

Research Article

Skin Cancer Detection Using Kernel Fuzzy C-Means and Improved Neural Network Optimization Algorithm

Jia Huaping ¹, Zhao Junlong², and A. M. Norouzzadeh Gil Molk ³

¹College of Computer, Weinan Normal University, Weinan, Shaanxi, China

²Rehabilitation Medicine Department, Weinan Central Hospital, Weinan, Shaanxi, China

³Department of Computer Engineering, University of Guilan, Rasht, Iran

Correspondence should be addressed to A. M. Norouzzadeh Gil Molk; norouzzadeh.ali@guilan.ac.ir

Received 20 May 2021; Revised 20 June 2021; Accepted 6 July 2021; Published 19 July 2021

Academic Editor: V. Rajinikanth

Copyright © 2021 Jia Huaping et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Early diagnosis of malignant skin cancer from images is a significant part of the cancer treatment process. One of the principal purposes of this research is to propose a pipeline methodology for an optimum computer-aided diagnosis of skin cancers. The method contains four main stages. The first stage is to perform a preprocessing based on noise reduction and contrast enhancement. The second stage is to segment the region of interest (ROI). This study uses kernel fuzzy C-means for ROI segmentation. Then, some features from the ROI are extracted, and then, a feature selection is used for selecting the best ones. The selected features are then injected into a support vector machine (SVM) for final identification. One important part of the contribution in this study is to propose a developed version of a new metaheuristic, named neural network optimization algorithm, to optimize both parts of feature selection and SVM classifier. Comparison results of the method with 5 state-of-the-art methods showed the approach's higher superiority toward the others.

1. Introduction

Cancer is fundamentally an uncontrolled cell division disease. Its growth is usually associated with a group of variations in the action of the cell cycle regularized. The inhibitors of the cell cycle stop cell division while situations are not proper; therefore, so little action of these inhibitors can cause cancer. Cancer cells disregard the signals that must cause cells to prevent dividing. For example, when common cells growing in a culture medium are surrounded on all sides by adjacent cells, they no longer divide (contact inhibition). On the contrary, cancer cells divide into layers in a mass and stack on top of each other, and contact inhibition does not prevent them from growing. One of the most dangerous cancers around the world is skin cancer. The skin, as the first cell layer of the body that is in connection with the outside environment, can suffer from many injuries and diseases. The most common cancer in the United States is skin cancer, which happens in the tissues of the largest section of the skin body. Skin cancer generally occurs on the

outer layer of the skin, which may first appear as a swelling, bulge, or different section of the skin [1]. The skin blocks heat, sunlight, sores, and infections; it also warms the body and stores fat water in the body and produces vitamin D. The skin owns 2 principal layers: the superficial layer of the skin (epidermis) and the inner skin (dermis) [2]. The skin's surface layer (epidermis) is mainly composed of flat, scaly cells named Squamous cells. Round cells, called basal cells, are located beneath the superficial cells of the squamous melanocyte in the surface layer [3]. Prolonged exposure to sunlight enhances melanoma skin cancer's risk over a lifetime. One of the most dangerous kinds of cancer is Melanoma, which is caused by damage caused by overexposure to the sun and some other factors. Melanoma can be diagnosed with skin biopsies. Figure 1 shows the statistical information of the lead cancers in quantity and death value in 2019 [4].

If melanoma is diagnosed and treated early, it can be treated, and late diagnosis can lead to the patient's death. Only experienced physicians can diagnose melanoma on

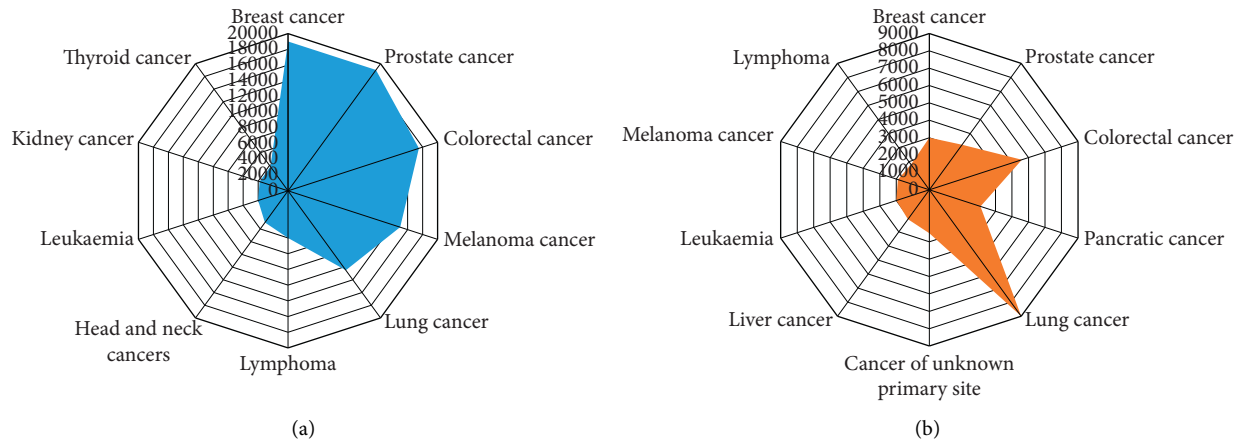


FIGURE 1: The statistical information of the lead cancers in (a) quantity and (b) death value in 2019 [4].

time using appropriate tools and histological reports. One of the devices used to diagnose melanoma is dermatoscopy. With this tool, changes in the pigmentation of skin lesions in diseases can be evaluated. With the development of science in recent years, digital dermatoscopy has been replaced by conventional dermatoscopes with the ability to capture and store skin images. With the development of digital dermatoscopes, the development of an algorithm for the diagnosis of melanoma was considered by researchers, and then, other methods were proposed to enhance the accuracy of skin lesion diagnosis. Some of the selected researches are explained in the following.

Astorino et al. [5] suggested an approach for melanoma diagnosis. The method uses multiple instance learning (MIL) approach on some dermoscopy images to categorize between melanomas and common nevi. The method was then analyzed by comparing it with some state-of-the-art classification methods, like support vector machine. Simulation outcomes demonstrated that the suggested approach accomplishes better precision. A leave-one-out authentication is carried out on a dataset and showed the method's higher efficiency.

Mohamed et al. [6] suggested an image processing approach for melanoma diagnosis. The method used a simple pipeline image preprocessing technique including methods like image conversion, noise reduction, and hair removal. Also, the melanoma area is threshold to segment the area, and then, feature extraction was used for taking out the main features of the image using the ABCD rule.

