

Research Article

Breast Cancer Diagnosis by Convolutional Neural Network and Advanced Thermal Exchange Optimization Algorithm

Xiuzhen Cai,¹ Xia Li ,² Navid Razmjooy ,³ and Noradin Ghadimi ⁴

¹Community Health Service Center, Taibei Street, Jiangan District, Wuhan, 430010 Hubei Province, China

²Wuhan Vocational College of Software and Engineering, Wuhan, 430205 Hubei, China

³Department of Engineering, Tafresh University, Tafresh, Iran

⁴Young Researchers and Elite Club, Islamic Azad University, Ardabil Branch, Ardabil, Iran

Correspondence should be addressed to Xia Li; lixia95272021@163.com, Navid Razmjooy; xnavid@gmail.com, and Noradin Ghadimi; noradin.ghadimi@iauardabil.ac.ir

Received 7 February 2021; Accepted 6 October 2021; Published 8 November 2021

Academic Editor: Maria Evelina Fantacci

Copyright © 2021 Xiuzhen Cai 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.

A common gynecological disease in the world is breast cancer that early diagnosis of this disease can be very effective in its treatment. The use of image processing methods and pattern recognition techniques in automatic breast detection from mammographic images decreases human errors and increments the rapidity of diagnosis. In this paper, mammographic images are analyzed using image processing techniques and a pipeline structure for the diagnosis of the cancerous masses. In the first stage, the quality of mammogram images and the contrast of abnormal areas in the image are improved by using image contrast improvement and a noise decline. A method based on color space is then used for image segmentation that is followed by mathematical morphology. Then, for feature image extraction, a combined gray-level cooccurrence matrix (GLCM) and discrete wavelet transform (DWT) method is used. At last, a new optimized version of convolutional neural network (CNN) and a new improved metaheuristic, called Advanced Thermal Exchange Optimizer, are used for the classification of the features. A comparison of the simulations of the proposed technique with three different techniques from the literature applied on the MIAS mammogram database is performed to show its superiority. Results show that the accuracy of diagnosing cancer cases for the proposed method and applied on the MIAS database is 93.79%, and sensitivity and specificity are obtained 96.89% and 67.7%, respectively.

1. Introduction

Breast cancer is common cancer and is the first cause of cancer mortality in women. Breast cancer is a malign tumor that develops from cells in the same organ. The disease usually begins in the lobules, or breast ducts, and can then penetrate the ducts and walls of the glands and attack the surrounding adipose tissue or even other parts of the body. There may be other lumps in the breast that are not cancerous, but in any case, the final diagnosis is up to the physicians. Science has proven that despite the existence of a way to prevent cancer and the lack of definitive treatment for this disease, early diagnosis of this disease helps physicians to at least prevent the progression of this disease. The number of cancer patients in the world is increasing. Based on the World

Health Organization (WHO), breast cancer has a great effect on about 2.1 million women annually. Based on this statistic, in 2018, 627,000 women have died of breast cancer which contains about 15% of all deaths among women cancer [1]. Figure 1 shows the statistical information of the cancer diagnosis and the cancer deaths in 2019 [2].

The best solution to decline breast cancer mortality is to diagnose it in the primary stage and treat it. Early diagnosis needs a precise and dependable diagnostic method. Among the various methods of diagnosing breast cancer, mammography is a very common and very popular method. Systematic screening of the female population with mammograms and early diagnosis of early stage breast cancer can also increase the patients' survival chances and reduce the negative side effects of necessary treatments. These results are

