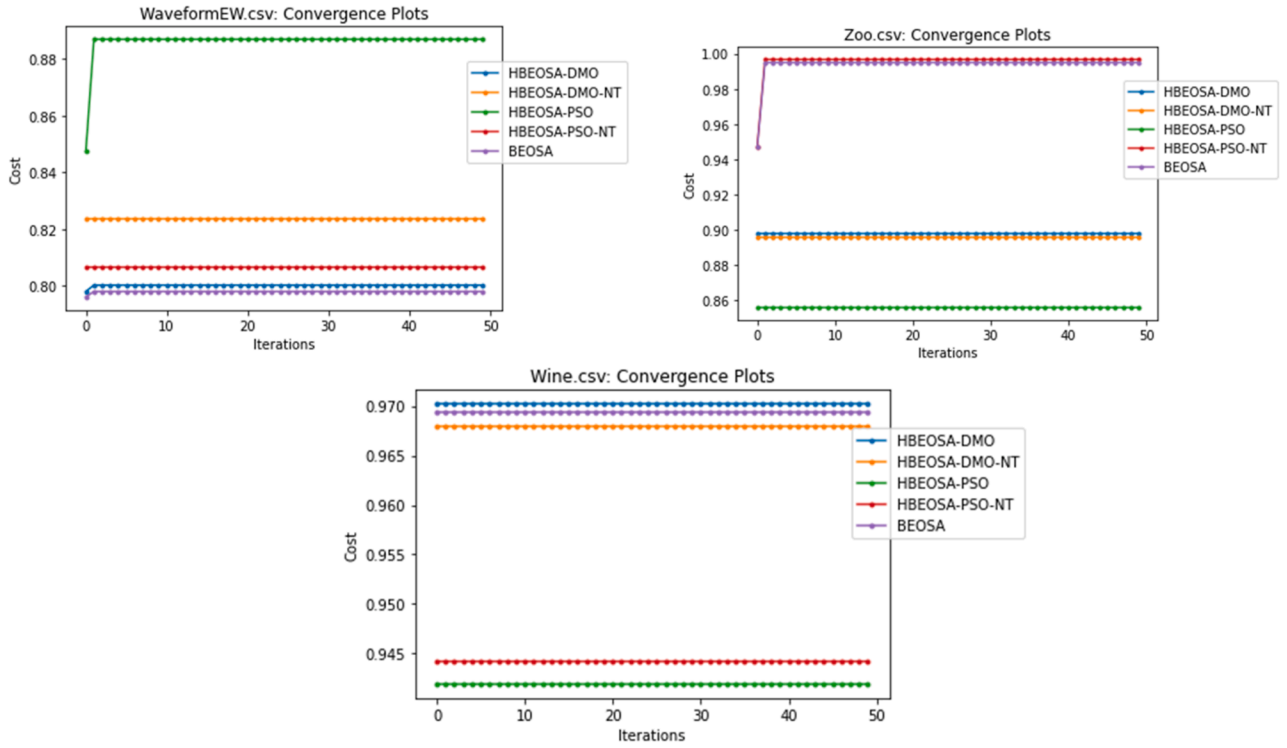


(a) 50-population size

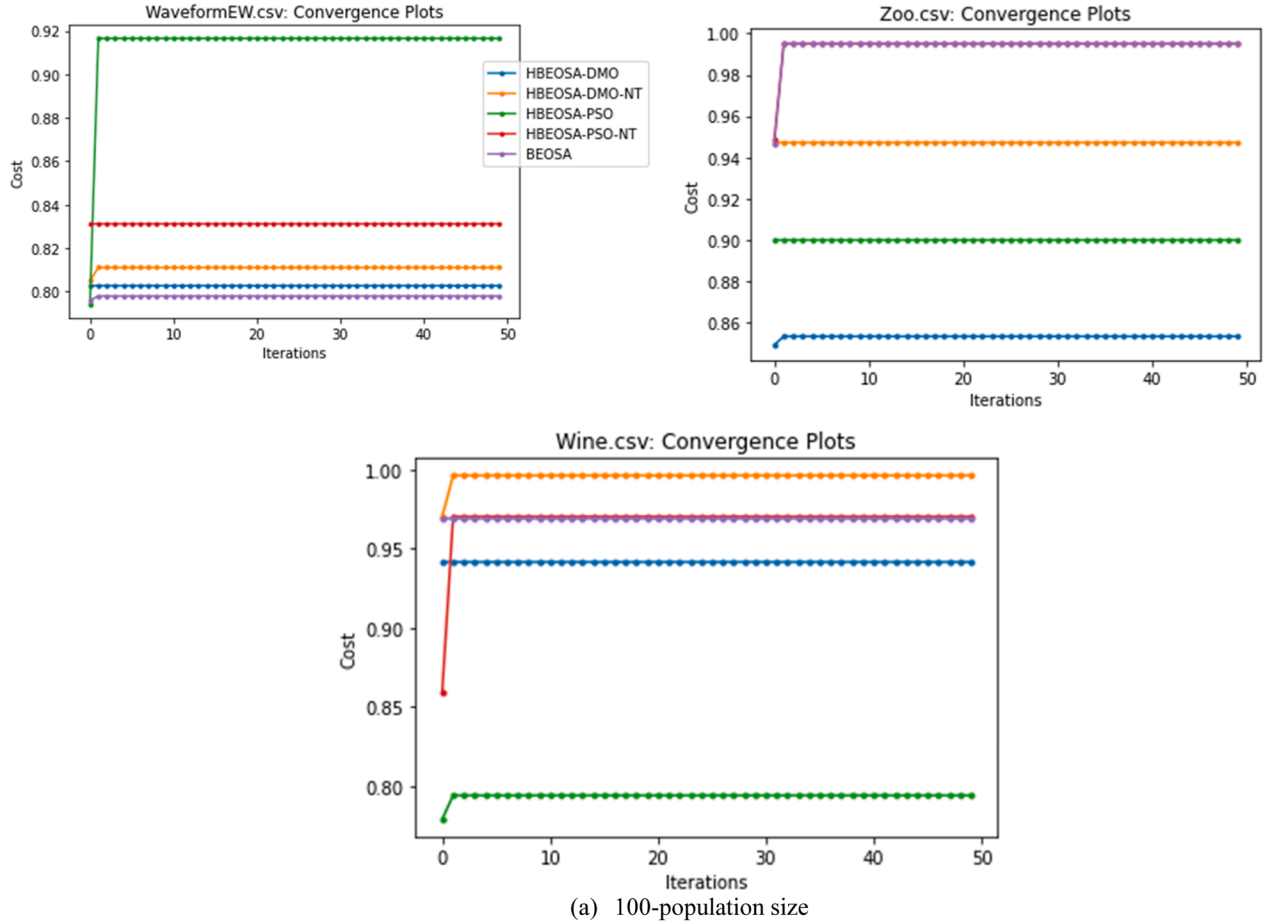


(a) 100-population size

Fig. 5. Line plot showing the fitness-function optimization visualization for the high-dimensional, medium-dimensional, and low-dimensional datasets during the 50 iteration points. (a) showing the fitness-function for 50-population size, and (b) showing the fitness-function for the 100-population size.



(a) 50-population size



(a) 100-population size

Fig. 6. Line plot showing the cost-function optimization visualization for the high-dimensional, medium-dimensional, and low-dimensional datasets during the 50 iteration points. (a) showing the cost-function for 50-population size, and (b) showing the cost-function for the 100-population size.

amounted to about 75658 instances. However, to ensure that this study overcomes overfitting and underfitting of the neural network models, transformational data augmentation technique was applied to the samples to increase the sample size to 31688 instances. This argumentation technique includes image flipping, rotating and some other operations. These provided sufficient image samples, in addition to the text-based datasets, for the experimentation that produced the results discussed in the following section.

5. Results and discussion

The experimentation described in the last section demonstrates the approach proposed in this study, and performance associated data were collected for results presentation in this section. To elaborate on this results presentation, in this section provided in three subsections which are: evaluation of the hybrid BEOSA methods with the non-hybridized binary variant of the EOSA algorithm; performance and discussion of the statistical analysis of the performances of the four binary algorithms considered in this study; and lastly, a discussion of the impact of the application of the hybrid binary optimizers to addressing neural network optimization architectures for better medical image analysis. Meanwhile, we provide a state-of-the-art comparison of the results of the application of the binary optimization algorithm with other studies which have applied the continuous optimization algorithm.

5.1. Evaluation of the hybrid BEOSA methods

The results on the performance of HBEOSA-DMO HBEOSA-DMO-NT HBEOSA-PSO HBEOSA-PSO-NT and BEOSA methods are investigated and comparatively analyzed in this subsection. Using the metrics of accuracy, fitness function, cost function, feature count, and computation time (CT), the five methods are evaluated for determining which is more superior in application to addressing binary optimization challenges in real-life problems. Recall that the classification accuracy demonstrates the measure of the binary optimizers in effectively selecting a minimal number of features which are optimal in helping the classifiers produce good result. Meanwhile, the fitness and cost functions allow for investigating how the methods effectively fits each solution in the search space and what is the domain-related cost value required in achieving this fitness function. On the other hand, the feature count demonstrates the approximate number of features required by the optimizer to provide the classifier for improved performance. Lastly, we investigate what is the computational time required to apply the algorithm to the binary optimization process to derive good fitness and cost values, and as well to extract the optimal feature counts. Meanwhile, the experimentation was applied to categories of population to understand what the impact of population size is in influencing the performance being understudy in this research. As a result, we have chosen to investigate this using population sizes of 50 and 100. All presentations of the performance and evaluation of the HBEOSA-DMO HBEOSA-DMO-NT HBEOSA-PSO HBEOSA-PSO-NT and BEOSA methods are based on dataset dimension. Therefore, our result presentation is further subdivided into high-dimensional, medium-dimensional, and low-dimensional datasets performance evaluation.

In Table 5, the performance evaluation of the high-dimensional datasets namely BreastEW, Colon, Ionosphere, KrVsKpEW, Leukemia, PenglungEW, Sonar, WaveformEW, are outlined. Th result obtained for the accuracy on 50 and 100 population sizes showed that the HBEOSA-PSO and HBEOSA-PSO-NT returned the best accuracy for 50 population size, while the HBEOSA-DMO-NT returned the best classification accuracy for 100 population size. Results obtained for the fitness and cost function values demonstrate that the HBEOSA-PSO-NT and HBEOSA-DMO-NT yielded better performance for the 50-population size on the fitness and cost respectively. However, when the 100-population size was applied, the HBEOSA-DMO-NT returned the best for fitness while HBEOSA-DMO and BEOSA returned the best performance for the cost function. Meanwhile, feature count results performance showed that HBEOSA-DMO-NT, HBEOSA-PSO-NT and BEOSA were more optimal and acceptable than the others. Interestingly, the BEOSA used the most minimal computational time compared to the other methods. Results obtained for the Colon and Ionosphere datasets were more interesting with the classification accuracy on 50 and 100 population sizes in each datasets showed that HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA for Colon, while only BEOSA demonstrated best performance on the Ionosphere dataset. For both datasets, BEOSA returned the minimal values for computational time and feature counts. However, the values returned for the fitness and cost functions showed that HBEOSA-DMO-NT and HBEOSA-PSO-NT were the best under 50 and 100 population sizes.