Barros et al. [7] proposed a real-time cancer diagnosis system based on hardware designing by the field-programmable gate arrays (FPGA). A multilayer perceptron artificial neural networks (ANN) is utilized for this purpose. The proposed image processing is used to extract the characteristics of the skin from the images and classify them based on the features and using the ANN classifier. Simulation results of the proposed approach are confirmed by an

open-access database. Ultimate results of the suggested approach were then compared with the results of another hardware-based technique using ARM A9 microprocessor to demonstrate the method's proper performance.

Santos and Espitia [8] presented an approach for the diagnosis of uveal melanoma (UM), which is a sort of intraocular cancer. The suggested technique combined iris segmentation methods and designed a method for UM diagnosis based on neural networks and fuzzy logic. Simulation results of the study indicated 96.04% accuracy for the artificial neural networks and 76% correct classification for the fuzzy logic system.

Wang [9] proposed an all-inclusive method to provide a proper segmentation technique. The method is applied based on deep convolution networks on the hyperspectral pathology images. The study employed a 3D fully convolutional network, called Hyper-net, to obtain the best performances of segmentation from the hyperspectral pathology images. The loss function was then modified for improving the model sensitivity. Ultimate outcomes demonstrated that the suggested approach has higher performance than the 2D models in the aim of segmentation.

The present research suggests a novel optimized pipeline methodology for automatic detection of skin cancer detection. The method includes four main phases. The first phase is preprocessing of the input image to prepare it for the main processing. The first phase here includes two parts: noise reduction and contrast enhancement. The next phase is to segment the region of interest. Afterwards, the features are extracted, and the best ones selected for injecting to a support vector machine (SVM) as the final phase. This investigation employs an improved version of a new meta-heuristic, called neural network optimization algorithm, to optimize both parts of feature selection and SVM classifier. This can improve the accuracy of the system as can be seen in the next sections. The main contribution of the paper has been highlighted as follows:

- (i) An early diagnosis system for skin cancer from dermoscopy images is proposed
- (ii) Kernel fuzzy C-means is used to segment the region of interest (ROI)
- (iii) Some features from the ROI are extracted for the final diagnosis
- (iv) An improved neural network algorithm is used for optimal feature selection
- (v) Support vector machine (SVM) is used for final classification
- (vi) Comparison results are compared with 5 state-of-the-art methods

Figure 2 illustrates the flowchart of the suggested approach.

2. Image Preprocessing

2.1. Noise Reduction. Ordinary nonpolarized light is used in medical images. Since, according to “Fresnel relations,” the reflection of light from the surface of matter and its underlying layers and scattering from uneven and rough surfaces is affected, as well as a function of the state of polarization of light, it is possible to eliminate some annoying and destructive reflected or diffused lights by using a suitable filter to achieve a high signal to noise ratio. Therefore, the first step for starting the image analysis of the cancer images is to eliminate, or, at least, reduce, this noise to have a high quality of images for processing. Different methods have been introduced for this purpose. One simple and popular method for noise reduction is Median filtering. However, this filter is nonlinear; it has some significant advantages such as

keeping the main edges of the image after filtering. In the median filter, the pixels have been substituted by the median value of their neighbors, i.e.,

$$Z_{m,n} = \text{median} \left(y_{i,j}; (i,j) \in \beta \right), \quad (1)$$

where β signifies the neighborhood of the considered pixel in position (m,n) .

The present study uses a 3×3 mask for the median filter. This selection is based on trial and error and may be different for other databases. Thus, it should be the point that using a higher mask can improve the probability of edges losing. Figure 3 shows an example of using the median filtering for noise reduction of the input images with considering 0.1 noise density salt and pepper noises.

2.2. Contrast Enhancement. In some situations, due to different reasons, like the sufficient user experience’s lack in imaging and the bad quality of measurement devices and sensors utilized in them, the contrast of the images has lessened, such that its intensity gets darkened, or overexposed. To correct this deficiency from the images, we need to implement the image contrast enhancement. Therefore, in this study, we can use an image contrast enhancement step on the low-quality images to improve the image contrast to

simplify the segmentation step. The present study considers global contrast enhancement based on Lookup Table to reach this purpose [10]. The method of global contrast enhancement is formulated as follows:

$$\text{Out}_H = \frac{\text{In}_H - \text{Min}_H}{\text{Max}_H - \text{Min}_H}, \quad (2)$$

where Out_H signifies the improved output image, and Min_H and Max_H demonstrate the minimum and the maximum levels of gray values of the original image histogram, orderly. We used an 8-bit lookup table for the method. A simple contrast enhancement on a medical image is shown in Figure 4.

3. Medical Image Processing

3.1. Segmentation Based on Kernel-Based Fuzzy C-Means.

One of the significant issues in medical image processing is the image segmentation of the medical image into its components. Image segmentation describes the success or the final failure of image analysis methods. However, there is no general method for successful segmentation of all medical images, and it still has good research areas due to its wide application. The accuracy of this study is crucial in areas such as medicine, which helps preserve and protect human life. Due to the wide range of applications, it provides image segmentation and the use of methods in various fields.

The importance of this part especially in mammographic images is too high, because medical images have a naturally low quality, which makes them a complicated problem for mass segmentation. This research uses an enhanced multistep fuzzy C-means method for performing on the skin cancer images for their mass segmentation. Based on this method, the first step is to generate superpixels using a simple linear iterative clustering (SLIC) algorithm [11]. The SLIC algorithm generates superpixels based on CIE LAB color space in 5D space. Afterwards, texture and color properties are employed to separate the superpixels and finally to find the image elements.

Then, the color feature from the superpixels has been extracted and has been arranged, such that, for all images in HSV space, three separate histograms for S, H, and V channels have been considered. To reduce time complexity, the histograms have been quantized into 8, 4, and 2 sub-distances. Afterward, to find the texture features, the NSCT algorithm has been used. Then, the generated data is clustered by a fuzzy kernel, and all superpixels with cluster tags have been placed in one cluster. This study uses kernel fuzzy C-means (KFCM) algorithm employing segmentation. This technique offers a kernel-based version of fuzzy C-means to evaluate the data point’s distance from the center of clusters. The kernel function in this study is achieved as follows:

$$K = \exp\left(\frac{-x - y^2}{\rho}\right). \quad (3)$$

First, the membership function has been evaluated. Then, based on a similarity measure of kernel fuzzy, the belonging of each data sample to the clusters has been

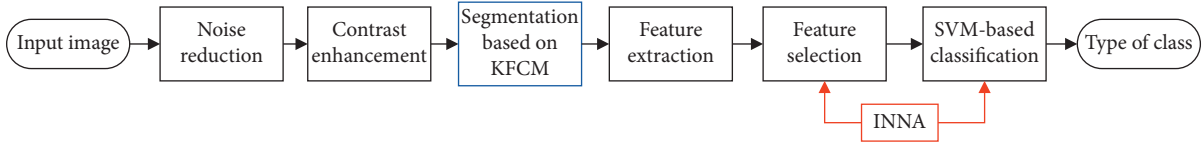


FIGURE 2: The flowchart of the suggested approach.

