IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 February 2023, accepted 2 March 2023, date of publication 6 March 2023, date of current version 13 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3253430

RESEARCH ARTICLE

Advanced Meta-Heuristic Algorithm Based on Particle Swarm and Al-Biruni Earth Radius Optimization Methods for Oral Cancer Detection

HADJOUNI MYRIAM[1], ABDELAZIZ A. ABDELHAMID[2,3], EL-SAYED M. EL-KENAWY[4], (Senior Member, IEEE), ABDELHAMEED IBRAHIM[5], (Member, IEEE), MARWA METWALLY EID[6], MONA M. JAMJOOM[1], AND DOAA SAMI KHAFAGA[1]

[1]Department of Computer Sciences, College of Computer and Information Science, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[2]Department of Computer Science, College of Computing and Information Technology, Shaqra University, Shaqra 11961, Saudi Arabia
[3]Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt
[4]Department of Communications and Electronics, Delta Higher Institute for Engineering and Technology (DHIET), Mansoura 35111, Egypt
[5]Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt
[6]Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 35712, Egypt

Corresponding authors: Abdelaziz A. Abdelhamid (abdelaziz@su.edu.sa), El-Sayed M. El-Kenawy (skenawy@ieee.org), and Mona M. Jamjoom (mfhaojouni@pnu.edu.sa)

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R193), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

ABSTRACT Oral cancer is a deadly form of cancerous tumor that is widely spread in low and middle-income countries. An early and affordable oral cancer diagnosis might be achieved by automating the detection of precancerous and malignant lesions in the mouth. There are many research attempts to develop a robust machine-learning model that can detect oral cancer from images. However, these are still lacking high precision in oral cancer detection. Therefore, this work aims to propose a new approach capable of detecting oral cancer in medical images with higher accuracy. In this work, a novel and robust oral cancer detection based on a convolutional neural network (CNN) and optimized deep belief network (DBN). The design parameters of CNN and DBN are optimized using a new optimization algorithm, which is developed as a hybrid of Particle Swarm Optimization (PSO) and Al-Biruni Earth Radius (BER) Optimization algorithms and is denoted by (PSOBER). Using a standard biomedical images dataset available on the Kaggle repository, the proposed approach shows promising results outperforming various competing approaches with an accuracy of 97.35%. In addition, a set of statistical tests, such as One-way analysis-of-variance (ANOVA) and Wilcoxon signed-rank tests, are conducted to prove the significance and stability of the proposed approach. The proposed methodology is solid and efficient, and specialists can adopt it. However, additional research on a larger scale dataset is required to confirm the findings and highlight other oral features that can be utilized for cancer detection.

INDEX TERMS Oral cancer, particle swarm optimization, Al-Biruni earth radius algorithm, deep belief network, convolutional neural network, metaheuristic optimization.

I. INTRODUCTION

Oral cancer is one of the most common forms of cancer worldwide [1]. The late detection of oral cancer can increase morbidity and high fatality rates. About half of all cases occur in South Asia [2], and two-thirds of all cases occur in low- and middle-income countries. The most common causes of oral cancer are heavy drinking and smoking. Survival rates are low because two-thirds of people in low- and middle-income countries who have an oral lesion appear at a late stage [3]. It is prohibitively expensive to treat cancer, especially in its advanced stages [4]. Late diagnosis is sometimes caused by a combination of medical professionals and patients' ignorance about oral lesions. The screening program has focused

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

mostly on diagnosing oral potentially malignant disorders (OPMD) due to the danger of cancerous transformation, its enormous value in reducing mortality and morbidity from oral malignancies, and its prevalence [5]. However, using this program in real-time situations has been challenging since it relies on visual examination, and healthcare workers often lack the knowledge or training to correctly detect this lesion [6], [7].

To lower cancer-related mortality and morbidity, researchers have turned to machine learning and deep learning models to enhance detection accuracy. The value of automated image analysis in aiding pathologists and clinicians in the earlier diagnosis of oral cancer and subsequent management decisions is undeniable. Significant variability in the presence of oral cancer makes diagnosis exceedingly hard for healthcare practitioners and is a typical cause of delays in inpatient referral to oral lesion experts [8]. In addition, patients may wait longer to seek medical attention since early-stage oral cancer and OPMD lesions are often asymptomatic and appear to be tiny, benign lesions [9], [10]. Oral cancer classification and detection performance is a focus for authors in [11] and [12], who strive to optimize these processes while reducing processing time. A fully automated method for identifying oral tumors in cytology slides was described by authors in [13] and [14]. By allowing for multi-class classification and reducing over-fitting to the data, the provided method successfully reduces the error in categorizing and predicting oral cancers.

Authors developed a deep learning training model in [15] and [16], which makes its prediction clear and helps the network maintain its attention on the tumorous area of the image, allowing for more accurate segmentation of that area. To photographically classify oral lesions, authors in [17] and [18] created a deep learning architecture named D'OraCa. To improve the accuracy of oral cancer classification, this study establishes a mouth landmark identification technique for the oral image and then applies it as guidance to the classification process. For early detection of cancer of the tongue, authors in [19] and [20] assessed the performance of six deep convolutions neural network (DCNN) models using transfer learning on a subset of data. The DCNN model can tell the difference between benign and precancerous lesions on the tongue and identify the five main types of tongue cancer.

Deep learning models have recently been the focus of research toward classifying melanomas. To provide more significant features for precise melanoma detection, the authors of [21] devised a novel method based on deep CNN and feature encoding techniques. The created model achieved an accuracy of 86.54% when stored against a dataset during its training phase. To differentiate between benign oral lesions and cancerous melanoma, the authors of [22] developed a deep CNN architecture for this task. A 91.97% sensitivity, 84.76% accuracy, and 787.1% specificity were achieved using the suggested method. To distinguish between benign moles and malignant ones, the authors of [23] used a three-classifier ensemble model. In [24], the authors introduced a novel technique based on spiking neural networks with time-dependent spike plasticity. Deep convolutional neural network architecture was used by the authors of [25] to classify melanomas. The strategy mentioned above was examined by utilizing three distinct data sets. In [26], the authors employ a CNN-based technique for detecting malignant lesions (CMLD). On the MED-NODE and Dermofit datasets, respectively, the generated model obtained 90.14 and 90.58% accuracy.

Using a deep neural network model dubbed NasNet Mobile, the authors of [39] were able to identify melanoma in patients. To address the issue of unequal classes, many augmentation strategies were implemented. Using the Nasnet-Mobile network, the suggested model achieved an accuracy of 89.20% without data augmentation and 97.90% with data augmentation. To determine whether an oral lesion was melanoma or a nevus, the authors of [40] relied on a pre-trained architecture known as ResNet50. The suggested model has a sensitivity to false positives of 77.9% and a specificity to false positives of 82.3%. The ResNet152 model was used to categorize different types of oral lesions by the authors of [41]. To classify lesions, the authors of [42] swapped out the last three AlexNet layers for fully linked layers, softmax, and an output layer. 96.86% accuracy was attained by the suggested method. The researchers in [43] classified oral lesions using a pre-trained model called Inception-v3. They used augmentation methods to expand the test dataset. The suggested model performed at a 71.2% success rate in a classification task. The compiled related works are summarized in Table 1.

Based on these studies, it can be noted that the detection accuracy of classifying oral cancer using the methods proposed by authors in the literature still needs more improvement. Therefore, this paper proposes a new approach based on metaheuristic optimization to robustly classify oral cancer in medical images with higher accuracy. The proposed approach's convolutional neural network structure is optimized using a new hybrid optimization algorithm for feature extraction. In addition, the structure of the deep belief network is optimized using the proposed optimization algorithm for oral cancer detection. The proposed hybrid optimization algorithm is based on two recently emerged metaheuristic optimization algorithms: particle swarm optimization and Al-Biruni earth radius optimization algorithms. The evaluation of the proposed approach is performed in terms of several criteria to confirm its effectiveness. In addition, statistical tests such as analysis of variance and Wilcoxon signed rank tests are performed to assess the statistical difference and significance compared to other competing approaches. The recorded results confirm the proposed approach's superiority and effectiveness in oral cancer detection.

II. METHODS AND MATERIALS

The primary methods employed in this work to implement the proposed approach are detailed in this section. These methods include the convolutional neural network, which is used in

TABLE 1. Summary of the recent research in the literature.

Paper	Description	Target	Dataset size
[27]	CT images of the oral cavity. Three-dimensional CNN for binary image classification.	Early diagnosis of oral cancer based on three-dimensional convolutional neural network.	7000 images.
[28]	Histopathological images. CNN was used for 4-class image classification.	Tissue Level Based Deep Learning Framework for early detection of dysplasia in oral squamous epithelium.	2688 images from 52 individuals.
[29]	Hyperspectral images of the oral cavity. CNN was used for 3-class image classification.	Computer-assisted medical image classification for early oral cancer diagnosis employing deep learning algorithm.	500 images.
[30]	Autofluorescence and white light images of close-ups of oral lesions captured with a mobile phone attachment. CNN is used for binary image classification.	Automatic classification of dual-modality, smartphone-based oral dysplasia and malignancy images using deep learning	Image pairs from 170 individuals.
[31]	In-depth details of mobile phone attachment device used to capture images.	Point-of-care, smartphone-based, dual-modality, dual view, oral cancer screening device with neural network classification for low resource communities.	150 images
[32]	Histopathological images. CNN and fully convolutional network used for 7-class segmentation.	Active deep learning: Improved training efficiency of convolutional neural networks for tissue classification in oral cavity cancer.	143 images.
[33]	Fluorescence outputs used to color augment white light images of close-ups of oral lesions captured using an oral imaging camera. The fully convolutional network is used for lesion segmentation.	Automated segmentation of gingival diseases from oral images	405 images from 150 individuals.
[34]	Standard white light images of oral cavity structures. 2-class CNN-based object detection and instance segmentation of lesions.	Utilizing Mask R-CNN for detection and segmentation of oral diseases.	40 images.
[35]	Standard white light images of oral cavity structures. Semi-automated active contour used for lesion segmentation. This is followed by texture-based features with a neural network used for 6-class image classification.	Texture analysis-based segmentation and classifier of oral cancer lesion in color images using NN.	16 images.
[36]	Standard white light images of oral cavity structures. CNN is used for binary image classification.	Deep learning for classifying orofacial diseases.	75 images.
[37]	Histopathological images. Texture-based features with fuzzy classifier used for 3 class image classification	Automated oral cancer identification using histopathological images: a hybrid feature extraction paradigm	158 images from 42 individuals
[38]	Confocal laser endomicroscopy providing in vivo cell structure images. CNN is used for binary image classification.	Automatic classification of cancerous tissue in endomicroscopy laser images of the oral cavity using deep learning	7894 images from 12 individuals.

feature extraction, the deep belief network, which is used in oral cancer classification; the meta-heuristic optimization algorithm, which is used to optimize the convolutional neural network and the deep belief network for boosting the oral classification accuracy. The details of these methods are described in the following sections.

A. BASELINE MODELS

The baseline models used in comparisons with the proposed approach are briefly described in this section. The models are support vector machines (SVM), Decision Trees (DT), K-Nearest Neighbors (KNN), and Linear Discriminant (LD).

1) SUPPORT VECTOR MACHINES (SVM)

Support vector machines (SVM) can be employed as one of several supervised learning models to classify images into two categories. This method can classify new, unseen images into one of two categories after being trained on labeled data sets characterizing items belonging to a specific category. This method can be used for multi-class classifications by taking the multiple-class problem and breaking it down into two-class problems [44].

2) DECISION TREES (DT)

A decision tree is a non-parametric supervised learning method that may be used for either classification or regression. It is organized like a tree with a trunk, branches, nodes in between, and leaves. For optimal split locations inside a tree, decision tree learning uses a divide-and-conquer technique by employing a greedy search. This partitioning is performed recursively from the top down until all or the vast majority of entries fit neatly into predetermined categories [45].

3) K-NEAREST NEIGHBORS (K-NN)

K-nearest neighbors (K-NN) is a supervised learning approach for regression and classification. K-NN calculates the distance between the test data and each training point to choose the best class for the test data. Next, select K samples that are statistically most comparable to the validation set. Each of the 'K' classes in the training data is evaluated for the possibility of containing the test data, and the most likely class is selected via the K-NN method. If 'K' training points are used, those points' value is an average [46].

4) LINEAR DISCRIMINANT (LD)

One popular method of reducing the number of dimensions in supervised classification tasks is linear discriminant analysis.

It transforms features from high-dimensional space to low-dimensional space. Modeling these distinctions across groups allows for defining the boundaries between the many classes in classification tasks [47].