The computational time obtained for the KrVsKpEW, Leukemia, PenglungEW, Prostate, Sonar, and WaveformEW datasets showed that the BEOSA method remains the most efficient algorithm, and as well yielding the most optimal feature counts. However, when the classification accuracy is for these datasets were investigated for the 50 and 100 population sizes, we observed that HBEOSA-DMO, HBEOSA-DMO-NT, and HBEOSA-PSO-NT were topping. Moreover, the performance for the fitness and cost function values showed that the BEOSA algorithm was more outperforming than others. The implication of our findings on the application of the high-dimensional datasets showed that while the hybrid methods and their corresponding NT-variants were more outperforming in classification accuracy, fitness, and cost function, they lagged the BEOSA method in the feature count and computational time. This demonstrates that hybrid methods are computationally expensive though they can help achieve better fitness of the search space and have a better understanding of yielding relevant features to the classifier for better performance.

Furthermore, experimental results obtained for the medium-sized datasets are reported in Table 6. The CongressEW, CongressEW, Lymphography, SpectEW, Vote, and Zoo datasets were also investigated with the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA methods. As is observed with the high-dimensional datasets, we found a reoccurring performance for BEOSA with respect to achieving the best in feature count and computational time for all the medium-dimensional datasets. Again, the implication of this is that in most datasets and real-life problems, the BEOSA method which does not hybridize with other methods are more desirable and optimal in helping to select the best combination and number of features in a dataset at a reduced computational time. However, when we desire to improve classification accuracy and obtain very good fitness and cost function values, hybrid methods are suitable for these. Meanwhile, the population sizes used were also found to minimally impact performance differences observed for these metrics.

The classification accuracy for the CongressEW and Lymphography showed that HBEOSA-PSO was best on 50 population size while HBEOSA-DMO-NT was suitable for population size of 100. However, when their performances were investigated on their capability to fit the solution space and reduce cost function, the BEOSA was optimal with respect to cost function for both 50 and 100 population sizes, while HBEOSA-DMO-NT was optimal for on fitness function for both 50 and 100 population sizes. Though the HBEOSA-PSO high performing under the Lymphography dataset for the fitness and cost, it could not yield superior performance with the classification accuracy. Furthermore, an investigation of the performance of the SpectEW, Vote, and Zoo datasets showed that the hybrid methods, particularly HBEOSA-DMO-NT, always outperformed the non-hybrid method on the fitness and cost function values. However, for the performance on classification accuracy, the HBEOSA-PSO and HBEOSA-PSO-NT are more competitive on the population size of 50 with the Zoo dataset. The summary of our findings of the performance evaluation on the medium-dimension datasets on the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA methods revealed that applying the hybrid methods to obtain better performance on the classification accuracy, fitness

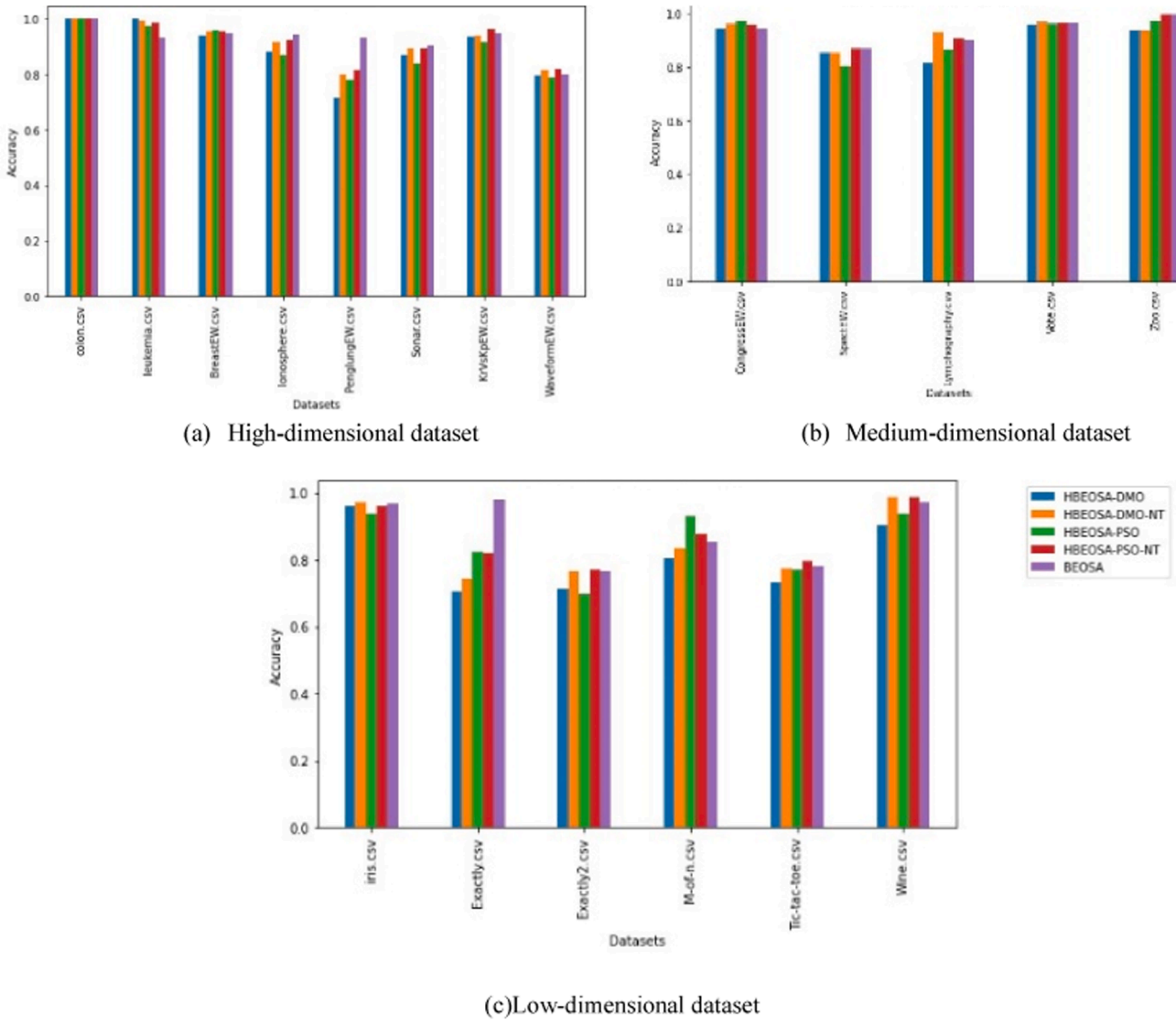


Fig. 7. Bar chart on the classification accuracy of the high-dimensional, medium-dimensional, and low-dimensional datasets.

function, and cost function is more optimal compared to the BEOSA methods. Moreover, it is desirable that our data instances are well classified with reduced cost and fitness values than having low feature counts which does not achieve the aim of performance enhancement. Therefore, our conclusion is that the hybrids method is more desirable when using high and medium dimensional datasets in real-life applications.

The performance of the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA methods on the low-dimensional datasets were also investigated and results listed in Table 7. The datasets considered in this category were Exactly, Exactly2, Iris, M-of-n, Tic-tac-toe, and Wine. An interesting reoccurring pattern observed for the high-dimensional, medium-dimensional, and low-dimensional datasets with respect to performance on number of feature counts returned and computational time required is that both HBEOSA-PSO-NT, and BEOSA were competing, though BEOSA remained at the topmost with respect to computational time. For the 50population size, BOESA was best for Exactly dataset, HBEOSA-PSO-NT best for Exactly2, Tic-tac-toe, and Wine datasets, HBEOSA-DMO-NT was the best for the Iris dataset, and HBEOSA-PSO was the best for the M-of-n dataset. This distribution of high performance among all methods for these low-dimensional datasets showed that any of the methods will be suitable for low-dimensional dataset. Interestingly, a similar performance is observed for the classification accuracy using the 100-population size. However, when the fitness and cost functions were evaluated on the datasets based on the five optimization methods, we found that the HBEOSA-PSO and HBEOSA-PSO-NT yielded better performance by obtaining good fitness values while the HBEOSA-DMO and HBEOSA-DMO-NT returned better cost function values for most low-dimensional datasets.

The summary of the findings on the investigation of the performances of the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA on high-dimensional, medium-dimensional, and low-dimensional datasets is that the BOESA method remains efficient with respect to computational time. However, performance evaluations on the classification accuracy, fitness function, and cost function values showed that the hybrid binary optimizers are more optimal and suitable for application to real-life optimization problem. This is important because while it is desirable to return a very minimal number of feature sets, it is also most desirable to have the optimizer select and combine the most relevant features to help the classifier perform well. Moreover, the fitness function is a factor for judging the quality of the search space to know what solution to select for solving an optimization problem. Therefore, since the hybrid methods such as the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, and HBEOSA-PSO-NT are returning the best fitness and cost values, the outcome from