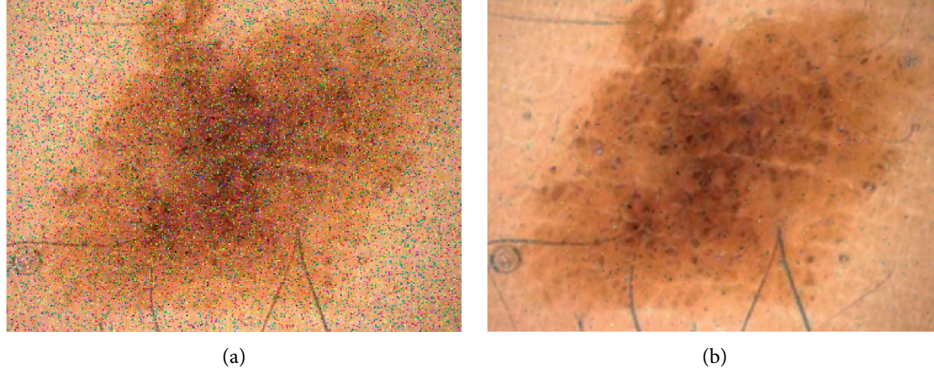


FIGURE 3: Image noise reduction: (a) image plus 0.1 salt and pepper noises and (b) image after noise reduction with median filtering.

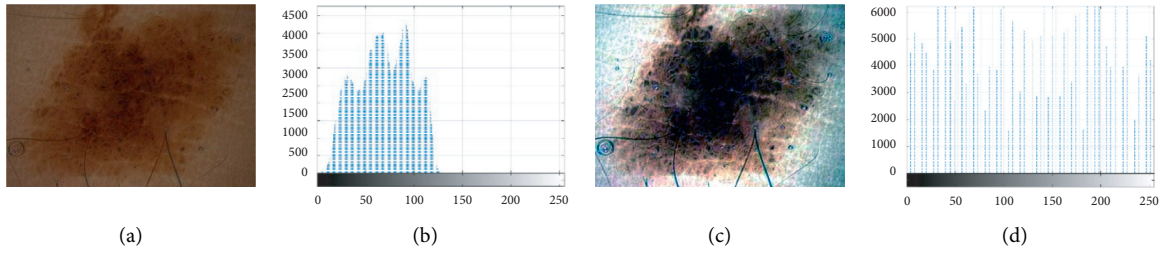


FIGURE 4: (a) Original Image with low contrast. (b) Histogram of (A). (c) Contrast improvement of (A). (d) Histogram of (C).

achieved. The pseudocode of the algorithm is given as follows:

- (1) Apply KFCM to cluster the set of objects and generating a U membership matrix [12]
- (2) In all elements, x_i, x_j , the number of t nearest neighbor is found.
- (3) If x_i, x_j are not t nearest neighborhood, $W_{ij} = 0$.
- (4) Else if they belong to a cluster, $W_{ij} = 1$.
- (5) Else, $W_{ij} = \exp(\ln 2 \times (u_i \oplus u_j))$, where \oplus defines the exclusive OR that totally indicates the overlap between two fuzzy sets.
- (6) Here, diagonal matrix, D , is formulated as follows:

$$D_{ij} = \sum_{j=1}^n W_{ij}. \quad (4)$$

And, it is normalized as follows:

$$L = D^{-1/2} W D^{1/2}. \quad (5)$$

The number of K bigger eigenvector than L (the first vector) is found (the first vector is selected), and the

matrix $P = [p^1, p^2, \dots, p^k]$ has been formed, and then, the algorithm normalizes the rows in the P matrix to form matrix Y .

- (7) Each line of Y has considered a point in space R^k and at then, the final clustering has been established by the K-means algorithm.

Figure 5 shows two sample examples of skin cancer segmentation based on the proposed KFCM methodology.

4. Improved Neural Network Algorithm

4.1. Optimization. Generally speaking, in most applications of engineering, optimization is a vital subject. Optimization is to make the best decision to get the optimal (minimum or maximum) result for the considered problem. Several methods have been introduced for optimization. However, classic methods as exact methods can solve these problems, and they fail in some cases that the problem is nonlinear or complicated. To overcome this problem, another technique, called metaheuristic, has been introduced [13]. Metaheuristics include a set of optimization algorithms that are inspired by nature, human behaviors, animals' competitions, etc. The main advantage of using metaheuristic algorithms is

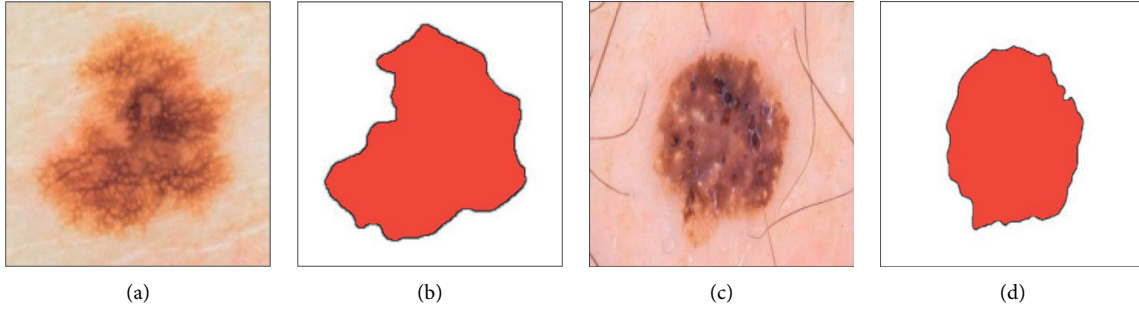


FIGURE 5: Two sample examples of skin cancer segmentation based on the proposed KFCM methodology.

that they use random structure instead of using gradient methods. This simplifies the optimization process [14]. Furthermore, the most complicated problems are that their number is increasing day by day and they cannot be solved by the classic methods, but these methods, based on their stochastic nature, can find a near-optimal result in a logical time. There are different types of these algorithms like ant lion optimizer (ALO) algorithm [15], chimp optimization (CO) algorithm [16], Harris Hawks optimization [17], and world cup optimization (WCO) algorithm [18]. Recently, a new metaheuristic algorithm, called neural network algorithm (NNA), is introduced, which is inspired by the concepts of biological nervous systems and artificial neural networks (ANNs). Based on [19], each ANN includes some artificial neurons, which are inspired by the biological nervous systems. The relationships among units principally specify the network function. The NNA approach is illustrated in the following.