B. CONVOLUTIONAL NEURAL NETWORK (CNN)

The outcomes of recent deep-learning experiments in several fields have been promising. With a large data set, deep learning models can identify and categorize several layers of information. They offer a considerable advantage compared to traditional machine learning, which takes a lot of effort and expert knowledge to fine-tune the features. Multiple architectures are proposed for use in deep learning. Convolutional Neural Networks (CNN) [48] are typically utilized in image processing due to their ability to detect patterns within images. Convolutional layers, max-pooling layers, and fully linked layers are only a few components that make up a typical CNN model. It is the convolutional layer that serves as the foundation of a CNN's design [49]. The input image's characteristics, such as its borders and colors, may be extracted with its help. Dimensionality reduction of the recovered features [50] is achieved with the help of the max-pooling layer, facilitating easier processing and less computational overhead. Linearity in the network is the goal of the final stage of the CNN design, the fully linked layer.

C. DEEP BELIEF NETWORK (DBN)

The deep belief network (DBN) model correctly assigns classes to images throughout the classification phase. The structure of DBN is composed of layers of Restricted Boltzmann Machine (RBM) and Back Propagation (BP) neural networks. Layers of output, n hidden layers, and the visible layer are all included [51]. When learning, the features go through several hidden layers before reaching the input/visible layer at the very end of the model. Eventually, the correct class label will be assigned to the output layer. RBM also has input and hidden layers with bidirectional linkages between them. The number of hidden units is denoted by n and m denotes the input units, where $h = h_1, h_2, \ldots, h_n$, and $v = v_1, v_2, \ldots, v_i, v_m$. The following equation describes the RBM energy function.

$$E(v, h; \theta) = -\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij} v_i h_j - \sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j \quad (1)$$

where x is RBM parameter, which include the input layer bias a_i, the hidden layer bias b_i, and the link weight w_{ij} between the nodes in the two layers. The RBM model's energy function provides the basis for the following definition of joint distribution:

$$p(v, h) = \frac{1}{R(\theta)} e^{-E(v,h)} \quad (2)$$

$$R(\theta) = \sum_{v,h} e^{-E(v,h)} \quad (3)$$

in which the factor $R(\theta)$ is used as a normalizer. The input layer's independent probability distribution is expressed as follows:

$$p(v) = \sum_{n} p(v, h) = \frac{1}{R(\theta)} \sum_{h} e^{-E(v,h)} \quad (4)$$

Given that the nodes in each equivalent layer are not connected, the conditional probability distribution of all layers is formulated as follows:

$$p(h_j = 1|v; \theta) = \sigma\left(\sum_{i=1}^{m} w_{ij} v_j + b_j\right) \quad (5)$$

$$p(v_i = 1|h; \theta) = \sigma\left(\sum_{j=1}^{n} w_{ij} h_j + a_i\right) \quad (6)$$

The sigmoid function is denoted by the equation $\sigma(x) = 1/(1 + exp(x))$. Using adjustments to the bias a_i, b_j, and weight w_{ij}, RBM seeks to maximize the probability $p(v)$. Using a maximum likelihood estimation strategy, the RBM parameters set $\theta = \{a_i, b_i, w_{ii}\}$ can be obtained from the training data. The contrast divergence model can recognize the parameter set θ.

$$w_{ij}^{(t+\Delta t)} = w_{ij}^{(t)} + \frac{\alpha}{\beta}\left(< v_i h_j >_{data} - < v_i h_j >_{model}\right) \quad (7)$$

$$a_i^{(t+\Delta t)} = a_j^{(t)} + \frac{\alpha}{\beta}\left(< v_j >_{data} - < v_j >_{model}\right) \quad (8)$$

$$b_j^{(t+\Delta t)} = b_j^{(t)} + \frac{\alpha}{\beta}\left(< h_j >_{data} - < h_j >_{model}\right) \quad (9)$$

As soon as the RBM's initial training is complete, the current concealed layer will be exposed as the next RBM's visible layer. When an RBM training session concludes, the resulting deep features are categorized. The typical architecture of a deep belief network is shown in Figure 1 where α and β are the learning rate and the batch size, respectively.

D. META-HEURISTIC OPTIMIZATION

This paper uses two optimization algorithms: Particle Swarm Optimization (PSO) [52] and Al-Biruni earth radius (BER) [53]. These algorithms are used to develop a new hybrid optimization algorithm for optimizing the design parameters of the convolutional neural network (CNN) used in feature extraction and for optimizing the design parameters of the deep belief network (DBN) used in oral cancer classification. This section presents these algorithms' basics and then explains the proposed hybrid optimization algorithm.

1) PARTICLE SWARM OPTIMIZATION

The Particle Swarm Optimization (PSO) methodology is based on mimicking the foraging behavior of flock animals like birds using an optimization tool that outperforms the group intelligence approach. Some technical optimization issues can be solved with the PSO method [54], [55], [56], [57], which was developed after observing this class of animal foraging patterns. All particles in the PSO algorithm have an adjustable value, and their search velocity and range are controlled by the speed at which they move. The particles look for the best solutions in a D-dimensional space, where

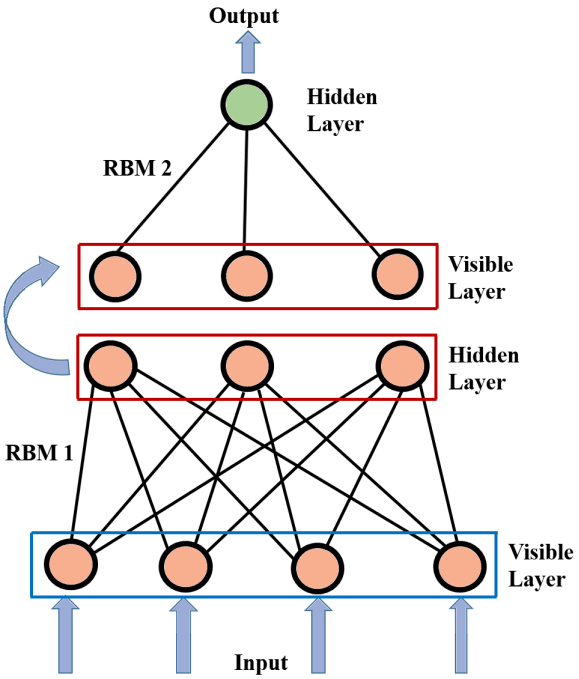


FIGURE 1. The architecture of a typical deep belief network (DBN).

a group of N particles lives together. Particles then search their solution space from the optimal particle's population position. Particles keep themselves current during the search process by monitoring two extreme values: the individual extremum P_{best}, which is the optimal solution discovered by the individual particles, and the global extremum g_{best}, which is the ideal solution currently found by the whole group. The *ith* particle's velocity throughout its exploration of the search space is represented by the following D-dimensional vector:

$$V_i = (V_{i1}, V_{i2}, \ldots, V_{iD}), i = 1, 2, \ldots, N \tag{10}$$

Each particle's particular extremum, defined as its best possible position in the solution space, is represented by the notation:

$$P_{best} = P_{i1}, P_{i2}, \ldots, P_{iD}, i = 1, 2, \ldots, N \tag{11}$$

As long as the particle locates both the local and global extrema, it can modify its current velocity and position using the following formulas.

$$V_i(t+1) = wV_i(t) + c_1 r_1(P_b - P_i(t)) + c_2 r_2(P^* - P_i(t)) \tag{12}$$

$$P_i(t+1) = P_i(t) + V_i(t+1) \tag{13}$$

where P^* denotes the optimal global position, w refers to the inertia weight, and the learning factors c_1 and c_2 are selected arbitrarily in the range between 0 and 2. V_{im} refers to the particle velocity, and r_1, r_2 are random numbers between 0 and 1. This algorithm is used to improve the search space exploration in the proposed hybrid optimization algorithm.

2) AL-BIRUNI EARTH RADIUS (BER) OPTIMIZATION

Finding the optimal solution within specified constraints is the work of optimization algorithms. Each member of the population can be represented by a S vector in BER,, $S = S_1, S_2, \ldots, S_d \in R$, where S_d is the dimension of the search space and d is the dimension of the parameter or feature being optimized. It is proposed that success up to some limit be measured using the fitness function F. Using these stages of the optimization procedure, populations may be searched for a fitness-maximizing vector S^*. To begin, a population sample is chosen randomly (solutions). Before BER can start optimizing, several factors must be specified, including the fitness function, the minimum, and maximum allowed solution sizes, the population size, the dimension, and the number of solutions [53].

- **Exploration Operation**: This operation is responsible for locating promising areas of the search space and breaking through local optimum stasis on the way to the best possible solution.
  - **Moving towards the best solution**: Using this strategy, the solitary explorer will scout out the immediate region around its current location for potentially fruitful new exploration sites. One way to accomplish this is by using an iterative process to search for a more optimal option (concerning fitness) from the numerous available choices nearby. The BER analysis uses the following formulas to accomplish this goal:

$$r = h\frac{cos(x)}{1 - cos(x)} \tag{14}$$

$$D = r_1(S(t) - 1) \tag{15}$$

$$S(t+1) = S(t) + D(2r_2 - 1) \tag{16}$$

    where $0 < x \leq 180$, h is a number that is randomly selected from the range [0, 2], r_1 and r_2 are coefficient vectors whose values are measured by equation (2), $S(t)$ is the solution vector at iteration t, and D is the diameter of the circle in which the search agent will look for promising areas.
- **Exploitation Operation**: The exploiting group is accountable for improving current solutions. When a cycle ends, the BER calculates each participant's fitness and awards those with the best scores. The BER employs two unique strategies to achieve the exploitation aim outlined in the following.
  - **Moving towards the best solution**: The following equation is employed to move in the direction of the best solution.

$$S(t+1) = r^2(S(t) + D) \tag{17}$$

$$D = r_3(L(t) - S(t)) \tag{18}$$

    where r_3 is a random vector calculated using equation (2) that controls the movement steps towards

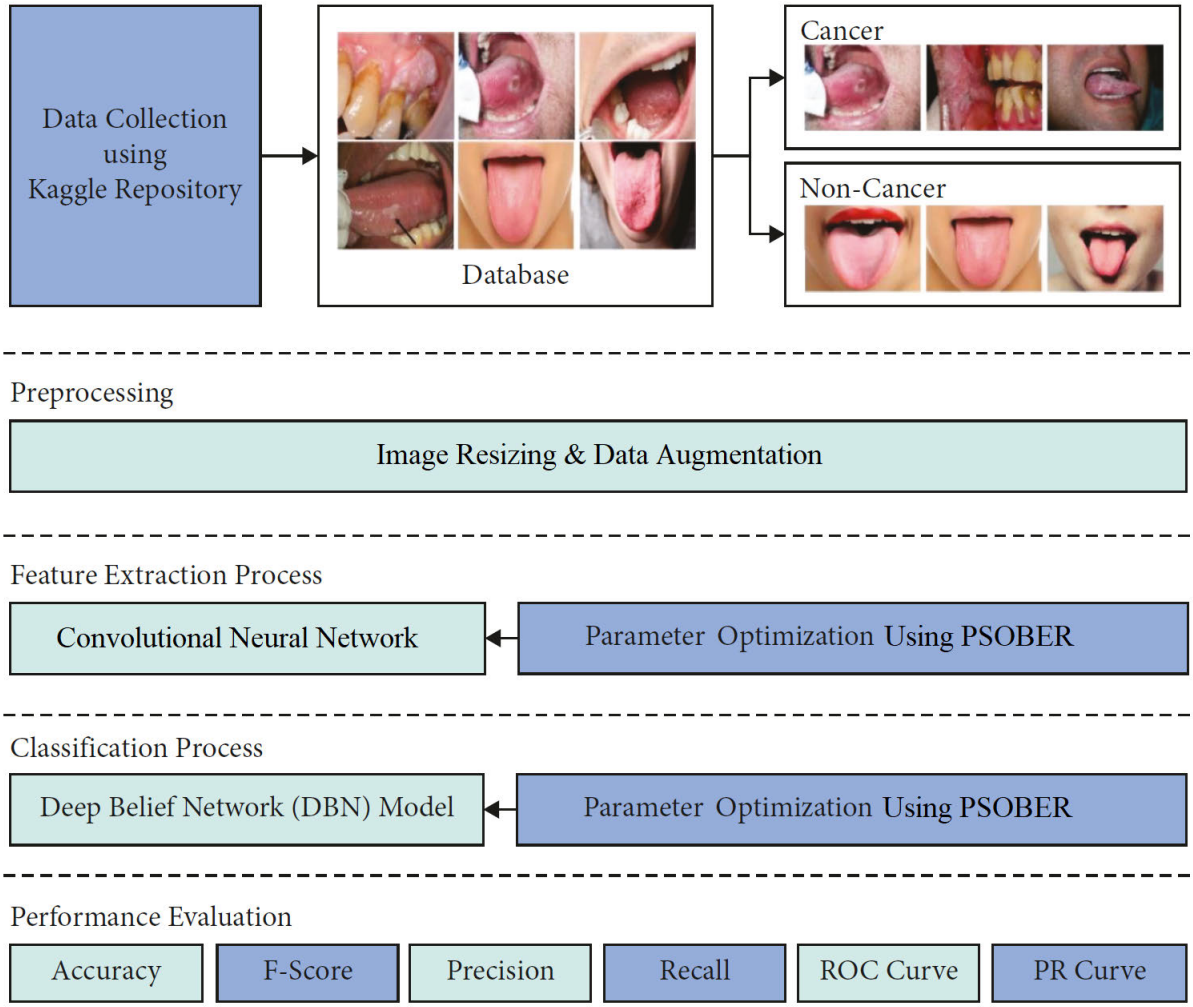


FIGURE 2. The architecture of the proposed methodology.

the best solution, $S(t)$ is the solution vector at iteration t, $L(t)$ is the best solution vector, and D refers to the distance vector.