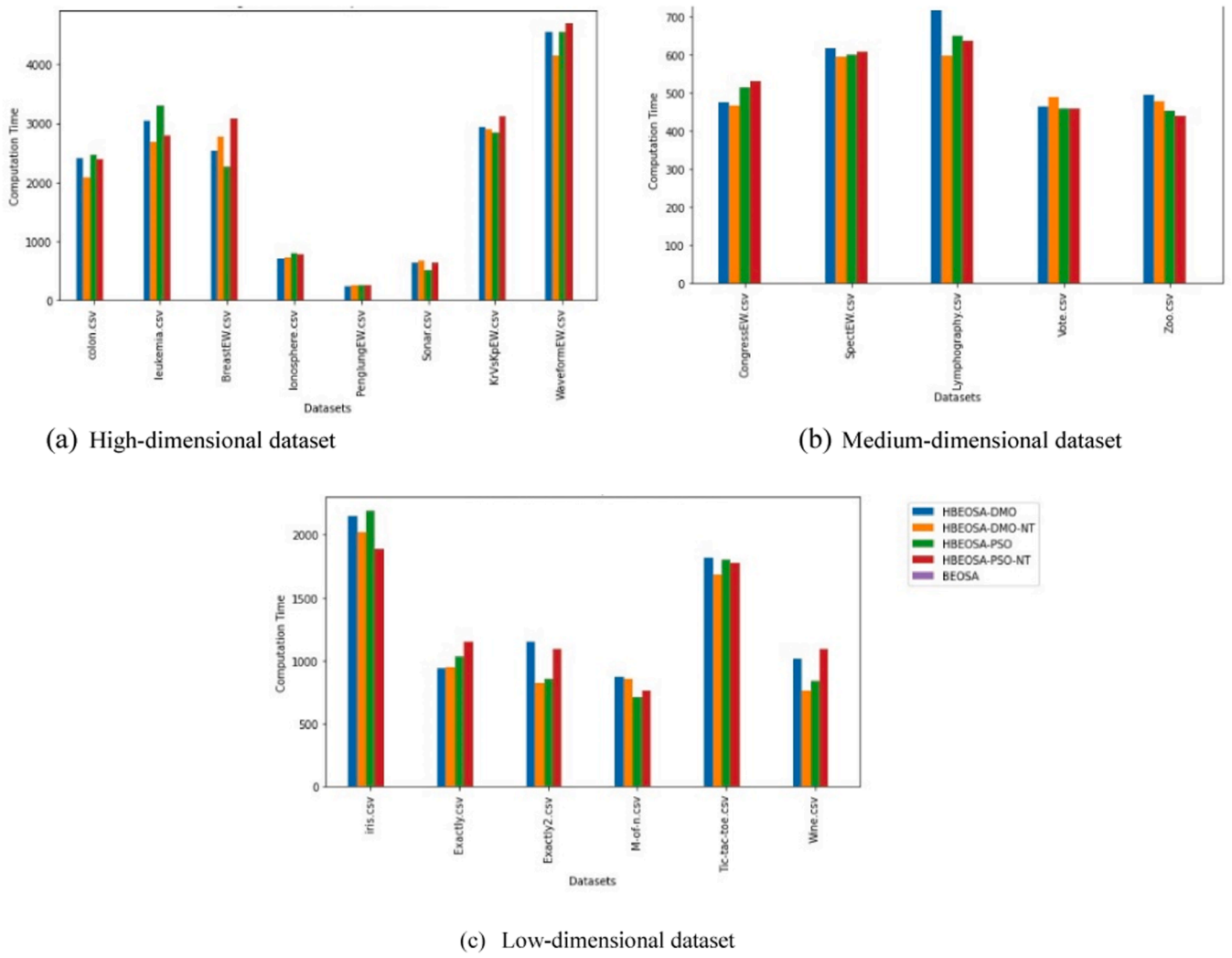


Fig. 8. Bar chart on the computational time comparison of the high-dimensional, medium-dimensional, and low-dimensional datasets.

our study demonstrates and confirms that these hybrids method are to be desirable over the basic BEOSA method.

The plot for the fitness values and cost function values obtained for the number of iterations experimented are graphed in Figs. 5 and 6 respectively. These plots are very important to help understand how the convergence curve for these fitness and cost functions flowed during the optimization process. For brevity, we examined the convergence curves for the fitness and cost functions for WaveformEW, Zoo, and Wine which belong to the high-dimensional, medium-dimensional, and low-dimensional dataset categories. The fitness function convergence curves for all datasets considered showed that the BEOSA method has a curve that always sits at the bottom of the graph except in the case of the WaveformEW – a generalization on the high dimensional datasets. Whereas the hybrid binary optimization methods such as the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT always compete by sitting at the top of the graph (for high dimensional datasets e.g WaveformEW), and mid-level (for medium dimensional datasets e.g Zoo), bottom (for low dimensional datasets e.g Wine). This is important to understand the implication of the position along which a curve runs. For instance, the curve of HBEOSA-DMO sits at the top position for the WaveformEW datasets both in the cases of 50 and 100 population sizes. The implication of this is that throughout the optimization process over the 50 and 100 iterations times, the HBEOSA-DMO algorithm returned minimal fitness values when compared with HBEOSA-DMO algorithm which returned best fitness for 50 and 100 population sizes. The performance of HBEOSA-DMO-NT and HBEOSA-PSO-NT methods on the medium-dimensional (e.g Zoo) datasets yielded the best in 100 population size and 50 population size respectively. Similarly, when the 50

Table 8
A comparison of the binary optimization method proposed in this study with continuous methods based on computational cost on some selected benchmark datasets.

	Hybrid binary method with nested transfer functions				**Single binary method with NO nested transfer functions**			
Dataset	HBEOSA-PSO-NT	HBEOSA-DMO-NT	HBEOSA-PSO	HBEOSA-DMO	BPSO	BDMO	BEOSA	BIEOSA
Colon	2549	2211	2597	2561	10,541	10,026	8971	9562
Leukemia	2838	2758	3378	3053	15,367	14,626	10,973	13,082
SpectEW	601	584	591	621	3386	3314	4389	4118
CongressEW	535	476	511	482	3293	3167	3681	3439
Iris	1821	2046	2401	2387	1569	2454	3638	4367
Exactly	1258	910	1107	889	5576	5139	6135	5239

Table 9
Summary of ANOVA: Single Factor.

	Fit-100 population				Accuracy-100 population			
Groups	Count	Sum	Average	Variance	Count	Sum	Average	Variance
HBEOSA-DMO	9	0.946397	0.105155	0.008399	9	7.596332	0.844037	0.015607
HBEOSA-DMO-NT	9	0.733902	0.081545	0.003983	9	8.080683	0.897854	0.008604
HBEOSA-PSO	9	0.693548	0.077061	0.008385	9	7.852659	0.872518	0.008505
HBEOSA-PSO-NT	9	0.745773	0.082864	0.005426	9	8.22799	0.914221	0.002913
BEOSA	9	0.614745	0.068305	0.003277	9	8.25254	0.916949	0.002236

Table 10
ANOVA for the single factor analysis.

	Fit-100 population						Accuracy-100 population					
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
Between Groups	0.006698	4	0.001674	0.284	**0.886557**	2.605975	0.0341	4	0.00853	1.12572	**0.3581**	2.6060
Within Groups	0.235757	40	0.005894				0.30292	40	0.00757			
Total	0.242454	44					0.33702	44				

and 100 population sizes are also considered, we found that the same HBEOSA-DMO, HBEOSA-DMO-NT and HBEOSA-PSO-NT methods returned a very good convergence curve which lies mostly at the bottom. However, the HBEOSA-PSO showed a fitness curve that underperformed when compared with other hybrid methods.

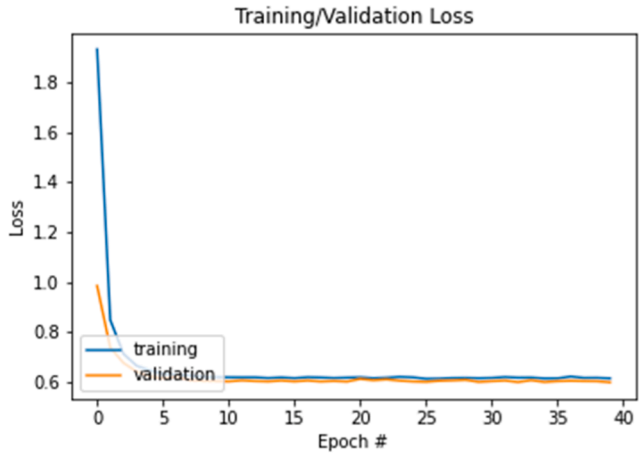
As another example, when the convergence curves for the cost function were examined, the hybrid methods HBEOSA-DMO, HBEOSA-DMO-NT and HBEOSA-PSO-NT showed a very strong competition because they obtained low-cost functions for the high-dimensional (e.g WaveformEW) datasets. This implies that the methods successfully optimized the cost objective function so that the process was completed at a reduced optimization computation cost and gaining good fitness values. Recall that these HBEOSA-DMO-NT and HBEOSA-PSO-NT are those methods using the proposed nested transform functions. Therefore, it is important to observe here that the proposed nested transfer functions have contributed to addressing the issue of maximization of the cost function objective. On the other hand, their corresponding variants namely the HBEOSA-DMO and HBEOSA-PSO also performed well especially in the cases of Zoo dataset (for HBEOSA-DMO) using 50 population size, Wine (in 50 and 100 population sizes), and Zoo (in 50 population sizes) datasets. Meanwhile, for the desirable performance among the curves showing well maximized objective cost function, the HBEOSA-DMO returned the best curves in the cases of 50 and 100 population sizes for the high-dimensional (e.g WaveformEW) datasets. The HBEOSA-DMO and HBEOSA-PSO curves were the best when used with 50 and 100 population sizes on Zoo dataset. Interestingly, the HBEOSA-PSO showed competitive performance on the 50 and 100 population sizes for the low-dimensional (e.g Wine) datasets.

The implication of these observations with respect to the convergence curves of the fitness and cost function values is that the optimization process over those number of iterations reveals that the hybrid binary optimization methods can find optimal binary solutions from the search space using consistent and quality fitness. The same interpretation holds for the cost values obtainable when the hybrid binary optimizers are applied to finding optimal binary solutions from the search space in real-life optimization problems.

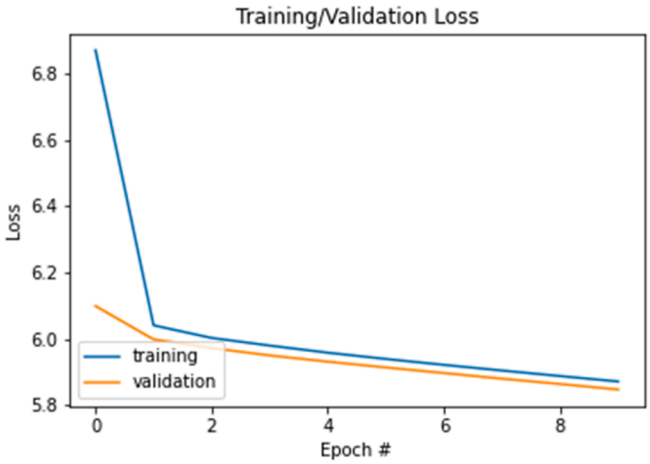
Another important graphing observed and plotted during the experimentation are those for the classification accuracy and those for the computational runtime. In Fig. 7, we illustrate the differences observed for the classification accuracy for the high-dimensional, medium-dimensional, and low-dimensional datasets. For the high-dimensional datasets all the methods showed that there was a strong competition among them for the Colon and WaveformEW datasets, however, the BreastEW, Ionosphere, KrVsKpEW, Leukemia, and Sonar

Table 11
Two-factor ANOVA analysis based on the fitness and accuracy metrics.