4.2. Neural Network Algorithm. Like any other metaheuristic algorithm, the NNA starts with an initial population that is called the pattern. In NNA, just like ANN, the main idea is to update the pattern population to minimize error among the forecasted data and the desired output data. Here, the best solution is the desired output that can be updated in each iteration for achieving the minimum error amount by moving the pattern population in the direction of the desired solution. In the following, the algorithm methodology is briefly explained.

4.2.1. Initialization. The first operation in this algorithm, like any other metaheuristic, is to generate some random population (decision variables) for initial evaluation. By considering a D dimensional problem, the considered pattern solution vector will be an array of $1 \times D$, representing input data in the NNA. Moreover, by considering the D dimension and N_{pop} several random candidates, the initial pattern population can be considered as follows:

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_D^1 \\ x_1^2 & x_2^2 & x_3^2 & x_D^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{N_{\text{pop}}} & x_2^{N_{\text{pop}}} & \dots & x_D^{N_{\text{pop}}} \end{bmatrix}, \quad (6)$$

where the matrix X is made randomly between the minimum and maximum limitations of a problem.

And, the cost value is achieved as follows:

$$C_i = f(x_1^i, x_2^i, \dots, x_D^i), \quad (7)$$

where f describes the objective value.

After evaluating the objective value of each pattern solution, the best one is selected as the best pattern solution. The NNA has N_{pop} input data and D dimension with only one target data. Figure 6 shows this structure.

After defining the target solution (X^T) among the other pattern solutions, its weight (W^T) has to be chosen from the population of weight (weight matrix).

In an ANN, the neurons connect with dendrite using a simple summation. The output is connected to the input layers based on weighted (w) interconnection. Initial weights are random values in this algorithm, and then, they have been updated based on some equations to provide the minimum network error. The initial weights in the algorithm are as follows:

$$W(t) = \begin{bmatrix} w_{11} & \dots & w_{i1} & \dots & w_{N_{\text{pop}}1} \\ w_{12} & \dots & w_{i2} & \dots & w_{N_{\text{pop}}2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{1N_{\text{pop}}} & \dots & w_{iN_{\text{pop}}} & \dots & w_{N_{\text{pop}}N_{\text{pop}}} \end{bmatrix}_{N_{\text{pop}} \times N_{\text{pop}}}, \quad (8)$$

where W includes uniformly distributed random values in the range $[0, 1]$. The first and the second subscripts of weight relate to the pattern solution and participated with the other solutions of the pattern. All pattern solutions have their weight value to generate a new candidate solution. These weights have also a constraint that should not exceed 1 and mathematically formulated as follows:

$$\sum_{j=1}^{N_{\text{pop}}} w_{ij}(t) = 1, \quad (9)$$

$$w_{ij} \in U(0, 1),$$

where $i, j = 1, 2, \dots, N_{\text{pop}}$.

Weight is randomly distributed values in the range between 0 and 1, where their aggregate in a solution of the pattern must not be more than one, which is because of the bias control of the movement and to generate the new

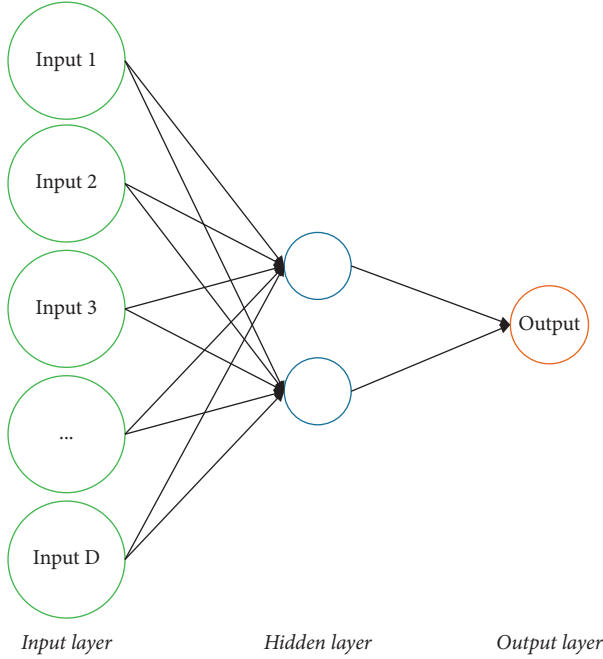


FIGURE 6: The structure of NNA.

pattern solutions. If the constraint is not considered, the probability of sticking in the local optimum by the values of the weight will be increased. After weight matrix formation, new pattern solutions (X^N) are evaluated as follows:

$$X_j^{\text{New}}(t+1) = \sum_{i=1}^{N_{\text{pop}}} w_{ij}(t) \times X_i(t), \quad (10)$$

$$X_i(t+1) = X_i(t) + X_i^{\text{New}}(t+1).$$

After achieving the new pattern solutions using the best weight amount that is named target weight, this updating is established by the following equation:

$$W_i^{\text{updated}}(t+1) = W_i(t) + 2 \times \text{rand} \times (W^{\text{Target}}(t) - W_i(t)). \quad (11)$$

Another operator for updating in NNA is the bias operator. This operator is employed for modifying the exploration part of the algorithm by using a specific percentage of the candidates in the new population ($X_i^{\text{New}}(t+1)$) and updated weight matrix ($W_i^{\text{updated}}(t+1)$). The pseudocode of this operator is as follows:

```

For I = 1 to  $N_{\text{pop}}$ 
  If  $\text{rand} \leq \gamma$ 
     $N_b = \text{round}(D \times \gamma)$ 
    For  $j = 1: N_b$ 
       $X_{\text{Input}}(I, \text{integer rand}[0, D]) = \text{LB} + (\text{UB} - \text{LB}) \times \text{rand}$ 
    End for
     $N_{\text{wb}} = \text{round}(N_{\text{pop}} \times \gamma)$ 
    For  $j = 1: N_{\text{wb}}$ 
       $W^{\text{Updated}}(j, \text{integer rand}[0, N_{\text{pop}}]) = U(0, 1)$ 

```

```

End for
End if
End for

```

Here, LB and UB represent the problem minimum and maximum bounds, orderly. γ describes the modification factor that defines the candidates' percentage. $\gamma = 1$ at first, and then, it decreases gradually during the process based on any decreasing formulation as follows:

$$\gamma(t+1) = 0.99 \times \gamma(t). \quad (12)$$

In the NNA, there is also a transfer function operator to transfer the new candidates to the new updated positions to provide better value against the target solution. The solution reimproved by giving the novel pattern solutions to the best solution direction (target solution). The transfer function operator (TF) can be formulated as follows:

$$X_i^*(t+1) = \text{TF}(X_i(t+1)) = X_i(t+1) + 2 \times \text{rand} \times (X^{\text{Target}}(t) - W_i(t+1)). \quad (13)$$