– **Searching the area around the best solution** Generally speaking, the region around the optimal response offers the most potential for success (leader). As a result, some people will search for ways to enhance the situation by investigating alternatives that are somewhat similar to the best. To implement the process mentioned above, the BER employs the following equation.

$$S'(t+1) = r(S^*(t) + k) \tag{19}$$

$$k = 1 + \frac{2 \times t^2}{Max_{iter}^2} \tag{20}$$

where the best solution is denoted by $S^*(t)$, which is selected after Comparing $S(t+1)$ and $S'(t+1)$. The following equation mutates the solution if the best fitness is not changed for the last two iterations.

$$S(t+1) = k * z^2 - h\frac{cos(x)}{1 - cos(x)} \tag{21}$$

where z is a random number in the range [0, 1] and t is the iteration number.

- Selection of the best solution: To ensure that the solutions are of excellent quality, the BER chooses the best to employ in the next cycle. While the elitism technique is more effective, it may lead to quick convergence [58], [59]. The BER can deliver cutting-edge exploration capabilities by taking a mutational approach and scanning the explorers' surrounding area. With its robust exploration capabilities, the BER can stave off convergence. First, parameters such as population size, mutation frequency, and iteration count are input into the BER. The BER then assigns individuals to either the exploration or exploitation groups. The BER method dynamically modifies the size of each group throughout the iterative process of locating the best solution. Each

group uses two approaches to complete their missions. The BER ensures diversity and thorough exploration by shuffling the order of responses between repetitions. In one iteration, a solution may be part of the exploration group, but by the next, it may be part of the exploitation group. Due to the BER's exclusive nature, the leader will not be deposed during the procedure.

III. THE PROPOSED METHODOLOGY

In this section, the methodology proposed for oral cancer classification is explained. The steps of the proposed methodology are depicted in Figure 2. These steps include preprocessing, feature extraction, classification, and performance evaluation. In the preprocessing step, image resizing and data augmentation are implemented. In the feature extraction process, a set of features are extracted using a convolutional neural network (CNN). Whereas the classification of the input image is performed in terms of a deep belief network (DBN). The key of the proposed methodology is the new optimization algorithm developed to optimize the CNN and DBN design parameters. This optimization algorithm is a hybrid of two optimization algorithms: particle swarm optimization and Al-Biruni earth radius optimization algorithms. The idea behind using these two algorithms is to exploit both advantages to improve the proposed optimization algorithm's exploration and exploitation process. More details about the steps of the proposed methodology are explained in the next sections.

A. DATASET PREPROCESSING

Preprocessing is applied to all input images of the dataset to provide increased consistency in the extracted features and the classification results. All images in the given dataset are available in a dimension of 1100×825. Using image resizing, the dimensions of the images are resized to 256×256. This step targets preparing the input images to fit the input layer of the CNN used in feature extraction. On the other hand, due to the limited number of images in the given dataset, data augmentation is employed to increase the number of images in the dataset. Various transformation operations were used to generate a new image to enrich the dataset. The applied transformation operations are listed in Table 2.

B. FEATURE EXTRACTION

To extract the set of significant features from the input image, the convolutional neural network (CNN) model is employed [60]. In the CNN concept, the block is the basis for construction, with several blocks working together to build a single cellular structure. The CNN uses a factorization of networks into cells and then separates those cells into blocks as the searching space. There is no set standard for the size or composition of the individual cells or blocks. They must, however, be tailored to the specifics of the data being used. The block likely uses operations like convolution, separable convolution, maximum pooling, average pooling, and an identification map. In other words, the block can convert a pair of inputs into a feature map. It is an adder that works element by element. If a block of size $H \times W$ is sent to a cell with a stride of 1, the resulting feature map will also be $H \times W$ in size. When the stride length reaches 2, the increment is halved. The structure of CNN used in this work is selected based on the proposed optimization algorithm that is discussed in the next section. The typical structure of CNN used in feature extraction is shown in Figure 3.

TABLE 2. The settings of the transformation operations used in data augmentation.

Transformation	Setting
Zoom transformation	$[0.2 - 0.6]$
Rotation transformation	$[25° - 75°]$
Scale transformation	[0 - 1]
Shear transformation	$[25° - 75°]$
Horizontal flip	True

C. THE PROPOSED PSOBER OPTIMIZATION ALGORITHM

To improve the classification performance, a new optimization algorithm is proposed based on a hybrid of the particle swarm optimization (PSO) [52] and Al-Biruni earth radius (BER) [53] optimization algorithms. The proposed algorithm is referred to as PSOBER. This optimization algorithm is used to optimize the parameters of the convolutional neural network used in feature extraction and the DBN used in classifying the input image. The steps of the proposed PSOBER algorithm are listed in Algorithm 1. The main advantage of the BER over the PSO algorithm is the fast convergence as reported in [53], which adds merit when combined with the PSO algorithm. The better exploitation of the search space distinguishes the standard PSO algorithm. However, the exploration process is time-consuming. On the other hand, the BER algorithm is distinguished by the better exploration of the search space. Therefore, the proposed PSOBER optimization algorithm exploits the advantages of the PSO and BER to achieve better exploration and exploitation of the search space. During the optimization process for the tested problem, the proposed PSOBER algorithm balances the exploration and exploitation of the individuals in the search space.

In the proposed method, 70% of the population is divided into two groups: exploration and exploitation. Having a large number of individuals involved in the exploration group helps in finding new search regions. Individual fitness increases when more individuals can increase their fitness levels, but the percentage of individuals in the exploratory group decreases rapidly from 70% to 30%. If a better solution cannot be identified, then an elitism approach guarantees convergence by keeping the process leader in subsequent populations. PSOBER may raise the size of the exploration

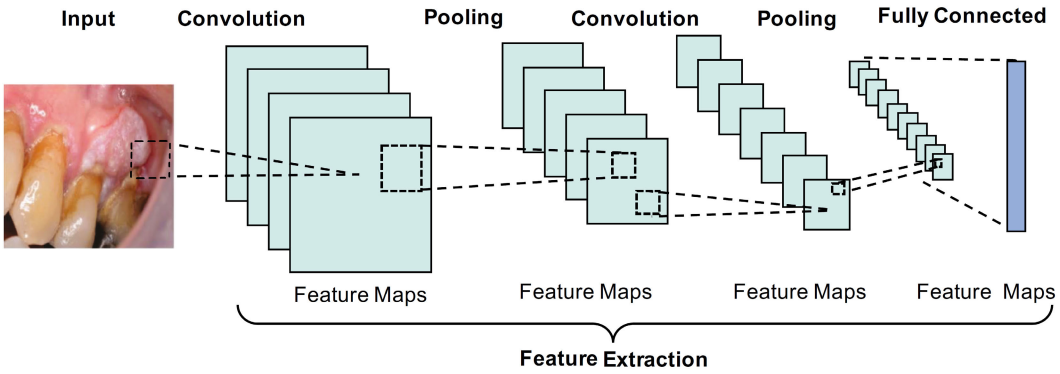


FIGURE 3. Architecture of Convolutional Neural Network (CNN) used in feature extraction.

group at any point so long as the leader's fitness has not increased sufficiently throughout three iterations.

D. FITNESS FUNCTION

The fitness function is used to measure the quality of the solutions resulting from the optimization algorithm. The formulation of the fitness function is represented by the following equation.

$$Fitness = v_1 Error + v_2 \frac{|S|}{|T|} \tag{22}$$

where $Error$ represents the classification error, $|S|$ denotes the number of selected features features, and $|T|$ refers to the number of features. The factors v_1 and v_2 are in the range of $[0, 1]$, where $v_1 = 1 - v_2$.

E. CNN PARAMETER OPTIMIZATION USING PSOBER

Optimizing the convolutional neural network (CNN) hyperparameters starts with initializing the population of the proposed PSOBER algorithm with a defined number of dimensions (number of hyperparameters to optimize). Therefore, images need to be normalized before being fed to CNN. During the optimization process, the loss function (cross-entropy) of the convolutional neural network is used as the objective function. Therefore, the proposed PSOBER algorithm's iterations begin after the objective function has been evaluated. The locations of individuals are then updated in a way that is determined by the proposed PSOBER algorithm. The best solution, defined as the hyperparameters that yielded the smallest loss, is evaluated alongside the new positions. Finally, it is checked if the stop condition (maximum number of iterations) is reached; otherwise, the process is rerun until a proposed solution is found.

F. DBN PARAMETER OPTIMIZATION USING PSOBER

To maximize the usefulness of any model, including deep neural networks, parameter tuning is an essential step. The PSOBER parameter optimization approach, described in the previous section, is used in the proposed method to determine the best settings for the deep belief network. PSOBER-DBN receives the dataset to train the model. PSOBER-DBN receives the images from the dataset and uses them to construct the model in two stages: pretraining, during which DBN parameters are optimized with PSOBER, and training, during which the DBN itself is constructed.

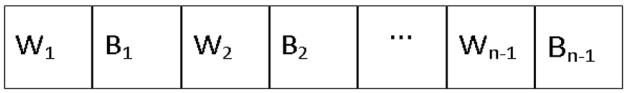


FIGURE 4. The coding scheme of population individuals.

1) PRETRAINING PHASE

The dataset is fed to the proposed PSOBER parameter optimization algorithm at the pretraining phase, at which point the optimal parameters for the DBN being constructed are determined. During this stage of PSOBER parameter tuning, a random particle population is established. A DBN's connection weight and bias parameters, represented by a particle's location in the population, are defined by each individual particle. There are several generations that PSOBER operates at. The position of each particle in the population is updated when a new generation's worth of data has been collected, which includes an assessment of each particle's current velocity. After each generation, the DBNs' training dataset is applied to the costs of each DBN in the population, and the optimal individual and population costs are updated.

Connection weights in each RBM and neuron biases are initialized randomly to provide a diverse population. Each member of the population stands for one of the DBN model's parameters. Figure 4 represents the coding scheme of each each individual in the population. Each individual is of length $\sum_{i=1}^{L-1} N_{i+1}(N_i + 1)$, where (L) is the number of DBN layers and $(L - 1)$ is the number of RBMs in the DBN. For the ith RBM, the number of neurons in the input layer is denoted by N_i and the number of neurons in the output layer is denoted by N_{i+1}. The biases and connection weights of the *ith* RBM's output neurons are represented by the vectors B_i and W_i in this encoding scheme.