SUMMARY	**HBEOSA-DMO**	**HBEOSA-DMO-NT**	**HBEOSA-PSO**	**HBEOSA-PSO-NT**	**BEOSA**	**Total**
fitness						
Count	9	9	9	9	36	9
Sum	0.733902	0.693548	0.745773	0.614745	2.787969	0.733902
Average	0.081545	0.077061	0.082864	0.068305	0.077444	0.081545
Variance	0.003983	0.008385	0.005426	0.003277	0.00485	0.003983
Accuracy						
Count	9	9	9	9	36	9
Sum	8.080683	7.852659	8.22799	8.25254	32.41387	8.080683
Average	0.897854	0.872518	0.914221	0.916949	0.900385	0.897854
Variance	0.008604	0.008505	0.002913	0.002236	0.005409	0.008604
Total						
Count	18	18	18	18	18	
Sum	8.814585	8.546208	8.973763	8.867284	8.814585	
Average	**0.489699**	**0.474789**	**0.498542**	**0.492627**	**0.489699**	
Variance	0.182313	0.175441	0.186877	0.193235	0.182313	
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Sample	12.1902	1	12.1902	2250.748	**1.37E−51**	3.990924
Columns	0.005519	3	0.00184	0.339673	**0.796696**	2.748191
Interaction	0.006889	3	0.002296	0.423973	**0.736459**	2.748191
Within	0.346628	64	0.005416			
Total	12.54923	71				

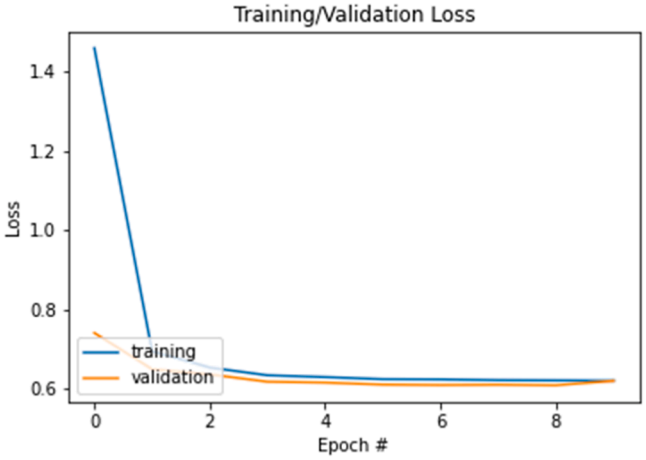


(a) 40-epoch training using Adadelta and learning rate of 1e-06



(b) 10-epoch training using Adadelta and learning rate of 1e-05

(c) 10-epoch training using Adam and learning rate of 1e-05



(d) 40-epoch training using Adam and learning rate of 1e-04

Fig. 9. Convergence curves for the different hyperparameters investigated on the CNN architecture to understand the suitable hyperparameters to use for the feature extraction task.

datasets have the hybrid binary algorithms rotating or sharing the superiority. The PenglungEW datasets showed that the BOESA was returning best classification accuracy. In summary, the hybrids are more impressive in their classification accuracy performance compared with BEOSA. Furthermore, the medium datasets showed that for CongressEW and Vote datasets, all algorithms are competing well. But for the Lymphography, SpectEW, and Zoo, we observed that HBEOSA-DMO-NT and HBEOSA-PSO-NT were more outperforming in classification accuracy compared with others. For the low-dimensional datasets, the BEOSA returned best for Exactly and Iris datasets, while the HBEOSA-PSO was best for M-of-n. The HBEOSA-DMO-NT and HBEOSA-PSO-NT competes Wine and Tic-tac-toe.

The runtime required by each algorithm was evaluated based on each category of the datasets used for the experimentation over the number of iterations. Using the bar chart in Fig. 8, the high-dimensional datasets, the highest computational time was demanded by the WaveformEW dataset, and followed by the BreastEW, Colon KrVsKpEW, and Leukemia whose runtime was in the average of about 2500 seconds. The Ionosphere, PenglungEW, and Sonar datasets utilized the lowest computational time among those in the category of high-dimensional datasets. The BOESA remained very insignificant based on computational time on the bar charts, though among the hybrid binary optimization algorithms, the HBEOSA-DMO-NT showed efficiency with respect to computational time among all the high-dimensional datasets. The HBEOSA-PSO and HBEOSA-PSO-NT were the costliest based on computational time for the high-dimensional datasets. Interestingly, for the medium-dimensional datasets, the HBEOSA-DMO were the costliest while the HBEOSA-DMO-NT were runtime efficient. Similarly, the HBEOSA-DMO-NT remained very efficient for computational time in the category of low-dimensional datasets. For this same low-dimensional category, the HBEOSA-DMO, HBEOSA-PSO and HBEOSA-PSO-NT methods were least cost effective considering their computational runtime.

In Table 8, a computer resource based comparative analysis is presented which investigated and reported the proposed hybrid method with other similar non-hybrid methods which does not use the nested transfer functions. Findings from this comparative analysis showed although the complexity of the proposed method is more than those other traditional methods, it however utilized a low computational resource on average high computing infrastructure. For instance, the proposed method was experimented on a system with memory 256 GB, external drive space as much as 4TB-960GB, and 2 extra graphical processing units (GPU) with each having 32 GB space. On the other hand, experimentation of the traditional single binary optimizers with no use of the proposed nested transfer functions, was conducted on a system CPU 1.70 GHz, 2.40 GHz; RAM of 16 GB; 64-bit Windows 10 OS.

Table 12
Performance of the trained CNN on prediction task using the test data partition.

Accuracy	Precision	Recall	F1-Score
0.881491	0.881491	0.881491	0.881491

Results obtained and listed in the table indicates that the computational cost of the proposed method will be high on low-resourced computing infrastructure while it will be insignificant on systems with average computational power.

Datasets were selected from the high-dimensional, medium-dimension and low-dimensional categories to compare across the hybrid binary method with nested transfer functions and the single binary method with NO nested transfer functions. Results reported in the table showed that there is an average 4-fold reduction in computational cost in running the approach proposed on systems with more processing power and memory.

The existing variation of computational time among the algorithms notwithstanding, we observed that the marginal differences are insignificant compared to performance improvements obtained using these algorithms for the classification accuracy, fitness function, and cost functions. It demonstrates that if we desire this high level of performance in addressing real-life optimization problems, it will be at the cost of computational time – a tradeoff that is worth the benefit of solving a challenging optimization.

5.2. Statistical analysis of the BEOSA and the HBEOSA methods

In the previous subsection, the focus of the evaluation was on the binary optimization algorithms with respect to their performances on standard metrics. However, it is also important to comparatively evaluate the statistical significance of their performance. As a result, in this subsection, the Analysis of Variance (ANOVA) test is applied to statistically analyze the difference between the means of all the five binary optimization methods. The one-way and two-factor ANOVA analysis were investigated to check the statistical differences among the means of the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA methods. In the two analyses, a null hypothesis is being tested: to determine if there is no significant difference among the means of the five binary optimization methods. To investigate this, we focused on using the fitness and classification accuracy of all the methods.

The single-factor or the one-way ANOVA test results obtained are listed in Tables 9 and 10 for the summary statistics and the ANOVA about the five binary optimization methods respectively. Using a p-value of 0.05, the summary table shows that for the fitness values over the 100-population size, the mean strengths range between 0.068305 for the BEOSA to the highest being 0.105155 for HBEOSA-DMO. The value of 0.081545 which was returned as the average mean for HBEOSA-DMO-NT is very close in range with that of HBEOSA-PSO-NT – a confirmation that the NT-variant of the hybrid binary optimizers may not really have a significant statistical difference between their mean. For the classification accuracy based on the 100-population size, the lowest mean is reported for 0.844037 while the highest remains 0.916949. This also shows that there is a significant statistical difference among the means of all the five binary optimization algorithms.

In the ANOVA table, the p-value of 0.886557 for fitness on 100-population is significantly higher than the 0.05 which we used. This therefore allows for accepting the null hypothesis to confirm that there is no significant difference in the mean-values of all the five binary optimization algorithms. Considering the p-value returned for the classification accuracy on the 100-population size, the value of 0.3581 is significantly higher than the value of 0.05 we used. This confirms that there is no significant statistical difference among the mean values of all the binary optimization methods with respect to the classification accuracy they reported.

The two-factor ANOVA analysis performed for the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA methods is reported on Table 11. In this two-factor analysis, the aim is to investigate if there are not any mean differences among the five binary optimization methods studied in this research based on the combination of their classification accuracy and fitness function value. The outcome of these two-factor ANOVA tests is divided into the summary of the analysis and the ANOVA analysis result. The values of 0.489699, 0.474789, 0.498542, 0.492627, and 0.489699 were obtained for the HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA respectively on the two-factor consideration for accuracy-fitness factors. The average mean for the HBEOSA-DMO, and BEOSA are similar when compared to those for the other three methods. Interestingly, we observed that the average mean for the HBEOSA-DMO was the lowest and far removed from the range of the remaining methods. This demonstrates that while looking at the performance of HBEOSA-DMO based on individual dataset, we assumed that it was yielding better performance, however here we understand from the analysis that a two-facto consideration of the performance reveals a lag.

The two-way ANOVA analysis shows that the p-values obtained are 1.37E-51, 0.796696, and 0.736459. Interestingly, the largest of these p-values is 0.796696, a value higher than the 0.05 applied for the analysis. Since the p-value of the interaction is not closer to 0.05, we may conclude that probably one or more of the fitness function value or the classification accuracy contributed to the improved performance reported for the methods investigated in this study. Now with this two-factor ANOVA result, and the one-factor ANOVA result discussed earlier, it is evident that the performance of the five binary optimization methods is relevant and has a real-life applicability.

5.3. Application of the HBEOSA methods to medical image analysis

In this subsection, we investigate the applicability of the hybrid binary optimization algorithms and the basic binary optimizers. The aim is to understand how efficient binary optimization algorithms can outperform or compete with continuous optimization algorithms. Experimentation design was aimed at optimizing the features extracted through the convolutional-pooling blocks using the binary optimizer and the related hybrid binary optimization algorithms. In Fig. 9(a-d), the convergence curves for the 1earning rate 1e-06 on 40 epochs using Adadelta, 1earning rate 1e-05 on 10 training epochs using Adadelta,

Table 13
Comparative analysis of the classification accuracy, precision, recall, F1-score, and AUC performances of features extracted when applied to KNN, RF, MLP and DTree algorithms for HBEOSA-DMO.