The pseudocode of the collaboration between bias and TF operators is tabulated below:

```

For  $i = 1$  to  $N_{\text{pop}}$ 
  If  $\text{rand} \leq \beta$ 
    Bias operator
  Else ( $\text{rand} > \beta$ )
    Apply the transfer function operator
  End if
End for

```

4.3. Improved Neural Network Algorithm. However, the neural network algorithm provides good results to solve different applications of optimization problems [20–22], and it is sometimes stuck in the premature convergence that gives a high impact on the solution accuracy. In this investigation, to modify this drawback, 2 mechanisms are considered. The first mechanism is to employ the Chaos theory. This mechanism explores novel situations, which are dynamic discrete-time, i.e.,

$$\beta_{m+1} = f(\beta_m), \quad m = 0, 1, 2, \dots \quad (14)$$

In this research, the logistic map function has been used as the modification mechanism. This mechanism could be mathematically modeled as follows:

$$\beta_{j,i+1,q} = \rho \times \beta_{j,i,q} (1 - \beta_{j,i,q}), \quad (15)$$

where i signifies the populations' number, $\beta_{j,i,q}$ defines the value for the i^{th} chaotic iteration, ρ describes a constant, which is set 4, q demonstrates the iteration number, j illustrates the generators' number in the system, and β_0 describes the initial value of β_i with a random value between 0 and 1 [23, 24]. With assuming the above explanation, the set of initial variables w_{ij} is rewritten as follows:

$$w_{ij} = L_j + (U_j - L_j) \times \beta_{j,i,q}. \quad (16)$$

This study also used Lévy flight (LF) as a second modification. The LF is a popular technique for improving optimization algorithms [25]. The LF uses random motion to control the local searching. This is obtained by the following equations:

$$\begin{aligned} Le(w) &\approx \frac{1}{w^{1+\tau}}, \\ w &= \frac{A}{\sqrt{|B|}}, \\ \sigma^2 &= \left\{ \frac{\Gamma(1+\tau)}{\tau\Gamma((1+\tau)/2)} \times \frac{\sin(\pi\tau/2)}{2^{(1+\tau)/2}} \right\}^{2/\tau}, \end{aligned} \quad (17)$$

where τ represents the Lévy flight index in the range [0, 2] (here, $\tau = 1.5$ [26]), w describes the step size, $A, B \sim N(0, \sigma^2)$, $\Gamma(\cdot)$ describes Gamma function, and the samples generate a Gaussian distribution with σ^2 variance and zero mean value. By assuming the LF mechanism, the updated equation for the new pattern solutions based on the best weight value is achieved as follows:

$$\begin{aligned} W_i^{\text{updated}}(t+1) &= W_i(t) + 2 \times \text{rand} \times Le(\delta) \\ &\quad \times (W^{\text{Target}}(t) - W_i(t)). \end{aligned} \quad (18)$$

4.4. Algorithm Validation. To assess the effectiveness of the proposed improved neural network algorithm, six standard benchmark functions have been utilized, and the results are compared with some new state-of-the-art algorithms, containing multiverse optimizer (MVO) [27], moth-flame optimization (MFO) algorithm [28], world cup optimizer (WCO) [19], and the original neural network algorithm (NNA). The parameter settings of the algorithms are as follows: for MVO: wormhole existence probability (WEP); $WEP_{\min} = 0.2$; $WEP_{\max} = 1$; Coefficient(P) = 6. For MFO: logarithmic spiral shaped(b) = 1; for WCO: $\alpha = 0.5$, payoff = 4%. The population size for whole algorithms is set to 100, and their maximum iteration is set to 100. Table 1 tabulates the studied benchmark functions for the analysis. The benchmark functions are considered with 30 dimensions.

To validate the algorithms based on the studied functions, four statistical indicators including minimum amount (Min), maximum amount (Max), mean amount (Mean), and standard deviation amount (Std) are employed, and the results are indicated in Table 2.

According to Table 2, the outcomes of the suggested INNA have the proper outcomes in terms of lower, higher, and amount. Because it provides the minimum value for these values that is the main purpose of these benchmark functions, therefore, the proposed INNA has the highest accuracy. Furthermore, the proposed method with a minimum standard deviation value provides the highest reliable solution for the studied benchmark functions.

5. Feature Extraction and Selection Based on INNA

Feature selection is one of the most important methods and techniques for data preprocessing and data mining. Due to the introduction of new programs for large data mining, media information retrieval, and medical data processing that requires the processing of large volumes of data, it is important to limit the number of features. The purpose of feature extraction here is to take out the important characteristics from the region of interest in skin cancer images to simplify the diagnosis process. In this paper, 19 different features have been used for extraction from the image. Table 3 indicates the used features for the extraction.

Therefore, here, we utilized a feature selection technique to reduce and select some more important features and to eliminate the low-cost features. Different methods have been introduced for this purpose. Metaheuristics are a kind of new technique for feature selection.

Feature selection using metaheuristics is one of the subsets of feature extraction and is discussed in various fields of machine learning and data mining. In general, this problem does not have a definite solution, and so far, no precise method has been proposed to solve it. Various classical approaches have been proposed to these problems, but usually, the quality of their solutions is generally not very good. In contrast, intelligent optimization methods can provide far better solutions to these problems. Therefore, one of the effective and constructive methods in solving feature selection problems and related issues is the use of metaheuristic optimization methods and evolutionary algorithms.

Here, we employed a cost function for optimum selection of the features using four indicators, including false positive (FP), false negative (FN), true positive (TP), and true negative (TN). The cost function is formulated as follows:

$$\text{fitness} = \frac{(\text{TP} \times \text{TN}) - (\text{FP} \times \text{FN})}{\sqrt{((\text{TN} + \text{FP}) \times (\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FN}))}} \quad (19)$$

The idea is to minimize the abovementioned function with selecting the high cost and important features. Based on the designed INNA, the above function has been minimized, and less-cost features are eliminated from the processing. The next step is to employ an efficient method for classifying these features.

After performing the binary optimization technique on the features, the following features with higher effectiveness are selected for training: elongation, area, mean, correlation, entropy, and elongation. This selection is performed based on a metaheuristic technique based on a stochastic nature.

6. Skin Cancer Image Classification

Classification is the process of separating an input image into predetermined classes. Image classification is considered the last part of the diagnosis in medical image analysis. This step

TABLE 1: The applied test functions for the analysis.