Algorithm 1 : The Proposed PSOBER Optimization Algorithm

Initialize BER population $P_i(i = 1, 2, \ldots, d)$ with size d, iterations Max_{iter}, fitness function F_n, $t = 1, n_1, n_2, a, r_1, r_2, r_3, r_4, r_5, r_6, r_7$
Calculate fitness function F_n for each P_i
Find best solution as P^*
while $t \leq Max_{iter}$ **do**
  if t%2 == 0 **then**
    for $(i = 1 : i < n_1 + 1)$ **do**
      Update $r = h\frac{cos(x)}{1-cos(x)}$
      Calculate $D = r_1(P(t) - 1)$
      Update positions to head toward best solution as $P(t + 1) = P(t) + D(2r_2 - 1)$
    end for
    for $(i = 1 : i < n_2 + 1)$ **do**
      Calculate $D = r_3(L(t) - P(t))$
      Update positions of best solution as $P(t + 1) = r_2(P(t) + D)$
      Calculate $k = 1 + \frac{2\times t^2}{Max_{iter}^2}$
      Investigate area around best solution as $P'(t + 1) = r_1(P^*(t) + k)$
      Compare $P(t + 1)$ and $P'(t + 1)$ to select best solution P^*
      if best fitness is not changed for last two iterations **then**
        Mutate solution as $P(t+1) = k*z^2 - h\frac{cos(x)}{1-cos(x)}$
      end if
    end for
    Update the fitness function F_n for each $P(t)$ using BER
  else
    Update Calculate fitness function Fn for each $P(t)$ using PSO
  end if
  Update BER and PSO parameters, $t = t + 1$
end while
Return P^*

DBN's connection weights and biases are evolved with the help of the proposed PSOBER. Every new generation improves upon the last by incorporating the current velocities and locations of the individuals into the population as a whole. In Algorithm 2, every generation is run on the training data to determine which of the potential models the population represents is the best to prepare for the development of the following generation. Each individual in the population is used to construct a DBN according to the encoding method, and then the cost of each DBN is calculated using the available training data. When the current generation's costs have been evaluated, the evolution process moves on to the next generation. At this point, each individual's best position and the whole population are updated.

Algorithm 2 : Cost Evaluation of Individuals in PSOBER

Input Training dataset, population individuals
for each individual i in population **do**
  Transform individual i into a DBN
  Construct the DBN D_i using the individual i
  Train the DBN using the dataset
  Evaluate Cost (Mean square error) using: $Cost(D_i) = \sum(y - o)^2/M$, where M is the number of training samples, y is the predicted output, and o is the actual output
end for

2) DBN TRAINING

In this stage, the parameter values determined in the previous stage are fine-tuned using the backpropagation approach (see Algorithm 3). In DBN, the input layer receives the training data and sends it on to the output layer. The error is determined at the output layer and sent via the layers. Here are the relevant equations: Derivative of the loss function (mean square error) is used to determine the error at the output layer (L), which is provided by the following equation.

$$\delta_j^L = y_j^L(1 - y_j^L)(y_j^L - o_j) \tag{23}$$

On the other hand, the error propagated to the hidden layers $l = L - 1, L - 2, \ldots 2$ is calculated using the following equation.

$$\delta_j^l = y_j^l(1 - y_j^l)\sum_{q=1}^{m} w_{jq}^{l+1}\delta_q^{l+1} \tag{24}$$

Based on the following negative gradients, the biases and network weights for layers $l = L, L - 1, L - 2, \ldots 2$ are updated:

$$\Delta b_j^l = -\gamma\delta_j^{l+1} \tag{25}$$

$$\Delta w_{ij}^l = -\gamma y_i^l\delta_j^{l+1} \tag{26}$$

where the lth RBM output layer is denoted by y^l. For the layer $l + 1$, the neuron connection weight from ith neuron to jth neuron is denoted by w_{ij}^{l+1}, and the learning rate is denoted by γ. The full process of optimizing the design parameters of the DBN is depicted in Figure 5.

Algorithm 3 : DBN Construction and Fine-Tuning Based on the Best Individual

Input Training dataset, best population individual
Build the DBN using the parameters of the best individual
for $i = 1$ to $Iter_{Max}$ **do**
  Feed the DBN with the training dataset
  Calculate the prediction error
  Backpropagate the error from the output layer to the previous layers
  Update the weights and biases
end for

TABLE 3. Key performance indicators.

No.	Metric	Formula
1	Accuracy	$= \frac{TP+TN}{TP+TN+FP+FN}$
2	Sensitivity	$= \frac{TP}{TP+FN}$
3	Specificity	$= \frac{TN}{TN+FP}$
4	Precision	$= \frac{TP}{TP+FP}$
5	NValue	$= \frac{TN}{TN+FN}$
6	F1-score	$= 2 \times \frac{Precision \times Sensitivity}{Precision+Sensitivity}$

G. PERFORMANCE METRICS

Several necessary measures, such as accuracy, sensitivity, specificity, precision, negative predictive value (NValue), and F1-score, are used to evaluate the effectiveness of the proposed methodology. TP denoted True Positive value in these measures, representing the proportion of samples correctly labeled as expected. Another quality control measure is the rate of false positives, which FP denotes. In addition, the number of instances that were wrongly labeled as non-defective is indicated by FN, or false negative rate. The percentage of correctly diagnosed cases as non-defective is denoted by True Negative (TN). Table 3 summarizes the metrics employed in the conducted experiments along with the corresponding formulas.

IV. EXPERIMENTAL RESULTS

A set of experiments was conducted to evaluate the proposed approach's effectiveness. These experiments are run on a machine with Nvidia GPU (8 GB), Intel Core i7 CPU, and 16 GB of main memory. This section explains the achieved results and provides a discussion on the recorded comparisons.

A. DATASET

In the conducted experiments, the publicly available dataset, on Kaggle repository [61], is employed to examine the proposed model's efficacy in classifying oral cancer. The images of the mouth and tongue are included in the collection and labeled as cancerous or non-cancerous. The dataset consists of 87 cancerous images and 44 non-cancerous images, totaling 131 images. Figure 6 depicts sample images from the dataset. To enrich this dataset, data augmentation is employed [62], [63]. The number of images in the dataset after augmentation is 1310 images. This dataset of images is split into two sets; the training set (90% with 10% of this set used for validation) and 10% for the testing set.

B. CONFIGURATION PARAMETERS

The configuration parameters of the employed optimization algorithms are listed in Table 4, and the configuration parameters of the baseline models are presented in Table 5. In addition, the run-time configuration parameters are set as learning rate (0.001), epochs (40), and batch_size (32).

TABLE 4. Configuration parameters of the employed optimization algorithms.

Algorithm	Parameter	Values
PSO [52]	Acceleration constants	[2,2]
	Inertia W_{max}, W_{min}	[0.6, 0.9]
	Particles	10
	Iterations	80
BER [53]	Iterations	80
	Mutation probability	0.5
	Exploration percentage	70
	k (decreases from 2 to 0)	1
GWO [64]	a	2 to 0
	Iterations	80
	Wolves	10
WOA [65]	r	[0, 1]
	Iterations	80
	Whales	10
	a	2 to 0
GA [66]	Cross over	0.9
	Mutation ratio	0.1
	Mechanism	Roulette wheel
	Agents	10
	Iterations	80
DE [67]	Similar to GA	

TABLE 5. Configuration parameters of the baseline classification models.

Model	Parameter	Value
SVM [44]	C	1
	kernel	'rbf'
	penalty	'l2'
	tol	1.0e-4
K-NN [46]	n_neighbors	5
	weights	'uniform'
	leaf_size	30
	p	2
LD [47]	solver	'svd'
	tol	1.0e-4
	shrinkage	[0-1]
DT [45]	splitter	'best'
	min_samples_split	2
	criterion	'gini' to 0
	min_samples_leaf	1

C. NETWORK STRUCTURE HYPERPARAMETERS

The structure of the CNN and DBN networks is optimized using the proposed optimization algorithm. Table 6 shows the structure hyperparameters of both networks along with the optimized parameters. These structures are used in feature extraction and image classification. The structure of the optimized networks is similar to the structure of the standard networks. However, the optimized structure best suits the task at hand as it is determined based on the best performance these structures can achieve.

D. ACHIEVED RESULTS

This section compares the proposed optimized DBN model to state-of-the-art machine learning models, including DBN,

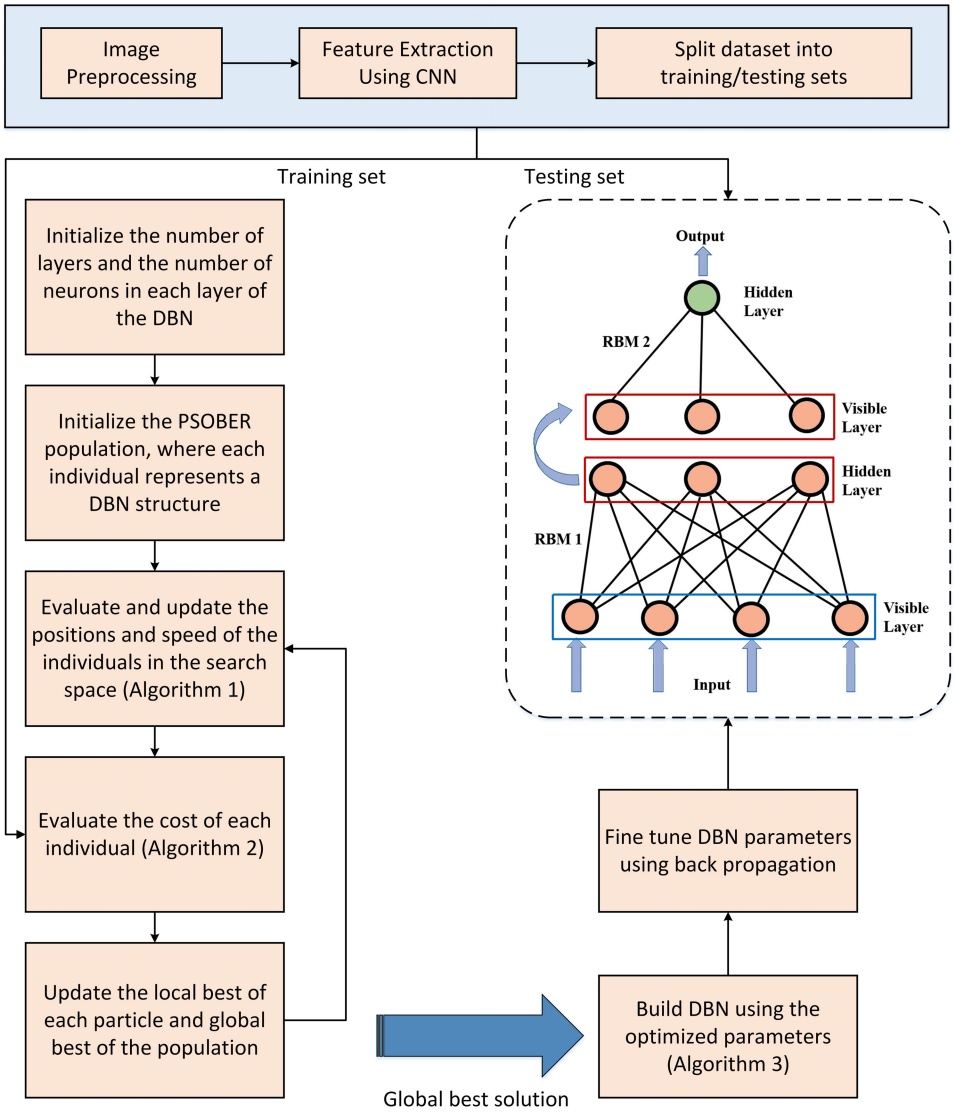


FIGURE 5. The steps of building DBN model using the proposed PSOBER methodology.

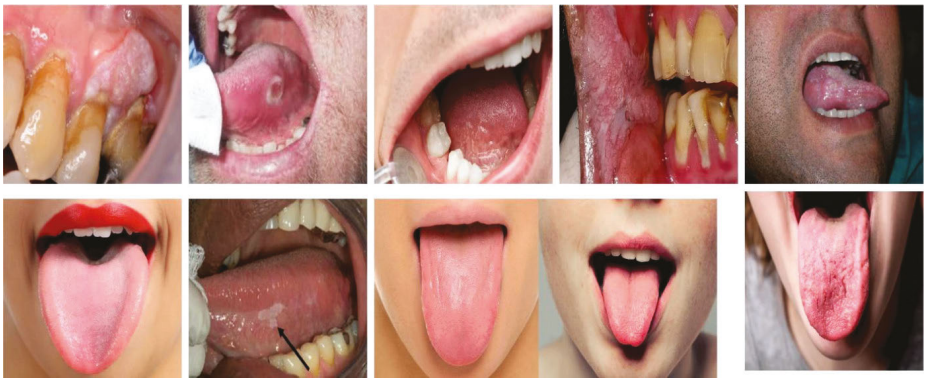


FIGURE 6. Sample images from the Kaggle dataset.

SVM-Linear, SVM-Gaussian, SVM-Cubic, K-NN, LD, and DT. Table 7 presents the measured values of the evaluation metrics based on the results achieved by each model. For these adopted metrics, the model that achieves the largest values is consider better than others. This table shows that the best accuracy (97.3%) is achieved by the proposed optimized DBN (PSOBER-DBN). Similarly, the other metrics prove the superiority of the proposed approach when compared with

TABLE 6. Network hyperparameters with the corresponding optimized values.