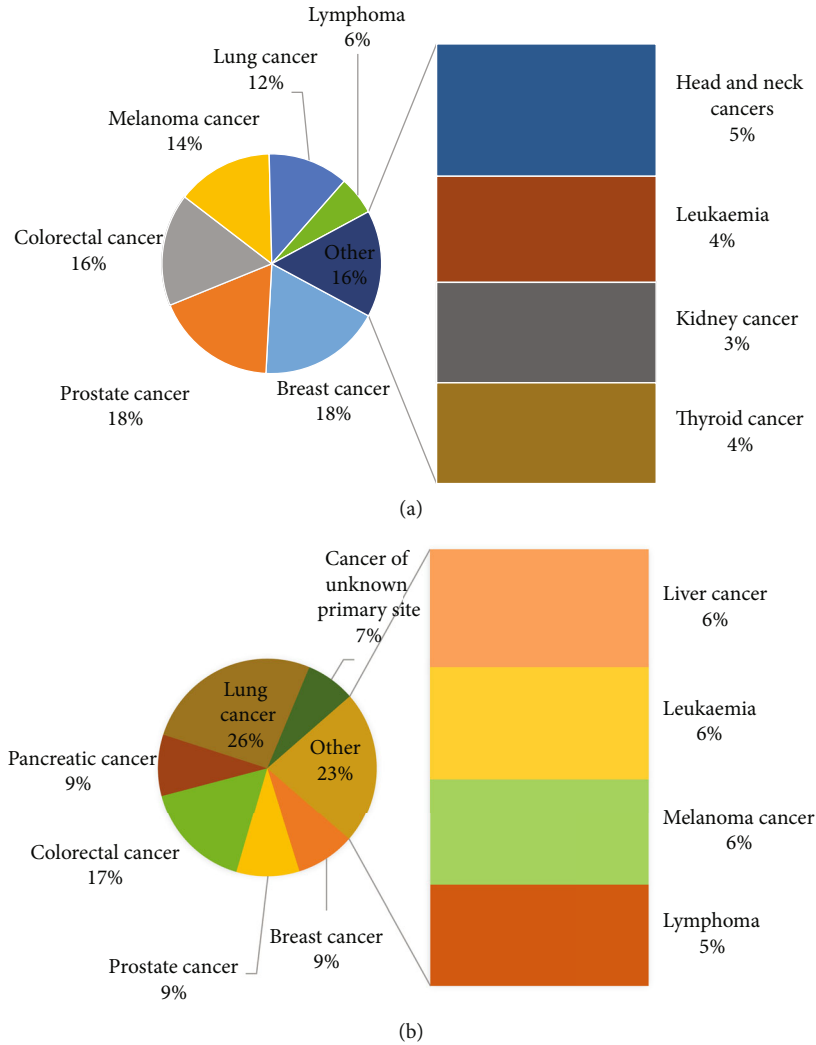


FIGURE 1: The statistical information of the cancers (a) and cancer deaths (b) in 2019 [2].

possible if the quality of services is provided in the best possible way.

On the other hand, the diagnosis of breast cancer based on mammography film has several problems. In some cases, there is a possibility that the film is damaged or the image is not suitable for diagnosis. Meanwhile, the film wears out over time, and the possibility of revision decreases. On the other hand, the only tool a doctor can use to diagnose a lesion is a visual video. The physician's visual observations to diagnose the lesion leads to two errors. The first is that a radiograph is shown twice to a physician or radiologist. If he does not know that both images are the same, his diagnosis may be different. Another error is showing an image to two physicians or two radiologists, each of whom has a different diagnosis. Although mammography is usually the best way to diagnose breast cancer, some classes of cancers are not diagnosed in this way. In this condition, providing the computer-aided systems can detect malignant lesions efficiently [1]. Based on the literature, the effectiveness of a computer diagnostic system is more precise than that of a physician. Recently, various kinds of research works were

performed in the area of the automatic early detection of breast cancers [3]. In other words, the higher efficiency of the computer-aided systems assists the physicians to diagnose cancer with lower complexity and higher speed. For instance, Liu et al. [4] proposed a proper image segmentation methodology for optimal breast cancer diagnosis regarding the interval uncertainties. For considering the indeterminacy, interval assessment was utilized. The method is guaranteed to providing suitable results in any changes in the imaging system. The main idea was to provide an interval-based Laplacian of Gaussian filter to simulate the intensity uncertainties. Final results have been performed to MIAS database, and a comparison of the results with some classic techniques was carried out to illustrate the method efficiency.

Toğaçar et al. [5] introduced a breast cancer diagnosis system by CNN. The method was improved by a technique called BreastNet. The image data was established by the expansion technique before implementing it into the model. Then, a precise classification method was accomplished based on the hypercolumn technique. Then, a comparison

of method results with some latest techniques was done to state the higher precision of the suggested system.