Function	Formula	Minimum	Limitation
Rosenbrock	$f_1 = \sum_{i=1}^{D-1} \{100(x_{i+1} - x_i)^2 + (x_i - 1)^2\}$	0	$[-30, 30]^D$
Sum squares	$f_2 = \sum_{i=1}^D ix_i^2$	0	$[-10, 10]^D$
Step 2	$f_3 = \sum_{i=1}^D (x_i + 0.5)^2$	0	$[-100, 100]^D$
Schwefel 2.22	$f_4 = \sum_{i=1}^D x_i - \prod_{i=1}^D x_i $	0	$[-10, 10]^D$
Schwefel 1.2	$f_5 = \sum_{i=1}^D (\sum_{j=1}^i x_j)^2$	0	$[-100, 100]^D$
Chung Reynolds	$f_6 = (\sum_{i=1}^D x_i^2)^2$	0	$[-100, 100]^D$

TABLE 2: The simulation results of the comparative algorithms on the studied benchmark functions, moth-flame optimization (MFO) algorithm [28], world cup optimizer (WCO) [19], and the original neural network algorithm (NNA).

Algorithm		f_1	f_2	f_3	f_4	f_5	f_6
MVO [27]	Min	17.58	0.0024	0.0031	19.38	0	$6.37e-8$
	Max	$7.29e+3$	436.15	$3.86e+5$	$2.37e+3$	$5.39e-7$	$4.19e-8$
	Mean	$2.96e+3$	280.16	$4.19e+4$	67.25	$6.15e-9$	$1.27e-7$
	Std	$5.79e+2$	95.37	$6.37e+4$	11.62	$3.19e-8$	$3.07e-7$
MFO [28]	Min	28.14	2.57	0.051	6.29	$4.38e-6$	8.39
	Max	129.08	4.18	9.27	53.49	0.015	110.35
	Mean	80.46	23.09	4.52	30.17	0.02	35.08
	Std	22.46	1.64	1.73	3.28	0.01	24.63
WCO [19]	Min	7.39	$5.13e-5$	$5.92e-5$	2.57	$1.94e-8$	$6.38e-9$
	Max	$1.38e+2$	0.264	0.017	4.12	$7.51e-8$	$7.29e-8$
	Mean	105.37	0.0573	0.024	3.28	$6.48e-9$	$1.18e-9$
	Std	14.83	0.0873	$3.28e-5$	1.23	$4.29e-9$	$3.19e-8$
NNA [29]	Min	4.91	$6.29e-21$	$9.75e-9$	0.017	$9.37e-16$	$5.9e-17$
	Max	45.22	$34.39e-18$	$1.19e-8$	4.62	$2.58e-12$	$5.19e-16$
	Mean	13.83	$6.30e-18$	$5.94e-8$	0.532	$3.42e-13$	$6.67e-16$
	Std	5.29	$3.62e-21$	$6.53e-9$	0.42	$1.97e-13$	$7.50e-17$
INNA	Min	5.16	$15.26e-22$	$6.37e-11$	$6.38e-15$	$7.26e-7$	$6.71e-39$
	Max	111.57	$6.76e-20$	$4.29e-10$	$3.95e-14$	0.0234	$9.43e-35$
	Mean	24.13	$2.48e-20$	$1.17e-10$	$1.09e-14$	$3.11e-5$	$4.57e-6$
	Std	12.28	$6.19e-21$	$5.64e-11$	$9.64e-15$	$11.97e-7$	$9.55e-36$

TABLE 3: Utilized features for the extraction.

Parameter	Equation	Parameter	Equation
Elongation	$\sqrt{\text{number of tumor pixels}/2 \times (\text{Max radius})^2}$	Compactness	$(2 \times \sqrt{\text{number of tumor pixels} \times \pi})/\text{Area}$
Area	$\sum_{i=1}^M \sum_{j=1}^N p(i, j)$	Mean	$1/MN (\sum_{i=1}^M \sum_{j=1}^N p(i, j))$
Perimeter	$\sum_{i=1}^M \sum_{j=1}^N B_p(i, j)$	Correlation	$\sum_{i=1}^M \sum_{j=1}^N p(i, j) - \mu_r \mu_c / \sigma_r \sigma_c$
Variance	$1/MN (\sum_{i=1}^M \sum_{j=1}^N (p(i, j) - \mu)^2)$	Solidity	$\text{Area}/\text{convex area}$
Invariant moments	$\varphi_1 = \eta_{20} + \eta_{02}$	Entropy	$-\sum_{i=1}^M \sum_{j=1}^N p(i, j) \log p(i, j)$
	$\varphi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$		
Elongation	$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$	Eccentricity	$2a^{-1} (a^2 - b^2)^{0.5}$
	$2\sqrt{\text{Area}/a\sqrt{\pi}}$		
Rectangularity	$\text{Area}/b \times a$	Energy	$\sum_{i=1}^M \sum_{j=1}^N p^2(i, j)$
Irrregularity index	$4\pi \times \text{area}/\text{perimeter}^2$	Standard deviation	$\text{Variance}^{1/2}$
Form factor	Area/a^2		

MN describes the image size, B_p defines the external side length for the boundary pixel, $p(i, j)$ defines the pixels intensity amount at position (i, j) , μ and σ describe the mean value and the standard deviation, orderly, and a and b present the major axis and the minor axis, respectively. However, some of the above features have a high impact, and some others have a low impact on feature extraction.

can decrease the time for the cancer diagnosis process by decreasing the search space in it. A popular and useful method in classification is to use a support vector machine (SVM). The SVM seeks the best hyperplane that operates as a multipart (here two-part) data separator in the input space.

For reaching the constraints of the optimization problem, the support vector machine has been employed to evaluate the normal vector, w , in the hyperplane, the bias b , and the slack variable, η for incorrectly assigned training patterns that support the generalization, which is defined as follows:

$$y = \min \frac{w^2}{2} + C \times \sum_{i=1}^n \eta_i. \quad (20)$$

With separation function:

$$y_i(q \times x_i + b) \geq 1 - \eta_i. \quad (21)$$

Feature space is highly affected by the separating function in equation (21), so that a function selection needs to be performed correctly to generate optimal output. Here, for minimizing equation (20), the suggested INNA has been utilized. Accordingly, the support vector machine uses a kernel that is employed to alter data to the feature space. In this research, three different kernels including linear, polynomial, radial basis function (RBF), and sigmoid have been used. Figure 7 shows a typical SVM.

As can be explained, the SVM technique is a proper tool for classifying the images. Medical images are also a part of this event. Therefore, after performing feature extraction on the images for decreasing the operation complexity, the skin cancer features for each image can be injected into the SVM, and the SVM will present the final classification, which will be healthy or cancer.