Network	Hyperparameter	Range	Optimized
CNN	Number of Filters 1	[32, 64, 96]	64
	Kernel Size 1	[3, 4, 5]	5
	Number of Filters 2	[64, 96, 128]	128
	Kernel Size	[3, 4, 5]	5
	Number of Filters 3	[64, 96, 128]	128
	Kernel Size 3	[3, 4, 5]	4
	Hidden Layer 1	[60, 100, 125,256]	125
	Hidden Layer 2	[60, 100, 125,256]	125
DBN	Number of RBMs	[3, 4, 5, 6, 7]	6

TABLE 7. The measured values of the evaluation criteria based on the results achieved by the proposed PSOBER-DBN and other competing methods.

	Accuracy	Sensitivity	Specificity	Precision	Nvalue	F1-Score
PSOBER-DBN	0.973	0.943	0.986	0.963	0.977	0.953
DBN	0.935	0.943	0.917	0.963	0.875	0.953
SVM-Linear	0.912	0.909	0.917	0.940	0.875	0.924
SVM-Gaussian	0.909	0.903	0.917	0.936	0.875	0.920
SVM-Cubic	0.904	0.893	0.917	0.929	0.875	0.911
K-NN	0.888	0.857	0.917	0.905	0.875	0.880
LD	0.874	0.813	0.917	0.872	0.875	0.841
DT	0.860	0.750	0.917	0.826	0.875	0.786

TABLE 8. ANOVA test of the results achieved by the proposed PSOBER-DBN compared to other models.

	SS	DF	MS	F (DFn, DFd)	P value
Treatment	0.08942	7	0.01277	F (7, 72) = 495.3	P<0.0001
Residual	0.001857	72	0.000026		
Total	0.09128	79			

the other seven machine learning models. In addition, the non-optimized DBN model gives high measured values of the adopted criteria. However, when it is optimized using the proposed optimization algorithm (PSOBER) its performance is improved as shown in Table 7.

In addition, the accuracy of the results achieved by the proposed optimized DBN with comparison to other standard models is shown in Figure 7. As shown in this figure, the proposed optimized model could achieve the highest accuracy which proves its effectiveness and superiority.

The statistical difference between the proposed and machine learning models is measured using the one-way analysis-of-variance (ANOVA) statistical test. Like the Wilcoxon test, the ANOVA test is based on two hypotheses, namely, H_0 and H_1. For H_0, the means of all models' accuracy are assumed equal; $\mu PSOBER-DBN = \mu DBN = \mu SVM - Linear = \mu SVM - Gaussian = \mu SVM - Cubic = \mu K - NN = \mu LD = \mu DT$. The results of this test are presented in Table 8, and the results of this test based on the objective function are shown in Figure 8.

On the other hand, the results of the comparison between the proposed optimized PSOBER-DBN and the other machine learning models are presented in Table 9. This test measures the significance of the proposed method

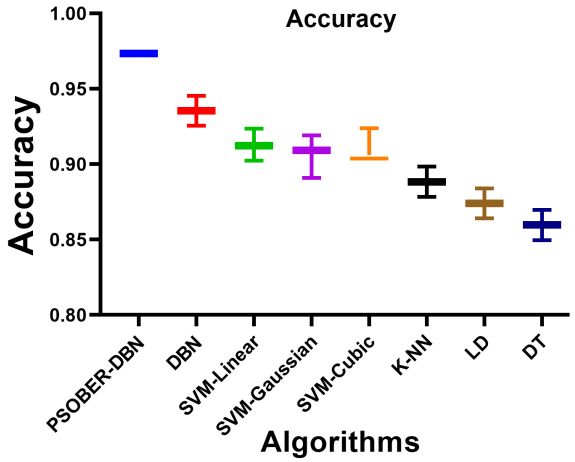


FIGURE 7. The accuracy achieved by the optimized DBN compared to the other standard models.

and the other models. If the achieved p-value of this test is smaller than 0.5, this reflects the statistical difference of the proposed method. The formulation of this test is based on two hypotheses; namely, null hypothesis (denoted by H_0), in which the means (μ) of all models accuracy are assumed as follows: $\mu PSOBER - DBN = \mu DBN$, $\mu PSOBER-DBN = \mu SVM-Linear$, $\mu PSOBER-DBN = \mu SVM - Gaussian$, $\mu PSOBER-DBN = \mu SVM - Cubic$, $\mu PSOBER-DBN = \mu K-NN$, $\mu PSOBER-DBN = \mu LD$, $\mu PSOBER - DBN = \mu DT$, and the alternate hypothesis (denoted by H_1), in which the means of models accuracy are assumed not equal. As presented in Table 9, the measured p-values between the proposed model and the other competing models are less than 0.5, which shows the proposed approach's statistical difference. In this case, the alternate hypothesis H_1 is accepted.

More detailed results are shown in Figure 8. These results include a residual plot, homoscedasticity plot, quartile-quartile (QQ) plot, and heatmap plot. The residual and homoscedasticity plots show the accuracy of the achieved results when the values approach the zero level. In addition, the QQ plot highlights the accuracy of the results when aligned close to the red dotted line. On the other hand, the heatmap is another perspective of results analysis. In this plot, the Y-axis represents the number of runs, the X-axis represents the employed model, and the cells represent the error values colored using the color map shown on the right. As the color of the cells gets black, this means minimum error values and thus represents high accuracy. As these plots show the performance of the proposed optimized model, it can be noted that the residual and homoscedasticity plots depict tiny errors in the recorded results, the QQ and heatmap show accurate results when compared to the other models, and finally the heatmap shows black cells corresponding to the proposed model. Based on these results, it is obvious that the proposed approach is superior to the other methods for classifying oral cancer cases.

TABLE 9. Wilcoxon signed rank test of the results achieved by the proposed PSOBER-DBN compared to other models.

	PSOBER-DBN	DBN	SVM-Linear	SVM-Gaussian	SVM-Cubic	K-NN	LD	DT
Number of values	10	10	10	10	10	10	10	10
Theoretical median	0	0	0	0	0	0	0	0
Actual median	0.973	0.935	0.912	0.909	0.904	0.888	0.874	0.860
Wilcoxon Signed Rank Test								
P value (two tailed)	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact
Sum of signed ranks	55	55	55	55	55	55	55	55
Sum of positive ranks	55	55	55	55	55	55	55	55
Sum of negative ranks	0	0	0	0	0	0	0	0
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Discrepancy	0.9735	0.9354	0.9123	0.9091	0.9037	0.8884	0.874	0.8596

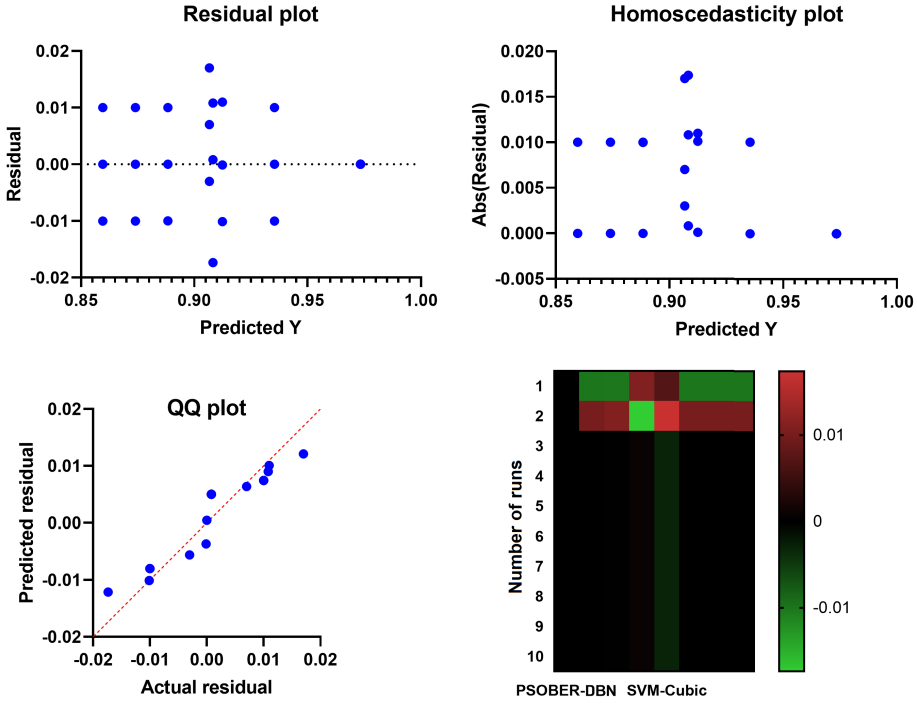


FIGURE 8. Analyzing the performance of the proposed PSOBER-DBN compared to the other machine learning models.

E. COMPARING THE PROPOSED OPTIMIZATION METHOD WITH STATE-OF-THE-ART METHODS

The superiority and significance of the proposed optimization algorithm are studied and analyzed in this section. Table 10 presents the results of the statistical measures calculated to profoundly investigate the efficiency of the proposed PSOBER-DBN in comparison with the other optimization methods. It can be noted from the results recorded in this table that the proposed approach is superior and can achieve better results in classifying oral cancer cases. In general, the values of the mean, standard deviation, median, ..., etc, calculated from the accuracy values achieved by the proposed approach are better than those of the other methods.

In addition, this study is based on statistical tests using the ANOVA and Wilcoxon signed-rank tests. The results of these tests are presented in Table 11 and Table 12. These tests are based on a similar hypothesis to those discussed in the previous section. The H0 hypothesis assumes the means of model accuracy are equal, and the H1 hypothesis assumes the means of model accuracy are different. The results shown in these tables prove the statistical difference of the proposed optimization algorithm compared to four other optimization algorithms: BER [53], PSO [52], grey wolf Optimizer (GWO) [64], and Whale Optimization Algorithm (WOA) [65].

Additional analysis of the achieved results is visualized in the plots depicted in Figure 9 based on a set of random samples from the Kaggle dataset. In this figure, the accuracy of the proposed approach is clearly shown based on the residual, homoscedasticity, QQ, and heatmap plots compared to the other meta-heuristic optimization algorithms included in the conducted experiments.

TABLE 10. Statistical results of the achieved results based on the proposed PSPBER-DBN.

	PSOBER-DBN	BER-DBN	PSO-DBN	GWO-DBN	WOA-DBN	DE-DBN	GA-DBN
Number of values	13	13	13	13	13	13	13
Upper 95% CI of geo. mean	0.9735	0.9658	0.9608	0.9528	0.9489	0.9482	0.9447
Lower 95% CI of geo. mean	0.9735	0.9609	0.9559	0.9479	0.9433	0.9432	0.9375
Harmonic mean	0.9735	0.9633	0.9584	0.9503	0.9461	0.9457	0.9411
Upper 95% CI of mean	0.9735	0.9658	0.9608	0.9528	0.9489	0.9482	0.9447
Lower 95% CI of mean	0.9735	0.9609	0.9559	0.9479	0.9433	0.9432	0.9375
Quadratic mean	0.9735	0.9634	0.9584	0.9503	0.9461	0.9457	0.9411
Upper 95% CI of quad. mean	0.9735	0.9658	0.9608	0.9528	0.9489	0.9482	0.9447
Lower 95% CI of quad. mean	0.9735	0.9609	0.9559	0.9479	0.9433	0.9432	0.9375
90% Percentile	0.9735	0.9694	0.9644	0.9563	0.9536	0.9517	0.9545
Upper confidence limit	0.9735	0.9634	0.9584	0.9503	0.9459	0.9457	0.9419
Lower confidence limit	0.9735	0.9634	0.9584	0.9503	0.9459	0.9457	0.9385
10% Percentile	0.9735	0.9574	0.9524	0.9443	0.9399	0.9397	0.9385
Geometric mean	0.9735	0.9633	0.9584	0.9503	0.9461	0.9457	0.9411
Upper 95% CI of harm. mean	0.9735	0.9658	0.9608	0.9528	0.9489	0.9482	0.9446
Lower 95% CI of harm. mean	0.9735	0.9609	0.9559	0.9479	0.9433	0.9432	0.9375
Minimum	0.9735	0.9534	0.9484	0.9403	0.9359	0.9357	0.9385
25% Percentile	0.9735	0.9634	0.9584	0.9503	0.9459	0.9457	0.9385
Median	0.9735	0.9634	0.9584	0.9503	0.9459	0.9457	0.9385
75% Percentile	0.9735	0.9634	0.9584	0.9503	0.9459	0.9457	0.9402
Maximum	0.9735	0.9734	0.9684	0.9603	0.9587	0.9557	0.9585
Range	0	0.02	0.02	0.02	0.02284	0.02	0.02
Std. Error of Mean	0	0.00113	0.00113	0.00113	0.00130	0.00113	0.00165
Actual confidence level	97.75%	97.75%	97.75%	97.75%	97.75%	97.75%	97.75%
Std. Deviation	0	0.00408	0.00408	0.00408	0.00469	0.00408	0.00595
Geometric SD factor	1	1.004	1.004	1.004	1.005	1.004	1.006
Coefficient of variation	0.000%	0.4238%	0.4260%	0.4296%	0.4960%	0.4317%	0.6328%
Mean	0.9735	0.9634	0.9584	0.9503	0.9461	0.9457	0.9411
Kurtosis		6	6	6	6.389	6	6.589
Skewness		0	0	0	0.8998	-4.821E-14	2.586
Sum	12.66	12.52	12.46	12.35	12.3	12.29	12.23

TABLE 11. The results of ANOVA test of the proposed (PSOBER-DBN) compared to the other models.