Stage	Classifier	Accuracy	Precision	Recall	F1-score	AUC
Pre-optimization	KNN	0.6623	0.85	0.94	0.89	0.7539
	RF	0.6731	0.73	0.67	0.84	0.7818
	MLP	0.6666	0.67	0.99	0.80	0.7918
	DTree	0.6147	0.72	0.92	0.81	0.7505
Post-optimization	KNN	0.7002	0.88	0.97	0.92	0.7954
	RF	0.8252	0.97	0.83	0.99	0.8291
	MLP	0.6743	0.76	0.69	0.81	0.7922
	DTree	0.6415	0.73	0.96	0.83	0.7591

Table 14
Comparative analysis of the classification accuracy, precision, recall, F1-score, and AUC performances of features extracted when applied to KNN, RF, MLP and DTree algorithms for HBEOSA-DMO-NT.

Stage	Classifier	Accuracy	Precision	Recall	F1-score	AUC
Pre-optimization	KNN	0.6277	0.85	0.92	0.88	0.7656
	RF	0.6883	0.92	0.97	0.94	0.7888
	MLP	0.6731	0.68	0.67	0.81	0.7904
	DTree	0.6341	0.74	0.95	0.83	0.7506
Post-optimization	KNN	0.7002	0.88	0.97	0.92	0.7954
	RF	0.8286	0.97	1.00	0.98	0.8280
	MLP	0.6743	0.68	0.68	0.81	0.7922
	DTree	0.6415	0.73	0.96	0.83	0.7592

Table 15
Comparative analysis of the classification accuracy, precision, recall, F1-score, and AUC performances of features extracted when applied to KNN, RF, MLP and DTree algorithms for HBEOSA-PSO.

Stage	Classifier	Accuracy	Precision	Recall	F1-score	AUC
Pre-optimization	KNN	0.6277	0.85	0.92	0.88	0.7656
	RF	0.6905	0.92	0.97	0.94	0.7832
	MLP	0.6732	0.68	0.67	0.68	0.7905
	DTree	0.6342	0.74	0.95	0.83	0.7520
Post-optimization	KNN	0.7002	0.87	0.97	0.92	0.7954
	RF	0.8212	0.97	1.00	0.98	0.8284
	MLP	0.6743	0.68	1.00	0.81	0.7922
	DTree	0.6415	0.73	0.96	0.83	0.7591

Table 16
Comparative analysis of the classification accuracy, precision, recall, F1-score, and AUC performances of features extracted when applied to KNN, RF, MLP and DTree algorithms for HBEOSA-PSO-NT.

Stage	Classifier	Accuracy	Precision	Recall	F1-score	AUC
Pre-optimization	KNN	0.6536	0.86	0.93	0.89	0.7605
	RF	0.6731	0.72	0.67	0.83	0.7761
	MLP	0.6688	0.67	0.67	0.80	0.7932
	DTree	0.6342	0.69	0.95	0.80	0.7494
Post-optimization	KNN	0.7002	0.88	0.97	0.92	0.7954
	RF	0.825	0.96	1.00	0.98	0.8288
	MLP	0.6743	0.68	1.00	0.81	0.7922
	DTree	0.6415	0.73	0.96	0.83	0.7592

1earning rate 1e-05 on 10 training epochs using Adam, and 1earning rate 1e-04 on 10 training epochs using Adam are listed. The 40-epoch training showed that the model was learning the feature representations so well such that curves for both the training and validation dataset samples were aligning from epoch 3–40. To verify if a reduced learning rate will improve learning and speed the process, we applied a learni9ng rate of 1e-05 on the same Adadelta through 10-epcohs. Since there was not much difference, the experimentation was further improved by investigating how the Adam optimizer will perform on 1e-05 and 1e-04 learning rates. Results returned as plotted in the figures indicate that the CNN architecture is suitable for the feature extraction task.

In Table 12, a summary of the performance of the CNN model is outlined. The aim is to evaluate the effectiveness of the model for effectively extracting the necessary features required for initialization of the optimization phase. Results obtained showed that the value of 0.881491 for classification accuracy, precision, recall, and F1-score were returned when the model was applied for prediction on the unseen dataset. The implication of these performance is that the current classification capability of the trained model is desirable and useful to progress the next phase of the study.

Experimentation for the feature selection and reduction using the hybrid binary optimizers was carried out using for different classifiers. Recall that the CNN model has been used for feature extraction while the vectorized features have been optimized for selection using the binary hybrid methods. Therefore, the random forest (RF), KNN, multilayer perceptron (MLP), and Decision Tree (DT) were applied as classifiers during this experiment. The K-fold value of 5 was used with each of the classifier on the test data when the training of the classifiers has been completed. The Random Forest classifier was trained using the *n_estimators*=300, Decision Tree classifier was trained with the

Table 17
Comparative analysis of the classification performance of the trained CNN using the Softmax, KNN, RF, MLP and DTree algorithms for HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA methods.

Classifier	**HBEOSA-DMO**		**HBEOSA-DMO-NT**		**HBEOSA-PSO**		**HBEOSA-PSO-NT**	
	Acc	**AUC**	**Acc**	**AUC**	**Acc**	**AUC**	**Acc**	**AUC**
KNN	0.7002	0.7954	0.7002	0.7954	0.7002	0.7954	0.7002	0.7954
RF	***0.8252***	**0.8291**	**0.8286**	***0.8280***	***0.8212***	***0.8284***	***0.825***	***0.8288***
MLP	0.6743	0.7922	0.6743	0.7922	0.6743	0.7922	0.6743	0.7922
DTree	0.6415	0.7592	0.6415	0.7592	0.6415	0.7591	0.6415	0.7592
Softmax	0.794	-	0.708	-	0.794	-	0.794	-

Table 18
Performance-based comparative analysis of the proposed method with methods using continuous optimization algorithms for neural network hyperparameter optimization on digital mammography.

Ref, Year	Approach	Category of Optimizer	Performance
[22], 2023	QLESCA for extraction of optimal feature subsets from the high dimensional space resulting from a shallow conventional neural network (SCNN) on X-ray samples.	Continuous optimization (metaheuristic) algorithm	Classification accuracy 97.8086 %,
[23], 2021	Investigated the role of Manta Ray Foraging based Golden Ratio Optimizer (MRFGRO) a continuous metaheuristic algorithm for feature selection through reduction of the noisy features	Continuous optimization (metaheuristic) algorithm	Classification accuracy 95.59 %,
[25], 2023	Investigated the use of PSO, CSO, FFA, BA, FPO, WOA, ASOA, HHOA, BOA, and GWO with CNN on feature selection.	Continuous optimization (metaheuristic) algorithm	Classification accuracy 95.7 %,
[27], 2022	FMCSA was used with CNN	Continuous optimization (metaheuristic) algorithm	F1-score = 87.5 %,
[28], 2024	Combined the Modified Cosine Similarity (MCS) algorithm with CNN for a CBMIR)	Continuous optimization (metaheuristic) algorithm	Accuracy=0.916082, precision=0.908542, sensitivity=0.923785, specificity=0.908481, F1-score=0.887539
[29], 2024	medical image segmentation can be improved using KOA and neural network on chest X-ray	Continuous optimization (metaheuristic) algorithm	accuracy=94.88 %, specificity=96.57 %, precision=95.40 %, and recall=95.40 %
[68], 2023	The SqueezeNet neural network model was applied for feature extraction while AOA selection of best hyperparameter combination.	Continuous optimization (metaheuristic) algorithm	Classification accuracy of 96.48 %
[69], 2023	The VGG16 architecture was applied for feature extraction. Then a hybrid of Social Ski-Driver (SSD) algorithm and Adaptive Beta Hill Climbing were applied to find the discriminant features for further classification using KNN.	Continuous optimization (metaheuristic) algorithm	Classification accuracy of 96.07 %
[70], 2023	The combination of PSO, dragon-fly optimization algorithm (DFOA), and crow-search optimization algorithm (CSOA) were applied for selecting optimal features from total features extracted using neural network based on transfer learning	Continuous optimization (metaheuristic) algorithm	Classification accuracy of 84.35 %
[26], 2022	The immunity-based Ebola optimization search algorithm (IEOSA) was applied to find the optimal features subset from the complete feature set extracted using a custom CNN architecture	Continuous optimization (metaheuristic) algorithm	F1-score and recall obtained are 0.58 and 0.72 for KNN.
[71], 2021	Features extracted using the combination of gray-level cooccurrence matrix (GLCM) and discrete wavelet transform (DWT) were further refined using Advanced Thermal Exchange Optimizer (ATEO)	Continuous optimization (metaheuristic) algorithm	Classification accuracy 93.79 %, sensitivity of 96.89 %, and specificity of 67.7 %
[72], 2023	Enhanced Ant Colony Optimization (EACO) was applied for finding optimal combination of hyperparameter of ResNet101	Continuous optimization (metaheuristic) algorithm	Classification accuracy of 98.63 %, sensitivity of 98.76 %, and specificity of 98.89 %
[73], 2023	The AlexNet architecture was applied for feature extraction, while the feature optimization was achieved using the Advanced Al-Biruni Earth Radius (ABER) optimization algorithm	Continuous optimization (metaheuristic) algorithm	average classification accuracy is 97.95 %
[74], 2022	Improved marine predators algorithm (IMPA) was applied to find the best selection and combination of ResNet50 hyperparameters	Continuous optimization (metaheuristic) algorithm	classification accuracy of 98.32 % accuracy, sensitivity of 98.56 %, and specificity of 98.68 %
[75], 2022	The GA, WOA, multiverse optimizer (MVO), satin bower optimization (SBO), and life choice-based optimization (LCBO) were applied for fine-tuning the weights, biases, and hyperparameters of a custom CNN architecture.	Continuous optimization (metaheuristic) algorithm	The best classification accuracy of 86.0 % was reported
[76] 2024	BEOSA was applied to optimize extracted features of a twin neural network	Binary optimization (metaheuristic) algorithm	Classification accuracy of 95.2 %
This study	**Hybrid binary optimization algorithms were applied for finding optimal subset of features from feature sets extracted using custom CNN architecture**	**Binary optimization algorithm**	**Classification accuracy of 0.8286, precision of 0.97, recall of 0.83, F1-score of 0.99, and AUC of 0.8291**

maximum depth = 2, the MLP classifier was trained using the parameters alpha=0.001, hidden_layer_sizes=(500,200,100), max_iter=200, and random_state=4. All these hyperparameters constitute the training of the classifiers on the selected and reduced feature set. Results were obtained by evaluating the optimized features with the RF, KNN, MLP and the softmax classifiers. The results from the experimentation are analysed and comparatively compared for discussion to understand how the hybrid binary optimization algorithms can perform in solving the medical image analysis. Here, we seek to understand how efficient the binary optimization methods over the continuous optimization methods are in selecting the discriminant features from digital mammography to ensure a good classification result is obtained. Presentation and discussion of the results obtained here are made to establish the difference between the pre-optimization phase, and the after effect when the optimization methods have been applied.