Carvalho et al. [6] used another method for breast cancer detection for utilizing in histopathological images. The authors presented a method for using phylogenetic diversity indexes to determine images to model creation and histopathological breast image classification into some classes. The method was then compared with several different latest techniques to demonstrate the technique's accuracy. The results showed significant robustness to the method to help experts at large medical centers.

It can be concluded from the previous studies that many works have been done for the automatic diagnosis of breast cancer. This paper proposes a new automated method for breast cancers diagnosed in mammogram images. Here, an optimized deep learning-based methodology based on a new improved metaheuristic, called Advanced Thermal Exchange Optimization algorithm, has been used for this purpose.

2. Image Preprocessing

The heterogeneity of light intensity in medical images has weakened the boundaries of medical images, especially the heterogeneity of light intensity in magnetic resonance images created by nonuniform magnetic fields by radiofrequency coils, which is why preprocessing is so important in medical research. Therefore, after obtaining the input information of the medical images, preprocessing operations should be performed, which are methods to eliminate noise and isolate and improve the differentiation of areas where there is a possibility of numerical information.

2.1. Image Contrast Improvement. Commonly, in images, several forms of contrast issues are existing, for instance, the inappropriate lighting and room conditions, the deficiency of enough applicant interface for imaging, and the inadequate quality of the measuring sensors and devices. These variables will disappear with some essential details, darkening or overexposure, and finally the image abnormalities. The presence of these variables increases their need for enhancement in numerous medical images. Lack of these parameters in most numbers of medical images increases their requirement for improvement. This improvement has been performed based on contrast enhancement.

Here, the contrast enhancement has been applied to the images to highlight the skin cancer areas with no changes on the other areas. A simple application to perform a piecewise linear contrast stretch operation on an image. The present study uses a 16-bit lookup table to improve the contrast of the images that are then stored on a disc. This is implemented based on the following formula:

$$y_{\text{hist}} = \frac{x_{\text{hist}} - \text{Min}_{\text{hist}}}{\text{Max}_{\text{hist}} - \text{Min}_{\text{hist}}}, \quad (1)$$

where Max_{hist} and Min_{hist} stand for the highest and the lowest levels for the gray magnitudes of the main image histogram, respectively, and x_{hist} and y_{hist} represent the input

image before contrast enhancement and the output image after image contrast enhancement, respectively.

2.2. Noise Reduction. As aforementioned, due to different conditions in the medical imaging, there are some kinds of noises in them that should be removed before processing. This is done by using noise reduction. Noises can be white, random, or Gaussian (that contains a large part of medical images). Noise is usually in the high-frequency bands of the image; the important edges and details of the image are in the same bands. Therefore, noise removal along with preserving the edges and important image information is the main problem in the image noise removal process. Many noise removal techniques have been proposed in recent years. One of the proper methods for noise removal in medical images is to use the Wang-Mendel algorithm. This algorithm is a beneficial technique based on fuzzy theory [7]. Because of the simple conception of the fuzzy theory, the method of the Wang-Mendel algorithm is so easy to understand. Also, due to the higher speed of this method, it is too valuable for initial fuzzy model creation [8]. The method for the rule database is achieved by the following:

- (1) Perform fuzzy separation from the input variable space which can be obtained based on the knowledge or using normalization technique. Afterward, it classifies it into two parts including equal or unequal by performing a fuzzy separation of the input variable space. Afterward, the membership function has been selected and the components have been given as a fuzzy package. Afterward, membership has been chosen and a fuzzy set is given to every part
- (2) Creation candidate language rules which can be formed by choosing the most all-encompassing laws for the samples
- (3) Level of validity assignment to the laws that are achieved based on multiplying the membership function values of the components and the membership function value of the result of the law
- (4) Provide final database rules from the collection of candidate language which is done by classifying the candidate rules into various groups, where each of them contains candidate rules with similar assumptions. For obtaining the final rule base, the maximum degree of verification law is achieved in every set