	SS	DF	MS	F (DFn, DFd)	P value
Treatment	0.006099	4	0.001525	F (4, 60) = 105.9	P<0.0001
Residual	0.0008643	60	0.0000144		
Total	0.006963	64			

TABLE 12. The results of Wilcoxon signed-rank test of the proposed (PSOBER-DBN) compared to the other models.

	PSOBER-CNN	BER-CNN	PSO-CNN	GWO-CNN	WOA-CNN	DE-CNN	GA-CNN
Number of values	13	13	13	13	13	13	13
Theoretical median	0	0	0	0	0	0	0
Actual median	0.9735	0.9634	0.9584	0.9503	0.9459	0.9457	0.9385
P value (two tailed)	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact	Exact
Sum of positive ranks	91	91	91	91	91	91	91
Sum of signed ranks (W)	91	91	91	91	91	91	91
Sum of negative ranks	0	0	0	0	0	0	0
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Discrepancy	0.9735	0.9634	0.9584	0.9503	0.9459	0.9457	0.9385

On the other hand, the accuracy of the classification results achieved by the proposed PSOBER-DBN and the other optimization algorithms is shown in Figure 10. As shown in this figure, the proposed method achieves the highest accuracy compared to the other methods. In addition, the histogram and the Receiver Operator Characteristic (ROC) are shown in Figure 11 and Figure 12, respectively. The ROC curve is a convenient plot for analyzing two classification models as it displays the trade-off between the sensitivity (proportion of positive tuples that are recognized) and the specificity (proportion of negative tuples that are incorrectly recognized as positive) for the proposed (PSOBER-DBN) model compared to another model (BER-DBN) as a proof of concept. In this plot, the closer the ROC curve of a model is to the diagonal

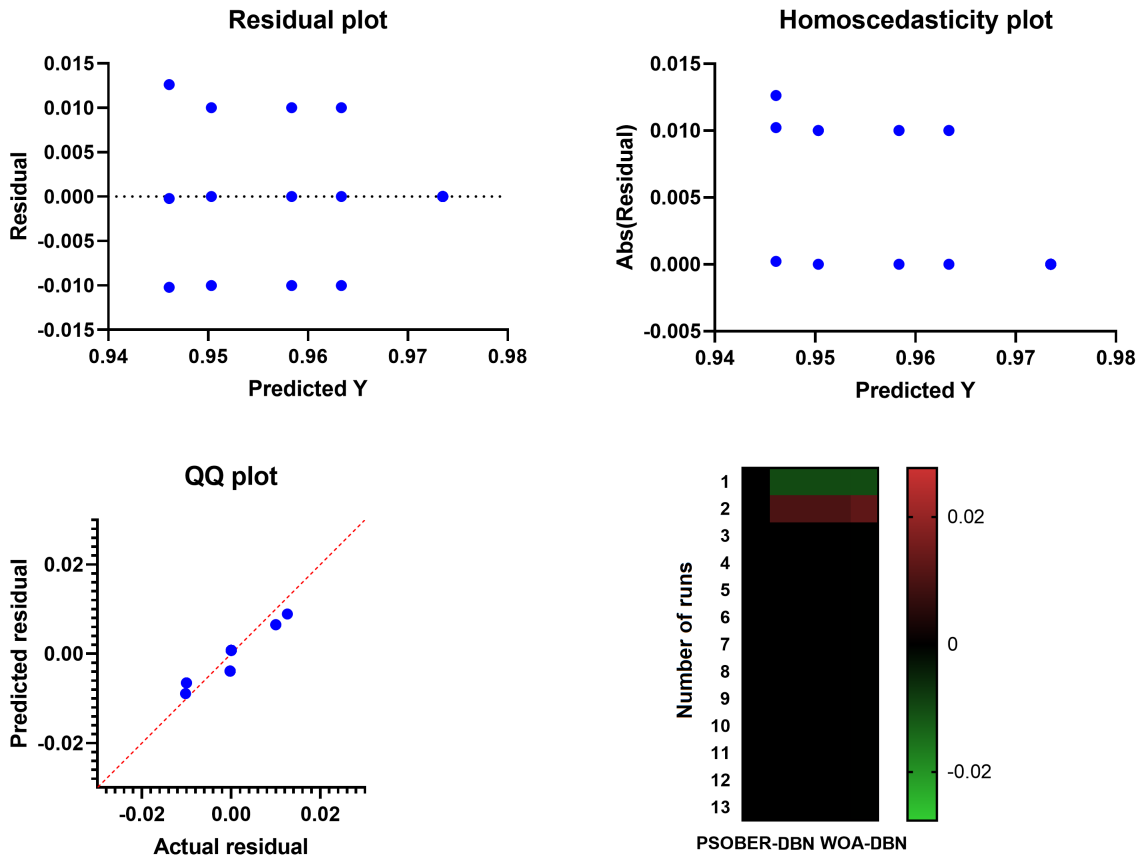


FIGURE 9. Analyzing the performance of the proposed PSOBER-DBN compared to the other optimization algorithms.

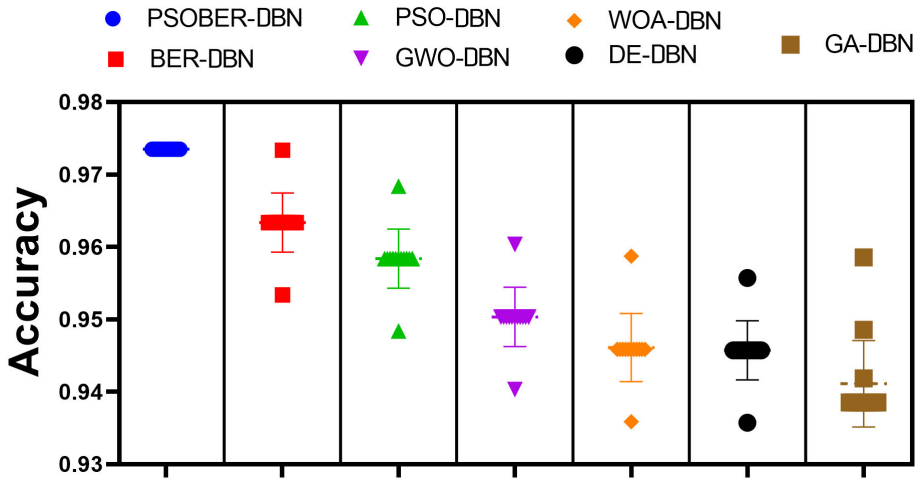


FIGURE 10. Accuracy plot of the achieved results using the proposed PSOBER-DBN and the other optimization algorithms.

line, the less efficient the model and vice versa. In Figure 12, it is obvious that the ROC curve (blue line) is far from the diagonal line (red dotted line), which proves the efficiency of the proposed model. Overall, the results represented by these figures show a clear superiority of the proposed approach in solving the problem of classifying oral cancer in images.

F. COMPARISON WITH DEEP LEARNING MODELS

To emphasize the superiority of the proposed approach, an additional experiment is conducted to compare its performance to the other deep learning models. In this experiment, the VGGNet, ResNet-50, and AlexNet deep models are considered in the conducted experiments. Table 13 presents the

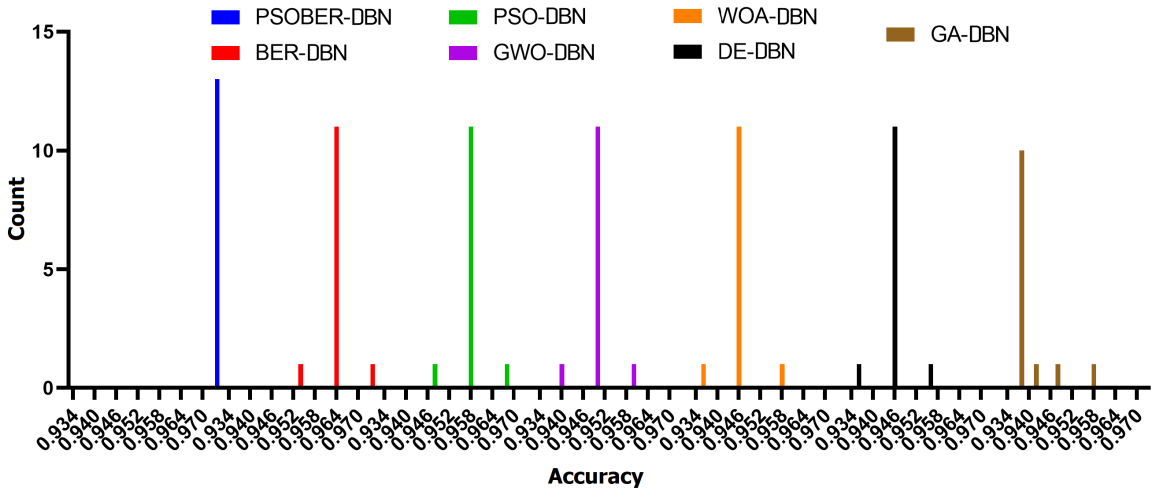


FIGURE 11. Histogram of the accuracy achieved by the proposed approach compared to other methods.

TABLE 13. Performance comparison between the proposed approach and other deep learning models.

	Accuracy	Sensitivity	Specificity	Precision	Nvalue	F-score
PSOBER-DBN	0.973	0.943	0.986	0.963	0.977	0.953
VGGNet	0.940	0.984	0.845	0.932	0.962	0.957
ResNet-50	0.944	0.984	0.871	0.932	0.969	0.957
AlexNet	0.945	0.984	0.879	0.932	0.971	0.957

TABLE 14. Analysis of the results achieved by the proposed and other deep learning models.

	PSOBER-DBN	VGGNet	ResNet-50	AlexNet
Mean	0.9735	0.9406	0.9439	0.9458
Std. Deviation	0.0000	0.0015	0.0018	0.0022
Std. Error of Mean	0.0000	0.0004	0.0005	0.0006

TABLE 15. Two-way ANOVA test.

	SS	DF	MS	F (DFn, DFd)	P-value
Row Factor	0.000073	12	0.000006	F (12, 36) = 4.576	P=0.0002
Column Factor	0.009007	3	0.003002	F (3, 36) = 2244	P<0.0001
Residual	0.000048	36	0.000001		

calculation of the adopted criteria for the achieved results. The largest values of these criteria indicate the superior performance of the corresponding model. As presented in this table, the accuracy of the proposed approach is 97.4%, which outperforms the accuracy achieved by the other deep networks. In addition, the sensitivity, specificity, precision, and Nvalue, which are measured for the proposed approach, outperform those of the deep learning models. These results emphasize the superiority of the proposed approach when compared to deep learning models.

On the other hand, the analysis of the results achieved by the proposed model and the deep learning models is illustrated by the measurements recorded in Table 14. This table presents the mean of the accuracy, standard deviation, and standard error mean when tested on different combinations of test samples. The values in this table emphasize the superiority of the proposed approach when compared to these deep learning models.

Additional tests are performed using the popular two-way ANOVA and Wilcoxon signed rank tests to assess the statistical difference between the proposed approach and the deep learning models. The results of these tests are presented in Table 15 and Table 16. In these tables, the most significant factor is the p-value which indicates the statistical difference when its value becomes lower than 0.05. It can be noted that this value satisfied both tests, which confirms the statistical difference between the proposed approach and the approaches.

TABLE 16. Wilcoxon signed rank test.

	PSOBER-DBN	VGGNet	ResNet-50	AlexNet
Number of values	13	13	13	13
Theoretical median	0	0	0	0
Actual median	0.97	0.94	0.94	0.95
Sum of positive ranks	91	91	91	91
Sum of signed ranks	91	91	91	91
Sum of negative ranks	0	0	0	0
P-value (two tailed)	0.0002	0.0002	0.0002	0.0002
Exact or estimate?	Exact	Exact	Exact	Exact
Discrepancy	0.97	0.94	0.94	0.95

TABLE 17. Evaluating the performance of the proposed optimization algorithm applied to various machine learning models.