In Table 13, the results for the classification accuracy, precision, recall, F-score, and area under the ROC (AUC) are outlined for the application of HBEOSA-DMO to the feature selection problem. Meanwhile, note that the interpretation to these metrics follows: accuracy implies the percentage of the predictions that are correct; precision is computes the percentage of real positives samples that our model correctly predicted; recall implies the percentage of positive predictions made that are actually correct; F1-score is the weighted average of the values for precision and recall for the predictions; and AUC showing the performance of the model as a value near 1.0 represents good performance. So, in the table, the performance of HBEOSA-DMO showed that the pre-optimization phase had the best classification accuracy of 0.6731, best precision is 0.85, best recall is 0.99, best F1-score 89 and the best AUC is 0.7918. however, after the optimizer method was applied to the feature selection process, result obtained showed that best classification accuracy of 0.8252, best precision is 0.97, best recall is 0.97, best F1-score 99 and the best AUC is 0.8291. The implication of this performance is that there is now a classification accuracy performance gain of 0.1521 because of the proposed approach of using the hybrid binary of BEOSA with DMO and with the nested function.

In a similar manner, the performance of HBEOSA-DMO-NT was investigated and reported in Table 14. The aim of these second investigation is to compare the result of HBEOSA-DMO and HBEOSA-DMO-NT to understand the relevance of the nested function when applied to the problem of feature reduction and selection. Results obtained showed that prior to using the HBEOSA-DMO-NT method for the optimization the feature selection process, recall and precision yielded 0.97 and 0.85 being the best, while maximum values for F1-score, AUC, and classification accuracy obtained are 0.94, 0.7904, and 0.6883. However, it was

Table 19
Comparative analysis of the classification accuracy, precision, recall, F1-score metrics with other benchmark related studies.

References	Accuracy	Precision	Recall	F1-score	AUC	Sensitivity	Specificity
[29]	0.948	0.954	**0.954**	-	-	-	0.966
[26]	-	-	0.720	0.580	-	-	-
[28]	0.916	0.906		0.887		0.924	0.906
[71]	0.938	-	-	-	-	**0.969**	0.677
[74]	**0.983**	-	-	-	-	0.986	0.987
[76]	0.952	-	-	-	0.638	-	-
This study	0.825	**0.96**	0.93	**0.98**	**0.8288**	0.93	-

interesting to discover that the application of the HBEOSA-DMO-NT showed that maximum performance obtained for recall, precision, F1-score, AUC, and classification accuracy are 1.0, 0.97, 0.98 0.8280, and 0.8286 respectively. This demonstrates that performance gain with respect to classification and AUC are 0.1403 and 0.0376 respectively.

Therefore, comparing the performance of HBEOSA-DMO-NT and HBEOSA-DMO, we found that there is a performance gain of 0.0118 based on classification accuracy. This demonstrates that the use of the nested function in feature selection on the hybrid solution is much beneficial than the basic binary hybrid method.

Furthermore, we investigated the performance of HBEOSA-PSO with HBEOSA-PSO-NT to understand if the use of the proposed nested function is supportive of improving classification accuracy, precision, recall, F1-score, and AUC. To understand this further, Table 15 is provided to outline the results obtained for the HBEOSA-PSO approach. Pre-optimization performance of this method shows that precision, recall, and F1-score are high though the best of these stood at 0.92, 0.97, and 0.88 respectively. The classicisation accuracy and the AUC at this stage also reported their best as 0.6905 and 0.7905. However, after the optimization method was applied to select the best representation of feature subset required for the classifier, we observed some measure of improvements. The result showed that the best for classification accuracy, precision, recall, F1-score, and AUC are 0.8212, 0.97, 1.00, 0.98 and 0.8284. this indicates that there was performance gain of 0.1307, 0.05, 0.05, 0.04, and 0.0379 for classification accuracy, precision, recall, F1-score, and AUC respectively due to the use of HBEOSA-PSO. The implication of this is that the combination of the hybrid with the proposed nested function improved performance across all metrics.

The result obtained for HBEOSA-PSO-NT for all metrics are listed in Table 16 for classification accuracy, precision, recall, F1-score, and AUC. The comparative analysis of the results with respect to the pre-optimization and post-optimization phases showed that there was also an improvement. For instance, performance for classification accuracy, precision, recall, F1-score, and AUC after the optimization returned the best values of 0.825, 0.96, 1.0, 0.98, and 0.8288 respectively. Computing the difference between this performance and those obtained prior to the application of the HBEOSA-PSO-NT, result shows performance gains of 0.1519, 0.1, 0.07, 0.9, and 0.0356 for classification accuracy, precision, recall, F1-score, and AUC respectively.

Comparing the performance of HBEOSA-PSO with HBEOSA-PSO-NT, we discovered that the HBEOSA-PSO which uses the nested function outperformed the latter across all the metrics computed. Furthermore, the performances of HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, and HBEOSA-PSO-NT were comparatively evaluated. Outcome of this comparison showed that the other of performance follows this sequence with the most performing method being the first and the least performing being the last: HBEOSA-PSO, HBEOSA-DMO, HBEOSA-PSO-NT and HBEOSA-DMO-NT. Overall the improvement observed for PSO-based hybrid over the DMO-based hybrid showed that a classification accuracy difference stood at 0.1401.

In Table 17, the summary of the classification accuracy and the area under curve (AUC) metrics are presented for the RF, KNN, MLP, and softmax classifiers for each optimizer namely: HBEOSA-DMO, HBEOSA-DMO-NT, HBEOSA-PSO, HBEOSA-PSO-NT, and BEOSA methods. The comparison reported in this table is to enable differentiating performances of all the classifiers applied. This is necessary to help guide the choice of classifier in future studies. We discovered that the random forest and the softmax classifiers showed significant performance increase compared to KNN, MLP and decision tree.

To further understand how relevant the use of the binary metaheuristic algorithms over the popular continuous metaheuristic algorithm for selection of optimal medical image features is, we compare the result obtained in this study with those recent related methods. In Table 18 the listing for the performance comparison with the state-of-the-art methods is based on the approach used, the type of optimizer used, and the result obtained. Note that all the related studies compared here are all based on digital mammography datasets.

The related works compared with this study may be categorized into two: studies optimizing the selection and combination of hyperparameters, weights, and biases of neural networks [22,23,25,27-29,68,72,74,75], and studies optimizing the features sets extracted using neural networks to obtain an optimal feature subsets [26,69-71,73], and [76]. Most of the studies reported the classification accuracy obtained using their approaches, and so, we compare using this metric. The classification accuracy performances obtained in [22,23,25,28], and [29] showed that these values range from 95.59 to 97.80 percent for different datasets such as chest X-ray and brain MRI. Results reported for [68,72,74], and [75] are 96.48 %, 98.63 %, 98.32 %, and 86.0 % classification accuracies respectively. These are the studies in the category of optimizing hyperparameters of neural networks. Compared with the result obtained in this study, approaches using metaheuristic algorithms to optimize neural network architectures outperformed our method which applied metaheuristic algorithm to optimize features sets extracted using neural network architectures. While these two approaches are different, we further compare the results obtained by approaches using continuous metaheuristic algorithms for optimizing features sets with those binary metaheuristic algorithms solving the same problem. Classification results reported for [26,69-71,73], and [76] are 96.07 %, 84.35 %, 72.0 % (recall), 93.79 %, 97.95 %, and 95.2 % respectively. The best classification accuracy for the second category is 97.95 % from the [73] study. In Table 19 we provide an outline that summarizes performance based on benchmark metrics, and have compared the performance of this study with related studies.

Comparing this best result with what is obtained in this study, we observed a very good competition between the approaches using continuous optimization algorithm and our approach using binary optimization method for finding optimal feature subset. Furthermore, we compared our result with the related work which also uses binary optimization method as described in this study, though the hybrid binary methods are investigated in this study. The result of the comparison showed that the hybrid binary optimization methods are also very competitive with those of the basic binary optimizer.

5.4. *Discussion of findings and limitation of the current approach*

Finding from this study confirmed that the binary optimization methods, particularly the hybrid binary methods, are very useful for finding optimal subset of features from those extracted using neural network architectures. The result also demonstrates that contrary to the popular methods of using only continuous metaheuristic (optimization)

algorithms for solving this same problem, the binary metaheuristics (optimization) algorithm are very suitable and applicable. Most importantly, the outcome from the study showed that formulation of the search space for the binary optimization algorithm can be achieved in a unique way as demonstrated in this study. Therefore, it is confirmation that the binary optimization algorithm is suitable for optimizing the selection and combination of hyperparameters, weights, and biases of neural networks.

Considering the findings and performance rsulting from the result analysis in previous subsections, the following highlights the main contributions of this study thereby making it the state-of-the-art in this domain. First, the novel approach applied to adapt the feature sets extracted by neural network into useful search space of binary optimizers is new. This also allows for rerepresenting these optimzed feature sets into formalism useful for the classifier. An exhaustive experimentation was performed to evaluate the four variants binary optimizers, HBEOSA-DMO HBEOSA-DMO-NT, HBEOSA-PSO and HBEOSA-PSO-NT. These new hybrid binary optimizers are now available for use among researchers for addressing similar problems in medical imaging.

The approach proposed in this study has demonstrated good performance and is representative of a novel adaptation of hybrid binary optimization algorithms to feature selection in medical image analysis. However, the role of transform function in binary optimization method is very important in converting the continuous search space into binary search space. Hence, it is desirable to investigate the current limitation of the proposed nested transfer functions so that the impediment resulting from the fundamental mathematical approach might be addressed for better performance. This is necessary because this study was not focused on comparative analysis of the influence of the nested transfer functions on the medical image feature selection and reduction in neural network architecture. Rather, it was focused on applying the new hybrid binary method and evaluating or comparatively analyzing the role of the whole hybrid method.