7. Evaluation and Results

7.1. Implementation Details. As mentioned before, the proposed skin cancer detection system contains different modules that are generated in MATLAB R2019b environment. The modules include image segmentation, image acquisition, feature extraction and choice, and the last classification. In the image acquisition stage for dermoscopic images, American Cancer Society (ACS) [30] and PH² [31] databases are used:

- (i) The ACS database: the ACS database includes 68 pairs of XLM and TLM images that are integrated by the Nervoscope system. Experimental analysis is established by a 16 GB RAM Intel Core i7 processor. This database is downloaded one by one from <https://www.cancer.org/cancer/skin-cancer.html>.
- (ii) The PH² database: dermoscopic images of this database are collected from the Dermatology Service of Hospital Pedro Hispano (Matosinhos, Portugal) under the same conditions. These images are 8-bit RGB color images with a resolution of 768 × 560 pixels. The image database includes a total of 200 dermoscopic images of melanocytic lesions, including 80 common nevi, 80 atypical nevi, and 40 melanomas. This database can be downloaded from: <https://www.fc.up.pt/addi/ph2%20database.html>.

The training and the test data for the benchmarks are set to 80% and 20%, respectively.

7.2. Results. As previously mentioned, the recommended simulations have been carried out to the ACS dataset to investigate the system efficiency. For numerical analysis of the suggested methodology, three popular measurement

indicators for classification problems have been used. These metrics are accuracy (ACC), precision (PR), and sensitivity (SN). By comparing the ground-truth with the output mask, the classification indicators are evaluated as follows:

$$\begin{aligned} \text{ACC} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100, \\ \text{PR} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100, \\ \text{SN} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100, \end{aligned} \quad (22)$$

where TP, TN, FP, and FN denote the true positive, true negative, false positive, and false negative.

So, by analyzing the efficiency of the INNA-SVM method by considering different kernels based on the abovementioned indicators, we have the following.

It can be observed from Table 4 that the linear kernel of the proposed INNA-SVM, in addition to simplicity, results in the best values for all indicators for both databases: sensitivity, accuracy, and precision. Therefore, this kernel is utilized here for the classification of the features. For a comprehensive investigation about the proposed method's efficiency, five evaluation measurements including specificity (SP), positive predictive value (PPV), negative predictive value (NPV), F1 score, and Matthews correlation coefficient (MCC) are considered, which are mathematically as follows:

$$\begin{aligned} \text{SP} &= \frac{\text{TN}}{\text{FP} + \text{TN}}, \\ \text{PPV} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{NPV} &= \frac{\text{TN}}{\text{TN} + \text{FN}}, \\ \text{F1 score} &= \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}, \\ \text{MCC} &= \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}. \end{aligned} \quad (23)$$

The general proposed method has been compared with five new methods, and the comparison results have been indicated in Table 5. The compared methods include Astorino's method [5] based on multiple instance learning (MIL), Hassan's method [6] based on a simple pipeline image preprocessing technique, Barros's method [7] based on hardware designing by the multilayer perceptron, Santos's method [8] based on neural networks and fuzzy logic, and Wang's method [9] based on deep convolution networks. The results are achieved from the mean value of both databases.

According to Table 5, the F1-score of the suggested method with 84.39% that defines its accuracy based on the precision and recall of the test data is the highest among all of

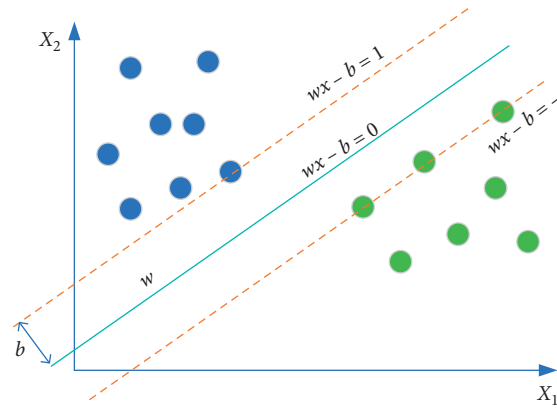


FIGURE 7: A typical model of an SVM.

TABLE 4: The classification results of INNA-SVM method based on different kernels in two databases.

Database	Kernel	ACC (%)	Pr (%)	SN (%)
ACS [30]	Linear	64.19	65.18	64.19
	RBF	37.57	38.09	37.57
	Polynomial	46.29	38.61	46.29
	Sigmoid	40.16	41.53	40.16
	Linear	70.00	69.00	71.00
PH ² [31]	RBF	42.00	43.00	42.00
	Polynomial	51.00	45.00	51.00
	Sigmoid	42.50	46.00	53.00

TABLE 5: The performance comparison between the suggested method and the analyzed methods.

Method	MCC (%)	SP (%)	PPV (%)	NPV (%)	F1-score (%)
The proposed method	92.35	88.67	81.43	92.66	81.69
Astorino's method [5]	72.29	87.49	81.37	86.72	69.29
Hassan's method [6]	75.46	80.76	69.35	85.87	63.49
Barros's method [7]	83.65	60.34	68.37	83.98	70.81
Santos's method [8]	85.74	60.41	71.95	86.69	76.76
Wang's method [9]	88.67	80.77	74.49	83.71	82.92

the compared methods. Also, the value of MCC in the proposed method with 94.21% is the highest among the others, and because it uses all of the analysis terms, it shows the method's higher efficiency than the others. Also, the higher value of the NPV and PPV with 95.57% and 85.09% against the other methods shows higher condition occurrence of the method to control the likelihood of a test identifying cancer toward the others. Finally, the higher SP value based on the proposed method reports higher occurrence-independent results of the algorithm.

8. Conclusions

Skin cancer is one of the most common cancers, and malignant melanoma is the most invasive and deadliest kind of this cancer. Early detection of this cancer by physicians has a

high effect on the treatment of this cancer. Recently, several machine vision-based techniques are used for helping physicians with accurate early detection. This paper proposed a new approach for optimal skin cancer detection. The first section in the proposed method was to establish a preprocessing technique including noise reduction and contrast enhancement. Then, the region of interest (ROI) was segmented by a kernel fuzzy C-means segmentation method. The features of the ROI were then extracted, and the most important features were selected optimally. Afterward, the selected features were injected into an optimal classifier using support vector machine (SVM) for classification. The main contribution of this study was to present a developed version of a new metaheuristic, called neural network optimization algorithm, to optimize both parts of feature selection and SVM classifier in the system. The superiority of

the proposed method was proved by performing a comparison among the suggested approach and five state-of-the-art methods.