3. Image Segmentation

3.1. Changing Color Space. However RGB has a good concept for the human, it is completely dependent on the three colors (red, green, and blue), as base colors of the RGB color space. This color space has also a high dependency on the ambient light intensity which limits its usages in a different application. To recover this issue, different color spaces have been introduced. In this study, the XYZ color space is utilized for the purpose after some trials and errors. The XYZ color space makes a link between the physiologically

- (i) Compute the histogram and probabilities of each intensity level
- (ii) Give primary values of $\omega_i(0)$ and $\mu_i(0)$ for all feasible threshold levels (highest intensity ($t = 1, 2, \dots$))
- (iii) Renew ω_i and μ_i
- (iv) Calculate $\sigma_b^2(t)$
- (v) The optimal threshold here is the maximum of $\sigma_b^2(t)$

PSEUDOCODE 1: The pseudocode of the Otsu algorithm.

supposed colors in human color vision and distributions of wavelengths in the visible spectrum of the electromagnetic. In the XYZ color model, Y states the luminance, and X and Z indicate the color information. The formula for converting the RGB to XYZ is as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \times \begin{bmatrix} 0.49 & 0.31 & 0.2 \\ 0.17697 & 0.8124 & 0.01063 \\ 0 & 0.01 & 0.99 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (2)$$

The most significant benefit of the XYZ color model is that it is completely independent of the device.

3.2. Method of Segmentation. The red (R) color space provides the main dimension in RGB color space to give nearly the image intensity in medical images. As previously mentioned, both X and Z values provide similar color information for the XYZ color space. So, for segmentation, X dimension and red dimension are only normalized, i.e.,

$$\begin{aligned} \hat{R} &= \frac{R}{\sqrt{R^2 + G^2 + B^2}}, \\ \hat{X} &= \frac{X}{\sqrt{X^2 + Y^2 + Z^2}}. \end{aligned} \quad (3)$$

These normalization values are carried out on every pixel in the input images. The Otsu threshold is used after this normalization to provide a low-cost segmentation in the sense of time complexity.

By intergroup variance optimization and lessening pixels' intragroup variance, the Otsu process is an efficient way that is established to automatically pick the optimal threshold. There is an issue with the global threshold when the resolution of the image background is insufficient. To remove the heterogeneity effect, it is possible to use a local threshold. To remove inhomogeneity and add a global threshold to the processed image, this problem is solved by image preprocessing.

Based on the Otsu method, the threshold value has been searched which minimizes the class-in-between variance as follows:

$$\sigma_\omega^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t), \quad (4)$$

where ω_i describes the probability for two different groups with a threshold magnitude of t and σ_i^2 represents the vari-

ance amount of the groups. In other words, Otsu indicates that variance minimization of a class is like maximization of the variance in class-within, i.e.,

$$\sigma_b^2(t) = \sigma^2 - \sigma_\omega^2(t) = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2, \quad (5)$$

where μ_i describes the mean value. The Otsu algorithm will be defined as the following pseudocode.

Subsequently, for better performance, postprocessing mathematical morphology, including filling, closing, and opening, was carried out on the images [9]. First, extra holes in the image are filled by applying the mathematical filling operator. The analytical model is as follows:

$$X_k = (X_{k-1} \oplus B) \cap A^c, \quad k = 1, 2, 3 \dots, \quad (6)$$

where A and B describe the area that should be processed and the constructing element, respectively.

Afterward, the mathematical opening has been employed to the filled image to eliminate the ignitor information with no adjustments on other gray surfaces. The mathematical model of this operator is as follows:

$$A \circ B = (A \ominus B) \oplus B. \quad (7)$$

Then, the mathematical closing is executed based on the following equation for linking the narrow parts:

$$A \cdot B = (A \oplus B) \ominus B. \quad (8)$$

This study uses a 5×5 identity matrix as a structural element. Figure 2 gives some examples of the breast segmentation of images depending on the approach suggested.

4. Feature Extraction

Feature extraction is the process of reducing the dimensional of the images by dividing and reducing an initial set of images to more controllable groups. So, the next processing of the images can be simpler by this process. Of these large datasets, the most important characteristic is that they have numerous parameters. To process them, these parameters need many calculating origins. Therefore, extraction of these features helps to select and combine variables into features to get the best feature from those big datasets, thereby effectively decreasing the data volume. In this study, two popular features including DWT and GLCM were employed for feature extraction that is described subsequently.