6. Conclusion

In this study, the challenge of finding the best optimal subset of features from those extracted using convolutional neural network architectures was addressed. The approach presented in the study is based on examining the influence of four related hybrid binary optimization algorithms for the search and selection of these optimal feature subsets. Meanwhile, the hybrid binary algorithms were first improved using some nested transform functions proposed in the study. An exhaustive experimentation was carried out on these optimizers to ascertain their usefulness in addressing the medical image feature selection problem. Furthermore, an approach to support the formulation of the CNN-based extracted feature space into a binary search space for the adaptation of the hybrid binary methods, was also introduced in the study. This is motivated by the difficulty of handling very high dimensional data representation of medical images, and the need to ensure that non-relevant features are eliminated so that only the discriminants are left for the classifier. It is important to address this problem to ensure that false positives and false negatives are reduced in the performances of neural networks detecting and characterizing abnormalities in medical image analysis. As proof of concept, this study investigated the applicability of the approach proposed on digital mammography, specifically the MIAS dataset. Meanwhile, the first part of the methodology and experimentation was focused on the design of the novel hybrid binary optimization algorithms which applies a new transfer function. The performance of the hybrid binary optimization algorithm was investigated through an exhaustive experimentation using several datasets. The evaluation of the hybrid binary metaheuristic algorithms was investigated and analyzed using metrics such as the classification accuracy, fitness function value, cost function value, number of feature count, and the computational runtime. Results obtained when the new hybrid binary optimizers were applied to medical image feature selection, showed that classification accuracy of 0.8286, precision of 0.97, recall of 0.83, and F1-score of 0.99, AUC of 0.8291. Findings from this evaluation showed that the hybrid binary metaheuristic algorithms yielded better performances in respect to the classification accuracy, fitness, and cost function values. However, we found the basic binary metaheuristic algorithm BEOSA was returning best computational time, lower than they hybrid methods, and as well yielded the minimal feature counts. Considering the impressive performances of the hybrid binary optimizers, we proceeded with experimentation to apply the method to search and select the optimal subsets of features sets from the extracted features in batches of medical image samples supplied into the neural network architecture. Result obtained showed that the hybrid binary optimization algorithms successfully selected the required discriminant subset of features sufficient to help the classifier return best classification accuracy. In addition to the findings in this study, from the study showed that contrary to the popular approach of using continuous metaheuristic algorithms for addressing the same problem, the binary metaheuristic algorithms are well suitable for handling the challenge. Comparative analysis of the study with the related works revealed that the approach proposed in this study can be investigated as a future work for optimizing the selection and combination of hyperparameters, weights, and biases of neural networks. The new hybrid binary optimizers HBEOSA-DMO HBEOSA-DMO-NT, HBEOSA-PSO and HBEOSA-PSO-NT represents new set of algorithms which can be applied for optimization tasks in neural networks. Future studies might consider applying these new methods for finding the best combination of hyperparameters in neural networks. Other considerations are to discover and design a new approach to adapting these new methods to the difficult challenge of neural architecture search (NAS). Moreover, the new hybrid binary methods were only applied to digital mammography samples. Studies have shown that different imaging methods presents varied challenges in detection and characterization of abnormalities represented in them. As a result, future work might consider seeking the implication of using these new hybrid solutions on samples from other medical imaging techniques. This will provide sufficient information on the applicability of the approach in medical image and signal analysis, thereby allowing for extending them to other domains.

CRediT authorship contribution statement

Hui Wang: Writing – review & editing, Writing – original draft, Supervision. **Enesi Femi Aminu:** Writing – review & editing, Writing – original draft, Conceptualization. **Olaide Oyelade:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Karen Rafferty:** Supervision, Writing – review & editing.

Declaration of Competing Interest

Authors declare that there is no conflict of interests.

Data availability

Data will be made available on request.

References

[1] Y. Zhao, X. Wang, T. Che, G. Bao, S. Li, Multi-task deep learning for medical image computing and analysis: a review, Comput. Biol. Med. (2023).
[2] P.K. Mall, P.K. Singh, S. Srivastav, V. Narayan, M. Paprzycki, T. Jaworska, M. Ganzha, A comprehensive review of deep neural networks for medical image processing: recent developments and future opportunities, Healthc. Anal. (2023).
[3] X. Xie, J. Niu, X. Liu, Z. Chen, S. Tang, S. Yu, A survey on incorporating domain knowledge into deep learning for medical image analysis, Med. Image Anal. (2021).
[4] M. Kolahdoozi, A. Amirkhani and M. Maroufi, "Fuzzy Cognitive Maps and a New Region Growing Algorithm for Classification of Mammography Images," in *2017*

24th National and 2nd International Iranian Conference on Biomedical Engineering (ICBME), 2017.
[5] J. Mozaffari, A. Amirkhani, S.B. Shokouhi, A survey on deep learning models for detection of COVID-19, Neural Comput. Appl. 35 (2023) 16945–16973.
[6] B. Sistaninejhad, H. Rasi, P. Nayeri, A review paper about deep learning for medical image analysis, Comput. Math. Methods Med. (2023).
[7] K. He, C. Gan, Z. Li, I. Rekik, Z. Yin, W. Ji, D. Shen, Transformers in medical image analysis, Intell. Med. 3 (1) (2023) 59–78.
[8] J. Mozaffari, A. Amirkhani, S.B. Shokouhi, ColonGen: an efficient polyp segmentation system for generalization improvement using a new comprehensive dataset, Phys. Eng. Sci. Med. 47 (2024) 309–325.
[9] A. Taleb, C. Lippert, T. Klein and M. Nabi, "Multimodal self-supervised learning for medical image analysis.," in *In International conference on information processing in medical imaging.*, 2021.
[10] S. Budd, E.C. Robinson, B. Kainz, A survey on active learning and human-in-the-loop deep learning for medical image analysis, Med. Image Anal. 71 (2021).
[11] M.A. Mazurowski, H. Dong, H. Gu, J. Yang, N. Konz, Y. Zhang, Segment anything model for medical image analysis: an experimental study, Med. Image Anal. 89 (2023).
[12] O.N. Oyelade, A.E.S. Ezugwu, A state-of-the-art survey on deep learning methods for detection of architectural distortion from digital mammography, IEEE Access (2020) 148644–148676.
[13] M. Becerra-Rozas, J. Lemus-Romani, F. Cisternas-Caneo, B. Crawford, R. Soto, G. Astorga, J. García, Continuous metaheuristics for binary optimization problems: an updated systematic literature review, Mathematics (2022).
[14] O.A. Akinola, A.E. Ezugwu, O.N. Oyelade, J.O. Agushaka, A hybrid binary dwarf mongoose optimization algorithm with simulated annealing for feature selection on high dimensional multi-class datasets, Sci. Rep. (2022).
[15] A.M. Khalid, H.M. Hamza, S. Mirjalili, K.M. Hosny, Knowledge-based systems BCOVIDOA: a novel binary coronavirus disease optimization algorithm for feature selection, Knowl. -Based Syst. (2022).
[16] H. Hichem, M. Elkamel, M. Rafik, M.T. Mesaaoud, C. Ouahiba, A new binary grasshopper optimization algorithm for feature selection problem, J. King Saud. Univ. Comput. Inf. Sci. (2022) 316–328.
[17] S.S. Shekhawat, H. Sharma, S. Kumar, A. Nayyar, B. Qureshi, bSSA: binary salp swarm algorithm with hybrid data transformation for feature selection, IEEE Access (2021) 14867–14882.
[18] O. Oyelade, A. Ezugwu, T. Mohamed, L. Abualigah, Ebola optimization search algorithm: a new nature-inspired metaheuristic optimization algorithm, IEEE Access (2022) 16150–16177.
[19] Z. Tian, S. Fong, Survey of meta-heuristic algorithms for deep learning training, Optim. Algorithms—Methods Appl. (2018).
[20] M. Abdel-Basset, L. Abdel-Fatah, A.K. Sangaiah, Metaheuristic algorithms: a comprehensive review, Comput. Intell. Multimed. big data cloud Eng. Appl. (2018) 185–231.
[21] M. Kaveh, M.S. Mesgari, Application of meta-heuristic algorithms for training neural networks and deep learning architectures: a comprehensive review, Neural Process. Lett. (2023) 4519–4622.
[22] Q.S. Hamad, H. Samma, S.A. Suandi, Feature selection of pre-trained shallow CNN using the QLESCA optimizer: COVID-19 detection as a case study, Appl. Intell. (2023) 18630–18652.
[23] A. Dey, S. Chattopadhyay, P.K. Singh, A. Ahmadian, M. Ferrara, N. Senu, R. Sarkar, MRFGRO: a hybrid meta-heuristic feature selection method for screening COVID-19 using deep features, Sci. Rep. (2021).
[24] A. Zafar, S.J. Hussain, M.U. Ali, S.W. Lee, Metaheuristic optimization-based feature selection for imagery and arithmetic tasks: an fNIRS study, Sens. (Basel) (2023).
[25] M.U. Ali, S.J. Hussain, A. Zafar, M.R. Bhutta, S.W. Lee, WBM-DLNets: wrapper-based metaheuristic deep learning networks feature optimization for enhancing brain tumor detection, Bioeng. (Basel) (2023).
[26] O.N. Oyelade, A.E. Ezugwu and A.K. Saha, "SIRO: A Deep Learning-Based Next-Generation Optimizer for Solving Global Optimization Problems," in *Metaheuristics International Conference 2024. Lecture Notes in Computer Science*, 2024.
[27] O.N. Oyelade, A.E. Ezugwua, Immunity-based Ebola optimization search algorithm for minimization of feature extraction with reduction in digital mammography using CNN models, Sci. Rep. 12 (2022).
[28] S.V. Kumari, K.U. Rani, FMCSA: metaheuristic method for optimal feature selection in medical image classification, J. XI Univ. ARCHITECTURE Technol. (2022).
[29] R. Shetty, V.S. Bhat, J. Pujari, Content-based medical image retrieval using deep learning-based features and hybrid meta-heuristic optimization, Biomed. Signal Process. Control (2024).
[30] M. Abdel-Basset, R. Mohamed, I. Alrashdi, K.M. Sallam, I.A. Hameed, CNN-IKOA: convolutional neural network with improved Kepler optimization algorithm for image segmentation: experimental validation and numerical exploration, J. Big Data (2024).
[31] P.-H. Dinh, Combining Gabor energy with equilibrium optimizer algorithm for multi-modality medical image fusion, Biomed. Signal Process. Control 68 (2021).
[32] P.-H. Dinh, A novel approach based on Grasshopper optimization algorithm for medical image fusion, Expert Syst. Appl. 171 (2021).
[33] P.-H. Dinh, Medical image fusion based on enhanced three-layer image decomposition and Chameleon swarm algorithm, Biomed. Signal Process. Control 84 (2023).
[34] T.-H.-H. Le, P.-H. Dinh, V.-H. Vu, N.L. Giang, A new approach to medical image fusion based on the improved extended difference-of-gaussians combined with the coati optimization algorithm, Biomed. Signal Process. Control 93 (2024).
[35] P.-H. Dinh, A novel approach based on three-scale image decomposition and Marine predators algorithm for multi-modal medical image fusion, Biomed. Signal Process. Control 67 (2021).
[36] O.C. Do, C.M. Luong, P.-H. Dinh, G.S. Tran, An efficient approach to medical image fusion based on optimization and transfer learning with VGG19, Biomed. Signal Process. Control 87 (2024).
[37] P.-H. Dinh, A novel approach based on marine predators algorithm for medical image enhancement, Sens. Imaging 24 (6) (2023).
[38] U.C. Aytaç, A. Güneş, N. Ajlouni, A novel adaptive momentum method for medical image classification using convolutional neural network, BMC Med. Imaging (2022).
[39] J.I. García-Mercado, G. Gutiérrez-Alcaraz, N. Gonzalez-Cabrera, Improved binary particle swarm optimization for the deterministic security-constrained transmission network expansion planning problem, Int. J. Electr. Power (2023).
[40] N. Ganesh, R. Shankar, R. Cep, S. Chakraborty, K. Kalita, Efficient feature selection using weighted superposition attraction optimization algorithm, Appl. Sci. (2023).
[41] O.O. Akinola, A.E. Ezugwu, J.O. Agushaka, R. Abu, Z. Laith, Multiclass feature selection with metaheuristic optimization algorithms: a review.,", Neural Comput. Appl. (2022) 19751–19790.
[42] P. Agrawal, T. Ganesh, A.W. Mohamed, A novel binary gaining – sharing knowledge-based optimization algorithm for feature selection, Neural Comput. Appl. (2020).
[43] A.S. DESUKY, S. HUSSAIN, S. KAUSAR, M.A. ISLAM, L.M.E. BAKRAWY, EAOA: an enhanced archimedes optimization algorithm for feature selection in classification, IEEE Access (2021).
[44] K. Zhu, S. Ying, N. Zhang, D. Zhu, Software defect prediction based on enhanced metaheuristic feature selection optimization and a hybrid deep neural network, J. Syst. Softw. (2021).
[45] W. Ma, X. Zhou, H. Zhu, L. Li, L. Jiao, A two-stage hybrid ant colony optimization for high-dimensional feature selection, Pattern Recognit. (2021).
[46] A. Li, B. Xue, M. Zhang, Improved binary particle swarm optimization for feature selection with new initialization and search space reduction strategies, Appl. Soft Comput. (2021) 568–4946.
[47] K. Hussain, N. Neggaz, W. Zhu, E.H. Houssein, An efficient hybrid sine-cosine Harris hawks optimization for low and high-dimensional feature selection, Expert Syst. Appl. (2021).
[48] L. Abualigah, A.J. Dulaimi, A novel feature selection method for data mining tasks using hybrid Sine Cosine Algorithm and Genetic Algorithm, Clust. Comput. (2021).
[49] A. Chaudhuri, T.P. Sahu, A hybrid feature selection method based on Binary Jaya algorithm for micro-array data classification, Comput. Electr. Eng. (2021).
[50] F.S. Gharehchopogh, M. Isa, Z.A. Dizaji, Chaotic vortex search algorithm: metaheuristic algorithm for feature selection, Evolut. Intell. (2021).
[51] G. Dhiman, D. Oliva, A. Kaur, K.K. Singh, S. Vimal, A. Sharma and K. Cengiz, "BEPO: A novel binary emperor penguin optimizer for automatic feature selection.," *Knowledge-Based Systems.*, 2020.
[52] M. Allam, M. Nandhini, Optimal feature selection using binary teaching learning based optimization algorithm, J. King Saud. Univ. -," *Comput. Inf. Sci.* (2022) 329–341.
[53] O.A. Akinola, J.O. Agushaka, A.E. Ezugwu, Binary dwarf mongoose optimizer for solving high-dimensional feature selection problems, PlosONE 17 (10) (2022).
[54] O. Akinola, O.N. Oyelade, A.E.-S. Ezugwu, Binary ebola optimization search algorithm (BEOSA) using novel S-V transformation functions for solving feature selection and classfication problem, Appl. Sci. (2022) 1–38.
[55] J. Li, S. Fong, L. Lian-sheng, N. Dey, A.S. Ashour, Dual feature selection and rebalancing strategy using metaheuristic optimization algorithms in X-ray image datasets (e. al. and), Multimed. Tools Appl. (2019) 20913–20933.
[56] N.K. Hussein, M. Qaraad, S. Amjad, M.A. Farag, S. Hassan, S. Mirjalili, M. A. Elhosseini, Enhancing feature selection with GMSMFO: A global optimization algorithm for machine learning with application to intrusion detection, J. Comput. Des. Eng. (2023) 1363–1389.
[57] Z. Ye, Y. Xu, Q. He, M. Wang, W. Bai, H. Xiao, Feature selection based on adaptive particle swarm optimization with leadership learning, Comput. Intell. Neurosci. (2022).
[58] R.S. Selvi, M.L. Valarmathi, P. Devadas, Improved meta-heuristic algorithm for selecting optimal features: A big data classification model, Concurr. Comput. Pract. Exp. (2022).
[59] J.L. Potharlanka, N. B. M, Feature importance feedback with Deep Q process in ensemble-based metaheuristic feature selection algorithms, Sci. Rep. (2024).
[60] J. Lee, Y. Yoon, J. Kim, Y.-H. Kim, Metaheuristic-based feature selection methods for diagnosing sarcopenia with machine learning algorithms, Biomimetics (2024).
[61] D. Połap, M. Woźniak and J. Mańdziuk, "Meta-heuristic Algorithm As Feature Selector For Convolutional Neural Networks," in *2021 IEEE Congress on Evolutionary Computation (CEC)*, 2021.
[62] D.A. Torse, R. Khanai, K. Pai, S. Iyer, S. Mavinkattimath, R. Kallimani, S. Shahpur, Optimal feature selection for COVID-19 detection with CT images enabled by metaheuristic optimization and artificial intelligence, Multimed. Tools Appl. (2023) 41073–41103.
[63] L. Meenachi, S. Ramakrishnan, Metaheuristic search based feature selection methods for classification of cancer, Pattern Recognit. (2021).
[64] S. Lee, S. Soak, S. Oh, W. Pedrycz, M. Jeon, Modified binary particle swarm optimization, Prog. Nat. Sci. 18 (9) (2008) 1161–1166.
[65] S. Granizo, M. Baldeon-Calisto, M. Iñiguez, D. Navarrete, D. Riofrío and N. Pérez-Pérez, "A Comparative Analysis of Vision Transformers and Convolutional Neural Networks in Cardiac Image Segmentation," in *2024 12th International Symposium on Digital Forensics and Security (ISDFS)*, 2024.