Data Availability

The data are available at <https://www.cancer.org/>.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] R. Navid, "Computer-aided diagnosis of skin cancer: a review," *Current Medical Imaging*, vol. 16, no. 7, pp. 781–793, 2020.
- [2] S. Ahuja, B. K. Panigrahi, N. Dey, V. Rajinikanth, and T. K. Gandhi, "Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices," *Applied Intelligence*, vol. 51, no. 1, pp. 571–585, 2021.
- [3] Q. Liu, Z. Liu, S. Yong, K. Jia, and N. Razmjooy, "Computer-aided breast cancer diagnosis based on image segmentation and interval analysis," *Automatika*, vol. 61, no. 3, pp. 496–506, 2020.
- [4] A. Costa, Y. Kieffer, A. Scholer-Dahirel et al., "Fibroblast heterogeneity and immunosuppressive environment in human breast cancer," *Cancer Cell*, vol. 33, no. 3, pp. 463–479, 2018.
- [5] A. Astorino, A. Fuduli, P. Veltri, and E. Vocaturo, "Melanoma detection by means of multiple instance learning," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 12, no. 1, pp. 24–31, 2020.
- [6] S. H. Mohamed, S. D. Noha, A. M. Tarek, and H. E. May, "Detection of melanoma using image processing techniques driven tds index," *Journal of Environmental Science*, vol. 46, no. 2, pp. 1–46, 2019.
- [7] W. K. P. Barros, D. S. Morais, F. F. Lopes, M. F. Torquato, R. D. M. Barbosa, and M. A. C. Fernandes, "Proposal of the CAD system for melanoma detection using reconfigurable computing," *Sensors*, vol. 20, no. 11, p. 3168, 2020.
- [8] D. F. Santos and H. E. Espitia, "Detection of uveal melanoma using fuzzy and neural networks classifiers," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 4, pp. 2213–2223, 2020.
- [9] Q. Wang, "Identification of melanoma from hyperspectral pathology image using 3D convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 218–227, 2020.
- [10] R. Navid, B. S. Mousavi, and F. Soleymani, "A real-time mathematical computer method for potato inspection using machine vision," *Computers & Mathematics with Applications*, vol. 63, no. 1, pp. 268–279, 2012.
- [11] K.-S. Kim, D. Zhang, M.-C. Kang, and S.-J. Ko, "Improved simple linear iterative clustering superpixels," in *Proceedings of the 2013 IEEE International Symposium on Consumer Electronics (ISCE)*, pp. 259–260, IEEE, Hsinchu, Taiwan, June 2013.
- [12] M. R. P. Ferreira and F. A. T. de Carvalho, "Kernel fuzzy c-means with automatic variable weighting," *Fuzzy Sets and Systems*, vol. 237, pp. 1–46, 2014.
- [13] R. Navid and R. Saeid, "Skin melanoma segmentation using neural networks optimized by quantum invasive weed optimization algorithm," in *Metaheuristics and Optimization in Computer and Electrical Engineering*, pp. 233–250, Springer, Berlin, Germany, 2021.
- [14] D. B. Mehdi, "A new improved model of marine predator algorithm for optimization problems," *Arabian Journal for Science and Engineering*, Springer, Berlin, Germany, 2021.
- [15] M. Mani, O. Bozorg-Haddad, and X. Chu, "Ant lion optimizer (ALO) algorithm," in *Advanced Optimization by Nature-Inspired Algorithms*, pp. 105–116, Springer, Berlin, Germany, 2018.
- [16] M. Khishe and M. R. Mosavi, "Chimp optimization algorithm," *Expert Systems with Applications*, vol. 149, Article ID 113338, 2020.
- [17] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019.
- [18] N. Razmjooy, M. Khalilpour, and M. Ramezani, "A new meta-heuristic optimization algorithm inspired by FIFA world cup competitions: theory and its application in PID designing for AVR system," *Journal of Control, Automation and Electrical Systems*, vol. 27, no. 4, pp. 419–440, 2016.
- [19] Z. Zhang, "Artificial neural network," in *Multivariate Time Series Analysis in Climate and Environmental Research*, pp. 1–35, Springer, Berlin, Germany, 2018.
- [20] Y. Zhang, Z. Jin, and Y. Chen, "Hybrid teaching-learning-based optimization and neural network algorithm for engineering design optimization problems," *Knowledge-Based Systems*, vol. 187, Article ID 104836, 2020.
- [21] M. Elsis, K. Mahmoud, M. Lehtonen, and M. M. F. Darwish, "An improved neural network algorithm to efficiently track various trajectories of robot manipulator arms," *IEEE Access*, vol. 9, pp. 11911–11920, 2021.
- [22] Y. Zhang, Z. Jin, and Y. Chen, "Hybridizing grey wolf optimization with neural network algorithm for global numerical optimization problems," *Neural Computing and Applications*, vol. 32, no. 14, pp. 10451–10470, 2020.
- [23] D. Yang, G. Li, and G. Cheng, "On the efficiency of chaos optimization algorithms for global optimization," *Chaos, Solitons & Fractals*, vol. 34, no. 4, pp. 1366–1375, 2007.
- [24] C. Rim, S. Piao, G. Li, and U. Pak, "A niching chaos optimization algorithm for multimodal optimization," *Soft Computing*, vol. 22, no. 2, pp. 621–633, 2018.
- [25] Y. Cao, Y. Wu, L. Fu, K. Jermstiparsert, and N. Razmjooy, "Multi-objective optimization of a PEMFC based CCHP system by meta-heuristics," *Energy Reports*, vol. 5, pp. 1551–1559, 2019.
- [26] X. Li, P. Niu, and J. Liu, "Combustion optimization of a boiler based on the chaos and Lévy flight vortex search algorithm," *Applied Mathematical Modelling*, vol. 58, pp. 3–18, 2018.

- [27] S. Mirjalili, S. M. Mirjalili, and A. Hatamlou, "Multi-verse optimizer: a nature-inspired algorithm for global optimization," *Neural Computing and Applications*, vol. 27, no. 2, pp. 495–513, 2016.
- [28] S. Mirjalili, "Moth-flame optimization algorithm: a novel nature-inspired heuristic paradigm," *Knowledge-based Systems*, vol. 89, pp. 228–249, 2015.
- [29] A. Sadollah, H. Sayyaadi, and A. Yadav, "A dynamic meta-heuristic optimization model inspired by biological nervous systems: neural network algorithm," *Applied Soft Computing*, vol. 71, pp. 747–782, 2018.
- [30] ACS, "Skin cancer database," 2019, <https://www.cancer.org/>.
- [31] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "PH 2-A dermoscopic image database for research and benchmarking," in *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5437–5440, IEEE, Osaka, Japan, July 2013.