Optimized Model	Accuracy	Sensitivity	Specificity	Pvalue	Nvalue	F-score
PSOBER-DBN	0.9735	0.9434	0.9856	0.9634	0.9774	0.9533
PSOBER-SVM	0.9315	0.6667	0.9429	0.3333	0.9851	0.4444
PSOBER-K-NN	0.9298	0.6000	0.9615	0.6000	0.9615	0.6000
PSOBER-LD	0.9125	0.6000	0.9333	0.3750	0.9722	0.4615
PSOBER-DT	0.9091	0.7500	0.9547	0.8257	0.9302	0.7860

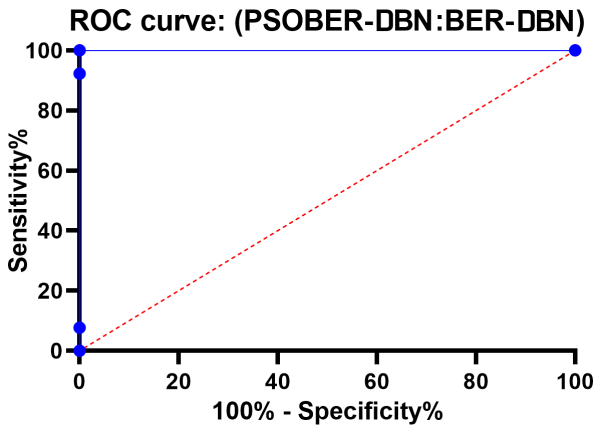


FIGURE 12. The ROC curve of the mapping between the sensitivity and specificity achieved by the proposed approach compared to BER-DBN approach.

To emphasize the effectiveness of the proposed optimization algorithm, additional experiment is presented in Table 17 of Appendix to show the effect of the proposed optimization algorithm on the performance of the set of baseline classification models.

V. CONCLUSION

In this work, a novel optimization algorithm is proposed to optimize the design structure of CNN network for feature extraction and the DBN network for oral cancer classification. These networks are adopted in the proposed approach after assessing other models such as SVM, KNN, LD, and DT and noted a negative impact on the sensitivity of these methods to the dynamic changes that occurred during the optimization process. The proposed algorithm is based on a hybrid of PSO and BER algorithms and is denoted by PSOBER. After data augmentation, a dataset on Kaggle is employed to evaluate the proposed approach. The features necessary for classification are extracted using the optimized convolutional neural network. The extracted features are classified in terms of the optimized deep belief network. To assess the proposed methodology, a set of experiments were conducted. The recorded results confirmed the superiority and effectiveness of the proposed when compared to state-of-the-art machine learning models and optimization algorithms. The best accuracy achieved by the proposed approach is 97.35% which is higher than all other models included in the conducted experiments. The future perspective of this work is to apply the proposed methodology to a larger dataset to investigate its strength and weakness deeply.

APPENDIX. A

An additional experiment is conducted to prove the proposed optimization algorithm's effectiveness in optimizing the baseline models' parameters. In this experiment, the proposed PSOBER is used to optimize the parameters of SVM, K-NN, LD, and DT models. The optimized parameters are then used to classify the images in the adopted dataset. The results of this experiment are presented in Table 17. It can be noted in this table that the classification results based on the optimized models are better than the classification results without model optimization. These results confirm the effectiveness of the proposed optimization algorithm in improving the performance of machine learning models for boosting classification accuracy.

ACKNOWLEDGMENT

The author would like to thank the Deanship of Scientific Research at Shaqra University for supporting this work.

REFERENCES

[1] S. He, R. Chakraborty, and S. Ranganathan, "Proliferation and apoptosis pathways and factors in oral squamous cell carcinoma," *Int. J. Mol. Sci.*, vol. 23, no. 3, p. 1562, Jan. 2022.

[2] R. A. Welikala, P. Remagnino, J. H. Lim, C. S. Chan, S. Rajendran, T. G. Kallarakkal, R. B. Zain, R. D. Jayasinghe, J. Rimal, A. R. Kerr, R. Amtha, K. Patil, W. M. Tilakaratne, J. Gibson, S. C. Cheong, and S. A. Barman, "Automated detection and classification of oral lesions using deep learning for early detection of oral cancer," *IEEE Access*, vol. 8, pp. 132677–132693, 2020.

[3] A. R. Javed, M. U. Sarwar, M. O. Beg, M. Asim, T. Baker, and H. Tawfik, "A collaborative healthcare framework for shared healthcare plan with ambient intelligence," *Hum.-Centric Comput. Inf. Sci.*, vol. 10, no. 1, p. 40, Dec. 2020.

[4] A. R. Javed, L. G. Fahad, A. A. Farhan, S. Abbas, G. Srivastava, R. M. Parizi, and M. S. Khan, "Automated cognitive health assessment in smart Homes using machine learning," *Sustain. Cities Soc.*, vol. 65, Feb. 2021, Art. no. 102572.

[5] J. Musulin, D. Stifanic, A. Zulijani, S. B. Segota, I. Lorencin, N. Andelic, and Z. Car, "Automated grading of oral squamous cell carcinoma into multiple classes using deep learning methods," in *Proc. IEEE 21st Int. Conf. Bioinf. Bioeng. (BIBE)*, Kragujevac, Serbia, Oct. 2021, pp. 1–6.

[6] A. Singh, A. Sahu, and S. Verma, "Computer intelligence in detection of malignant or premalignant oral lesions: The story so far," in *Computational Intelligence in Oncology: Applications in Diagnosis, Prognosis and Therapeutics of Cancers* (Studies in Computational Intelligence), K. Raza, Ed. Singapore: Springer, 2022, pp. 187–200.

[7] K. Saleem, M. Saleem, R. Z. Ahmad, A. R. Javed, M. Alazab, T. R. Gadekallu, and A. Suleman, "Situation-aware BDI reasoning to detect early symptoms of COVID 19 using smartwatch," *IEEE Sensors J.*, vol. 23, no. 2, pp. 898–905, Jan. 2023.

[8] M. K. Hasan, S. Islam, I. Memon, A. F. Ismail, S. Abdullah, A. K. Budati, and N. S. Nafi, "A novel resource oriented DMA framework for Internet of Medical Things devices in 5G network," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 8895–8904, Dec. 2022.

[9] S. Y. Siddiqui, A. Haider, T. M. Ghazal, M. A. Khan, I. Naseer, S. Abbas, M. Rahman, J. A. Khan, M. Ahmad, M. K. Hasan, A. A. Mohammed, and K. Ateeq, "IoMT cloud-based intelligent prediction of breast cancer stages empowered with deep learning," *IEEE Access*, vol. 9, pp. 146478–146491, 2021.

[10] R. F. Mansour, N. M. Alfar, S. Abdel-Khalek, M. Abdelhaq, R. A. Saeed, and R. Alsaqour, "Optimal deep learning based fusion model for biomedical image classification," *Expert Syst.*, vol. 39, no. 3, Mar. 2022, Art. no. e12764.

[11] B. Bhandari, A. Alsadoon, P. W. C. Prasad, S. Abdullah, and S. Haddad, "Deep learning neural network for texture feature extraction in oral cancer: Enhanced loss function," *Multimedia Tools Appl.*, vol. 79, nos. 37–38, pp. 27867–27890, Oct. 2020.

[12] T. Y. Rahman, L. B. Mahanta, A. K. Das, and J. D. Sarma, "Histopathological imaging database for oral cancer analysis," *Data Brief*, vol. 29, Apr. 2020, Art. no. 105114.

[13] J. Lu, N. Sladoje, C. R. Stark, E. D. Ramqvist, J.-M. Hirsch, and J. Lindblad, "A deep learning based pipeline for efficient oral cancer screening on whole slide images," in *Image Analysis and Recognition* (Lecture Notes in Computer Science), A. Campilho, F. Karray, and Z. Wang, Eds. Cham, Switzerland: Springer, 2020, pp. 249–261.

[14] N. Panda and S. K. Majhi, "Oppositional salp swarm algorithm with mutation operator for global optimization and application in training higher order neural networks," *Multimedia Tools Appl.*, vol. 80, nos. 28–29, pp. 35415–35439, Nov. 2021.

[15] K. C. Figueroa et al., "Interpretable deep learning approach for oral cancer classification using guided attention inference network," *Proc. SPIE*, vol. 27, no. 1, Jan. 2022, Art. no. 015001.

[16] A. K. Sharma, A. Nandal, A. Dhaka, and R. Dixit, "Medical image classification techniques and analysis using deep learning networks: A review," in *Health Informatics: A Computational Perspective in Healthcare* (Studies in Computational Intelligence), R. Patgiri, A. Biswas, and P. Roy, Eds. Singapore: Springer, 2021, pp. 233–258.

[17] J. H. Lim, C. S. Tan, C. S. Chan, R. A. Welikala, P. Remagnino, S. Rajendran, T. G. Kallarakkal, R. B. Zain, R. D. Jayasinghe, J. Rimal, A. R. Kerr, R. Amtha, K. Patil, W. M. Tilakaratne, J. Gibson, S. C. Cheong, and S. A. Barman, "D'OraCa: Deep learning-based classification of oral lesions with mouth landmark guidance for early detection of oral cancer," in *Medical Image Understanding and Analysis* (Lecture Notes in Computer Science), B. W. Papiez, M. Yaqub, J. Jiao, A. I. L. Namburete, and J. A. Noble, Eds. Cham, Switzerland: Springer, 2021, pp. 408–422.

[18] F. Martinez, F. Martínez, and E. Jacinto, "Performance evaluation of the NASNet convolutional network in the automatic identification of COVID-19," *Int. J. Adv. Sci., Eng. Inf. Technol.*, vol. 10, no. 2, pp. 662–667, 2020.

[19] M. Z. M. Shamim, S. Syed, M. Shiblee, M. Usman, S. J. Ali, H. S. Hussein, and M. Farrag, "Automated detection of oral pre-cancerous tongue lesions using deep learning for early diagnosis of oral cavity cancer," *Comput. J.*, vol. 65, no. 1, pp. 91–104, Jan. 2022.

[20] H.-A. Li, J. Fan, K. Yu, X. Qi, Z. Wen, Q. Hua, M. Zhang, and Q. Zheng, "Medical image coloring based on Gabor filtering for Internet of Medical Things," *IEEE Access*, vol. 8, pp. 104016–104025, 2020.

[21] Z. Yu, X. Jiang, F. Zhou, J. Qin, D. Ni, S. Chen, B. Lei, and T. Wang, "Melanoma recognition in dermoscopy images via aggregated deep convolutional features," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 1006–1016, Apr. 2019.

[22] R. Rokhana, W. Herulambang, and R. Indraswari, "Deep convolutional neural network for melanoma image classification," in *Proc. Int. Electron. Symp. (IES)*, Sep. 2020, pp. 481–486.

[23] G. Liberman, D. Acevedo, and M. Mejail, "Classification of melanoma images with Fisher vectors and deep learning," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (Lecture Notes in Computer Science), R. Vera-Rodriguez, J. Fierrez, and A. Morales, Eds. Cham, Switzerland: Springer, 2019, pp. 732–739.

[24] Q. Zhou, Y. Shi, Z. Xu, R. Qu, and G. Xu, "Classifying melanoma skin lesions using convolutional spiking neural networks with unsupervised S learning rule," *IEEE Access*, vol. 8, pp. 101309–101319, 2020.

[25] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks," *Multimedia Tools Appl.*, vol. 79, nos. 33–34, pp. 24029–24055, Sep. 2020.

[26] S. Mukherjee, A. Adhikari, and M. Roy, "Malignant melanoma classification using cross-platform dataset with deep learning CNN architecture," in *Recent Trends in Signal and Image Processing* (Advances in Intelligent Systems and Computing), S. Bhattacharyya, S. K. Pal, I. Pan, and A. Das, Eds. Singapore: Springer, 2019, pp. 31–41.

[27] S. Xu, Y. Liu, W. Hu, C. Zhang, C. Liu, Y. Zong, S. Chen, Y. Lu, L. Yang, E. Y. K. Ng, Y. Wang, and Y. Wang, "An early diagnosis of oral cancer based on three-dimensional convolutional neural networks," *IEEE Access*, vol. 7, pp. 158603–158611, 2019.

[28] R. K. Gupta, M. Kaur, and J. Manhas, "Tissue level based deep learning framework for early detection of dysplasia in oral squamous epithelium," *J. Multimedia Inf. Syst.*, vol. 6, no. 2, pp. 81–86, Jun. 2019.