[66] M. Marches, "Megapixel Size Image Creation using Generative Adversarial Networks," *ArXiv*, 2017.
[67] P. Xi, C. Shu and R. Goubran, "Abnormality Detection in Mammography using Deep Convolutional Neural Networks.," *arXiv:1803.01906v1 [cs.CV]*, 2018.
[68] M. Basheri, Intelligent breast mass classification approach using archimedes optimization algorithm with deep learning on digital mammograms, Biomimetics 8 (6) (2023) 463.
[69] P. Pramanik, S. Mukhopadhyay, S. Mukhopadhyay, R. Sarkar, Deep feature selection using local search embedded social ski-driver optimization algorithm for breast cancer detection in mammograms, Neural Comput. Appl. 35 (2023) 5479–5499.
[70] S.S. Chakravarthy, N. Bharanidharan, H. Rajaguru, Deep learning-based metaheuristic weighted Kk-nearest neighbor algorithm for the severity classification of breast cancer, IRBM 44 (3) (2023).
[71] X. Cai, X. Li, N. Razmjooy, N. Ghadimi, Breast cancer diagnosis by convolutional neural network and advanced thermal exchange optimization algorithm, Comput. Math. Methods Med 2021 (2021).
[72] S. Thirumalaisamy, K. Thangavilou, H. Rajadurai, O. Saidani, N. Alturki, S. k Mathivanan, P. Jayagopal, S. Gochhait, Breast cancer classification using synthesized deep learning model with metaheuristic optimization algorithm, Diagnostics 13 (18) (2023) 2925.
[73] A.A. Alhussan, A.A. Abdelhamid, S.K. Towfek, A. Ibrahim, N. Khodadadi, D. S. Khafaga, S. Al-Otaibi, A.E. Ahmed, Classification of breast cancer using transfer learning and advanced al-biruni earth radius optimization, Biomim. (Basel) 8 (3) (2023) 270.
[74] E.H. Houssein, M.M. Emam, A.A. Ali, An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators algorithm, Neural Comput. Appl. 34 (2022) 18015–18033.
[75] O.N. Oyelade, A.E. Ezugwu, Characterization of abnormalities in breast cancer images using nature-inspired metaheuristic optimized convolutional neural networks model, Concurr. Comput. Pract. Exp. 34 (4) (2022).
[76] O.N. Oyelade, E.A. Irunokhai, H. Wang, A twin convolutional neural network with hybrid binary optimizer for multimodal breast cancer digital image classification, Sci. Rep. (2024).

Olaide Nathaniel OYELADE. Dr Olaide Nathaniel Oyelade is currently a Senior Lecturer of Computer Science and AI at the Department of Engineering, Computer and Mathematics, University of Chichester, United Kingdom. The focus of his research is the application of AI models and algorithms to lead new disruptive practices in medicine, health & medical imaging. He is particularly interested in machine learning, deep learning and knowledge graph-based natural language processing. He has years of experience working application of neural networks to medical image analysis; design and application metaheuristic algorithms to optimization of AI models and algorithms when applied to the field of medicine. He also has experience working on knowledge representation and reasoning algorithm design, hybridization of knowledge graph and large language models for drug repurposing.

Enesi Femi AMINU. Dr Enesi Femi Aminu is a Senior Lecturer and is affiliated to the Department of Computer Science, Federal University of Technology Minna, Nigeria. His publication topics includes Authentication, Back Propagation Neural Network, Bit Error Rate, Cloud Computing, Cognitive Complexity, Concepts In Order, Copyright Protection, Cover Image, Crop Yield, Descriptive Characteristics, Digital Content. He is passionate about research in knowledge representation and reasoning algorithm design when applied to smart agriculture and food technology. He also has research experience in machine learning and deep learning, with evolving publications in computational intelligence.

Hui WANG. Professor Hui Wang is a full professor in School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast. His research interests are machine learning, knowledge representation and reasoning, and their applications in image, video, spectra and text data analyses and more recently education. He played an important role in the development of an algebraic framework for machine learning and knowledge representation, Lattice Machine; the original concept of contextual probability for probabilistic modelling, which can be used for uncertainty reasoning/quantification, probability estimation and machine learning; a generic similarity measure, neighbourhood counting, and its specialisations on multivariate data, sequences, tree and graph structures. He is currently leading two research programmes: (1) contextual probability for decentralised learning and causal learning; (2) discovery AI to support the discovery of signatures (e.g., biomarkers), causalities and hypotheses. He has over 370 publications in these areas.

Karen RAFFERTY. Karen Rafferty is currently the Head of the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast. She has over fifteen years experience working within the fields of software engineering, sensor fusion, and real time software development and over ten years experience working within the areas of virtual and augmented reality and multi-sensorial systems. Her research interests include application of tools and technologies to lead new disruptive practices and systems for many application areas with a main focus on the Health and Training and Industry and Automation.