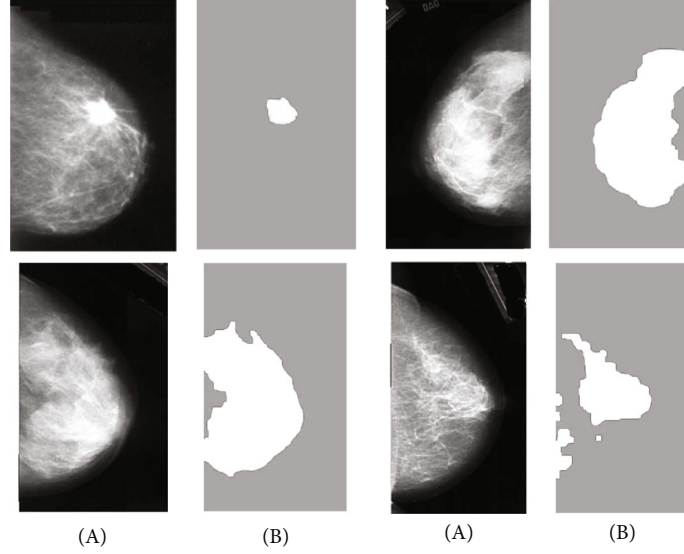


FIGURE 2: Several instances of the breast segmentation of images depending on the approach suggested: (a) basic image and (b) segmented image.

4.1. Discrete Wavelet Transform (DWT). Wavelet transform is an efficient tool for various applications in image processing and is utilized in numerous fields such as image noise removal, pattern recognition, coding, image compression, and feature extraction. Wavelet transform is a method in the frequency domain. In this method, instead of using sine and cosine functions (such as Fourier transform), a function called wavelet is used. When implementing a wavelet transform, the wavelet function retains its shape but is displaced along with the signal and compressed and opened during the displacement, thus encapsulating the entire signal. Unlike short-time Fourier transform, this method can create different resolutions for low- and high-frequency ranges.

There are various solutions for implementing DWT, the most common of which is the implementation of Multiresolution Analysis (MRA). In this method, the implementation of discrete wavelet transform is done with the help of a series of consecutive operations that each step of this operation includes signal filtering and downsampling. At each stage of the discrete wavelet process, the signal content is decomposed into two orthogonal subspaces including a low-pass filter (LL) and a high-pass (HH) filter [10], which is then split into four classifications: LH, LL, HH, and HL.

To increase the frequency resolution, this decomposition is rendered consecutively, such that the approximation signal is passed through a pair of high- and low-pass filters and decomposed into two new information and approximate signals. Afterward, the read rapidity was reduced by 50 percent. To give more information, the HL subbands with further efficiency are applied. This process is mathematically formulated as follows:

$$P_{dwt}(s) = \begin{cases} d_{i,j} = \sum f(s) \times H * i(s - 2 \times i \times j), \\ d_{i,j} = \sum f(s) \times L * i(s - 2 \times i \times j), \end{cases} \quad (9)$$

where $d_{i,j}$ signifies the feature of the component in signal $f(s)$, L and H describe the coefficients of low-pass and high-pass filters, respectively, and i and j stand for the wavelet and the translation factor scales, respectively.

4.2. Gray-Level Cooccurrence Matrix. To study the structure of different tissues, Haralick has proposed properties based on the GLCM, which is one of the most successful methods for studying the properties of different tissues [11]. In the gray surface cooccurrence matrix method, it is assumed that the texture image information is determined by a specific matrix. This method is relied on manipulating the gray surfaces of an image. In this method, in addition to examining the gray surface of the desired pixel, the gray surfaces of its neighboring pixels are also examined, and by creating a new matrix of gray surfaces of the pixel neighbors at different angles and distances, the image properties are identified and defined. The coevent matrix is a square matrix and its size is the amount to the gray surfaces' number.