[29] P. R. Jeyaraj and E. R. S. Nadar, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," *J. Cancer Res. Clin. Oncol.*, vol. 145, no. 4, pp. 829–837, Apr. 2019.

[30] B. Song, S. Sunny, R. D. Uthoff, S. Patrick, A. Suresh, T. Kolur, G. Keerthi, A. Anbarani, P. Wilder-Smith, M. A. Kuriakose, P. Birur, J. J. Rodriguez, and R. Liang, "Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning," *Biomed. Opt. Exp.*, vol. 9, no. 11, pp. 5318–5329, Nov. 2018.

[31] R. D. Uthoff, B. Song, S. Sunny, S. Patrick, A. Suresh, T. Kolur, G. Keerthi, O. Spires, A. Anbarani, P. Wilder-Smith, M. A. Kuriakose, P. Birur, and R. Liang, "Point-of-care, smartphone-based, dual-modality, dual-view, oral cancer screening device with neural network classification for low-resource communities," *PLoS ONE*, vol. 13, no. 12, Dec. 2018, Art. no. e0207493.

[32] J. Folmsbee, X. Liu, M. Brandwein-Weber, and S. Doyle, "Active deep learning: Improved training efficiency of convolutional neural networks for tissue classification in oral cavity cancer," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 770–773.

[33] A. Rana, G. Yauney, L. C. Wong, O. Gupta, A. Muftu, and P. Shah, "Automated segmentation of gingival diseases from oral images," in *Proc. IEEE Healthcare Innov. Point Care Technol. (HI-POCT)*, Nov. 2017, pp. 144–147.

[34] R. Anantharaman, M. Velazquez, and Y. Lee, "Utilizing mask R-CNN for detection and segmentation of oral diseases," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 2197–2204.

[35] B. Thomas, V. Kumar, and S. Saini, "Texture analysis based segmentation and classification of oral cancer lesions in color images using ANN," in *Proc. IEEE Int. Conf. Signal Process., Comput. Control (ISPCC)*, Sep. 2013, pp. 1–5.

[36] R. Anantharaman, V. Anantharaman, and Y. Lee, "Oro vision: Deep learning for classifying orofacial diseases," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Aug. 2017, pp. 39–45.

[37] M. M. R. Krishnan, V. Venkatraghavan, U. R. Acharya, M. Pal, R. R. Paul, L. C. Min, A. K. Ray, J. Chatterjee, and C. Chakraborty, "Automated oral cancer identification using histopathological images: A hybrid feature extraction paradigm," *Micron*, vol. 43, nos. 2–3, pp. 352–364, Feb. 2012.

[38] M. Aubreville, C. Knipfer, N. Oetter, C. Jaremenko, E. Rodner, J. Denzler, C. Bohr, H. Neumann, F. Stelzle, and A. Maier, "Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning," *Sci. Rep.*, vol. 7, no. 1, p. 11979, Sep. 2017.

[39] M. Cakmak and M. E. Tenekeci, "Melanoma detection from dermoscopy images using nasnet mobile with transfer learning," in *Proc. 29th Signal Process. Commun. Appl. Conf. (SIU)*, Jun. 2021, pp. 1–4.

[40] T. J. Brinker, A. Hekler, A. H. Enk, C. Berking, S. Haferkamp, A. Hauschild, M. Weichenthal, J. Klode, D. Schadendorf, T. Holland-Letz, C. von Kalle, S. Fröhling, B. Schilling, and J. S. Utikal, "Deep neural networks are superior to dermatologists in melanoma image classification," *Eur. J. Cancer*, vol. 119, pp. 11–17, Sep. 2019.

[41] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park, and S. E. Chang, "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," *J. Investigative Dermatol.*, vol. 138, no. 7, pp. 1529–1538, Jul. 2018.

[42] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Classification of skin lesions using transfer learning and augmentation with Alex-Net," *PLoS ONE*, vol. 14, no. 5, May 2019, Art. no. e0217293.

[43] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[44] G. Anthony, H. Greg, and M. Tshilidzi, "Classification of images using support vector machines," 2007, *arXiv:0709.3967*.

[45] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 20–28, Mar. 2021.

[46] Y. Dang, N. Jiang, H. Hu, Z. Ji, and W. Zhang, "Image classification based on quantum *K*-nearest-neighbor algorithm," *Quantum Inf. Process.*, vol. 17, no. 9, p. 239, Aug. 2018.

[47] M. Zhou, S. Samiappan, E. Worch, and J. E. Ball, "Hyperspectral image classification using Fisher's linear discriminant analysis feature reduction with Gabor filtering and CNN," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 493–496.

[48] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," Dec. 2015, *arXiv:1511.08458*.

[49] S.-H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, and H. Cheng, "Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling," *J. Med. Syst.*, vol. 42, no. 5, p. 85, Mar. 2018.

[50] S. Wang, Y. Jiang, X. Hou, H. Cheng, and S. Du, "Cerebral micro-bleed detection based on the convolution neural network with rank based average pooling," *IEEE Access*, vol. 5, pp. 16576–16583, 2017.

[51] C. Li, Y. Wang, X. Zhang, H. Gao, Y. Yang, and J. Wang, "Deep belief network for spectral–spatial classification of hyperspectral remote sensor data," *Sensors*, vol. 19, no. 1, p. 204, Jan. 2019.

[52] R. Bello, Y. Gomez, A. Nowe, and M. M. Garcia, "Two-step particle swarm optimization to solve the feature selection problem," in *Proc. 7th Int. Conf. Intell. Syst. Des. Appl. (ISDA)*, Oct. 2007, pp. 691–696.

[53] E.-S. M. El-kenawy, A. A. Abdelhamid, A. Ibrahim, S. Mirjalili, N. Khodadad, M. A. Al duailij, A. Ali Alhussan, and D. Sami Khafaga, "Al-Biruni Earth radius (BER) metaheuristic search optimization algorithm," *Comput. Syst. Sci. Eng.*, vol. 45, no. 2, pp. 1917–1934, 2023.

[54] M. Ouyang, J. Xi, W. Bai, and K. Li, "Band-area resource management platform and accelerated particle swarm optimization algorithm for container deployment in Internet-of-Things cloud," *IEEE Access*, vol. 10, pp. 86844–86863, 2022.

[55] H.-E. Tseng, C.-C. Chang, and T.-W. Chung, "Applying improved particle swarm optimization to asynchronous parallel disassembly planning," *IEEE Access*, vol. 10, pp. 80555–80564, 2022.

[56] M. E. C. Bento, D. Dotta, R. Kuiava, and R. A. Ramos, "A procedure to design fault-tolerant wide-area damping controllers," *IEEE Access*, vol. 6, pp. 23383–23405, 2018.

[57] S. Fahad, S. Yang, S. U. Khan, S. A. Khan, and R. A. Khan, "A hybrid smart quantum particle swarm optimization for multimodal electromagnetic design problems," *IEEE Access*, vol. 10, pp. 72339–72347, 2022.

[58] D. S. Khafaga, A. A. Alhussan, E.-S.-M. El-Kenawy, A. Ibrahim, M. M. Eid, and A. A. Abdelhamid, "Solving optimization problems of metamaterial and double T-shape antennas using advanced meta-heuristics algorithms," *IEEE Access*, vol. 10, pp. 74449–74471, 2022.

[59] A. A. Alhussan, D. S. Khafaga, E.-S.-M. El-Kenawy, A. Ibrahim, M. M. Eid, and A. A. Abdelhamid, "Pothole and plain road classification using adaptive mutation dipper throated optimization and transfer learning for self driving cars," *IEEE Access*, vol. 10, pp. 84188–84211, 2022.

[60] C. Öksüz, O. Urhan, and M. K. Güllü, "Brain tumor classification using the fused features extracted from expanded tumor region," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103356.

[61] S. Baro. (2020). *Oral Cancer (Lips and Tongue) Images Dataset*. Accessed: Oct. 4, 2022. [Online]. Available: https://www.kaggle.com/code/shivam17299/oral-cancer-lips-and-tongue-images-dataset/data

[62] A. Saber, M. Sakr, O. M. Abo-Seida, A. Keshk, and H. Chen, "A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique," *IEEE Access*, vol. 9, pp. 71194–71209, 2021.

[63] J. Rashid, M. Ishfaq, G. Ali, M. R. Saeed, M. Hussain, T. Alkhalifah, F. Alturise, and N. Samand, "Skin cancer disease detection using transfer learning technique," *Appl. Sci.*, vol. 12, no. 11, p. 5714, Jun. 2022.

[64] Q. Al-Tashi, S. J. A. Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Binary optimization using hybrid grey wolf optimization for feature selection," *IEEE Access*, vol. 7, pp. 39496–39508, 2019.

[65] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.

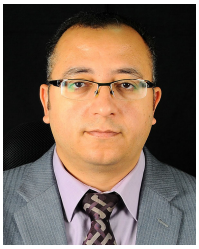
[66] S. D. Immanuel and U. K. Chakraborty, "Genetic algorithm: An approach on optimization," in *Proc. Int. Conf. Commun. Electron. Syst. (ICCES)*, Jul. 2019, pp. 701–708.

[67] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, Dec. 1997.

HADJOUNI MYRIAM has academic and work experience with the Computer Science Department, Faculty of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University. Her main research interests include data science, artificial intelligence, machine learning, data mining, and software engineering. She is currently a reviewer for some journals.

ABDELAZIZ A. ABDELHAMID received the M.Sc. degree in computer science from the Faculty of Computer and Information Sciences, Ain Shams University, and the Ph.D. degree in computer engineering from the Faculty of Engineering, Auckland University, New Zealand. He is currently an Assistant Professor with the Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University. He is also an Assistant Professor with the Computer Science Department, College of Computing and Information Technology, Shaqra University. His research interests include speech and image processing and machine-learning-based intelligent systems.

EL-SAYED M. EL-KENAWY (Senior Member, IEEE) is currently an Assistant Professor with the Higher Institute for Engineering and Technology (DHIET), Mansoura, Egypt. He has published more than 35 articles with more than 1200 citations, and his H-index is 22. He has launched and pioneered independent research programs. He motivates and inspires his students in different ways by providing a thorough understanding of various computer concepts. He explains complex concepts in an easy-to-understand manner. His research interests include artificial intelligence, machine learning, optimization, deep learning, digital marketing, and data science. He is a Reviewer of *Computers, Materials, and Continua* journal, IEEE Access, and other journals.



ABDELHAMEED IBRAHIM (Member, IEEE) received the bachelor's and master's degrees in engineering from the Computer Engineering and Systems Department, in 2001 and 2005, respectively, and the Ph.D. degree in engineering from the Faculty of Engineering, Chiba University, Japan, in 2011. He was with the Faculty of Engineering, Mansoura University, Egypt, from 2001 to 2007, where he is currently an Associate Professor of computer engineering. He has published more than 90 publications with over 2600 citations, and his H-index is 28. His research interests include machine learning, optimization, swarm intelligence, metaheuristics, and pattern recognition. He serves as a Reviewer for *Engineering Applications of Artificial Intelligence*, IEEE Access, IEEE Journal of Biomedical and Health Informatics, *Biomedical Signal Processing and Control*, *Computer Systems Science and Engineering*, *Sensors*, *PLOS ONE*, *Knowledge and Information Systems*, and other respected journals.

MARWA METWALLY EID received the Ph.D. degree in electronics and communications engineering from the Faculty of Engineering, Mansoura University, Egypt, in 2015. She was an Assistant Professor with the Delta Higher Institute for Engineering and Technology, from 2011 to 2021. She has been an Assistant Professor with the Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura, Egypt, since 2022. Her current research interests include image processing, encryption, wireless communication systems, and field programmable gate array (FPGA) applications.

MONA M. JAMJOOM has an academic and work experience with the Computer Science Department, Faculty of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University. Her main research interests include data science, artificial intelligence, machine learning, data mining, and software engineering. She is currently a reviewer for some journals.

DOAA SAMI KHAFAGA received the B.Sc. degree (Hons.) in computer and information sciences, the M.Sc. degree in computer science, and the Ph.D. degree in computer science from the Faculty of Computers and Artificial Intelligence, Helwan University, Egypt, in 2003, 2008, and 2013, respectively. She has 18 years of academic experience. She has worked with the Computer Science Department, College of Information Technology and Artificial Intelligence, Misr University for Science and Technology, Egypt; the Computer Science Department, Institute of Public Administration, Saudi Arabia; and the Computer Science Department, Faculty of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University. Her main research interests include data science, artificial intelligence, machine learning, data mining, and software engineering. She has FHEA and the fellow recognition from U.K. Higher Education Academy. She is currently a reviewer for some Journals.