The cooccurrence matrix of an image is defined using radius d and angle θ . Usually, d is selected in the range of 1 and 2. Since each pixel has 8 neighbors at θ equal to 0, 45, 90, 135, 180, 225, 270, and 315 degrees to define the cooccurrence matrix, so the angle selection may be up to 4 adjacent pixels at θ equal to 0, 45, 90, 135, and 180 degrees (horizontal, right diagonal, vertical, and left diameter). In addition to the radius and angle parameters used to define the cosmopolitan matrix, the gray number parameter can also be defined. In this study, to define the cooccurrence matrix, a radial distance of 1 with four zero angles and the number of 256 gray surfaces were used, for which a new matrix was extracted. Subsequently, the information about the utilized characteristics was explained. The first characteristic is *Contrast* that describes the intensity magnitude of the pixels and their neighborhood. The second feature is *Entropy* which defines the image selected interference. The third feature is *Energy* that describes the repetitive pixel pair

quantity. The fourth feature is *Correlation* that defines the spatial feature reliance among the pixels. Finally, *Homogeneity* as the last feature as a local uniformity feature creates multiple/single intervals for accusing the nontextured/textured characteristics. Table 1 indicates five gray-level cooccurrence matrix features extracted from the samples.

5. Convolutional Neural Networks

After feature extraction from the segmented images, they should be classified properly as the final step of diagnosis. In this study, convolutional neural network (CNN) was used for this purpose. CNNs are significant deep learning techniques where several layers are prepared strongly. This technique is very effective and is a usual technique in different applications of computer vision. An outline of convolutional neural network architecture is depicted in Figure 3.

Generally, a CNN is made of three major layers: the convolutional layer, the pooling layer, and the fully connected layer. Various layers do various tasks. Each convolution neural network includes two stages: feedforward and backward for preparation.

In the first step, the features enter the network, and this operation is the point multiplication between the input and the variables of each neuron, and finally, the application of convolution operations in every layer.

The output of the network is then computed. Here, to establish the variables related to network training, network output results are applied to compute the network error rate. To do this, a comparison of the network output to the correct solution (optimal solution) is carried out by an error function and the error rate is computed. In the later phase, by the computed error rate, the postrelease phase begins. The gradient of each variable is computed in this phase based on the chain rule, and all variables are altered by the influence they have on the error created in the network. Following parameters' updating, the feed-forward phase starts. Afterward, repeating a good number of these phases, the network preparing finishes. In this study, CNN is employed for local feature extraction in breast mammogram images. To offer optimum weighting among network connections, the backpropagation technique has been established. As the activation mechanism, a rectified linear unit (ReLU) is used.

With multiplying filter matrices by the images, feature maps are generated. To generate the feature map, the filter moves from left to right and up to down with a specific stride size to extract high-level features (like edges) until it finishes the full width. Here, the Max-pooling process uses the maximum value of the matrix in the feature maps to decrease the output neurons and the cross-entropy loss value based on backpropagation, which is formulated as follows:

$$L = \sum_{j=1}^N \sum_{i=1}^M -d_j^{(i)} \log z_j^{(i)}, \quad (10)$$

$$d_j = \left(0, \dots, 0, \underbrace{1, \dots, 1}_k, 0, \dots, 0 \right),$$

TABLE 1: Five GLCM features extracted from the samples.

Feature name	Mathematical equation
<i>Contrast</i>	$\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (i-j)^2 f(i, j)$
<i>Entropy</i>	$-\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \log_2 f(i, j)$
<i>Energy</i>	$\sqrt{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} f^2(i, j)}$
<i>Correlation</i>	$\frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (i, j) f(i, j) - \mu_i \mu_j}{\sigma_i \sigma_j}$
<i>Homogeneity</i>	$\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \frac{1}{1 + (i-j)^2} f(i, j)$

where d_j signifies the proper output vector and z_j determines the achieved output vector for the m^{th} class. The softmax function is achieved as follows:

$$z_j^{(i)} = \frac{e^{f_j}}{\sum_{i=1}^M e^{f_i}}, \quad (11)$$

where M describes the sample number.

To adjust function followed by keeping higher values, a weighting penalty (ρ) is added that is illustrated in the following equation:

$$L = \sum_{j=1}^N \sum_{i=1}^M -d_j^{(i)} \log z_j^{(i)} + \frac{1}{2} \rho \sum_K \sum_L W_{k,l}^2, \quad (12)$$

where W_k defines the weight of connections and L and K define the overall number of layers and the layer l connections, respectively.

The CNN layouts were usually used based on trials and errors, which yielded inaccurate results. Numerous automated and optimized works have been implemented to address this problem [12]. The use of metaheuristic algorithms is one of the normal approaches. A new, optimized metaheuristic was used in this study to present an effective CNN based on the previously described cases.

6. The Modified Thermal Exchange Optimizer

6.1. The Concept of Newton Law of Cooling. Heat transfer that occurs simultaneously with the movement of a fluid is called convection heat transfer. Depending on the process, heat transfer is divided into two categories: free and forced. In free movement, the energy transferred is due to natural factors such as Archimedes' force. But in forced displacement, external forces such as a pump or fan cause the fluid to move. The heat transfer analysis is complex due to the simultaneous process of thermal conductivity and fluid motion. The higher the fluid velocity, the higher the heat

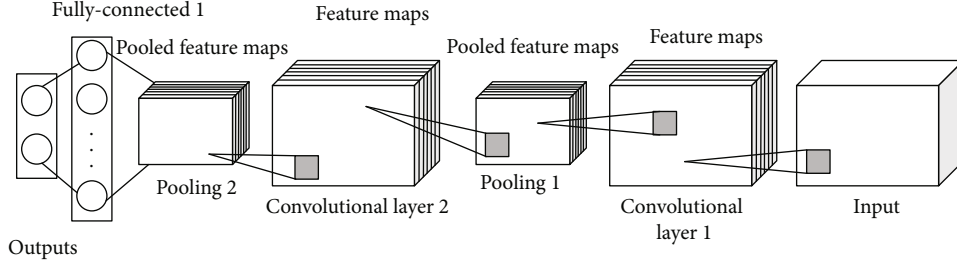


FIGURE 3: An overview of convolutional neural network architecture.

transfer rate. The transfer heat transfer velocity can also be expressed using Newton's law of cooling by the following formula:

$$\dot{Q} = \beta \times A \times (T_s - T_a), \quad (13)$$

where A describes the surface of the body that transfers heat, Q determines the heat, α signifies the coefficient of the heat transfer that relates to numerous cases like surface state, heat transfer mode, and object geometry, and T_b and T_a represent the body and the ambient temperatures.

According to the above equation, heat losing time is $\beta \times A \times (T_a - T) dt$ that defines reserved heat changing once the temperature dT falls, i.e.,

$$V \times \rho \times c \times dT = -\alpha \times A \times (T - T_b) dt, \quad (14)$$

where V represents the volume (m^3), c defines the specific heat ($J/kg/K$), and ρ describes the density (kg/m^3). Therefore,

$$\frac{T - T_b}{T_{eh} - T_b} = \exp\left(\frac{-\beta \times A \times t}{V \times \rho \times c}\right), \quad (15)$$

where T_{eh} describes the early high temperature. By considering the $(\alpha \times A \times t)/(V \times \rho \times c)$, a time-independent value, i.e.,

$$\zeta = \frac{\alpha \times A}{V \times \rho \times c}. \quad (16)$$

That ζ is a constant, the main equation can be reformulated as follows:

$$\frac{T - T_b}{T_{eh} - T_b} = \exp(-\zeta t). \quad (17)$$

Accordingly,

$$T = (T_{eh} - T_b) \times \exp(-\zeta t) + T_b. \quad (18)$$

6.2. Thermal Exchange Optimization Algorithm. After explanations about the concept of Newton's law of cooling, it is time to explain the concept of optimization and the relation of the Newton law of cooling and the optimization [13]. Generally, optimization contains all techniques that are used for finding the best solution for optimization problems. Several methods

of optimization techniques have been introduced for this aim. Classic methods give exact results for the optimization problems, but recently, by increasing the complexity of these problems, the ability to solve the problems with these algorithms is decreasing. Metaheuristics are intelligent algorithms that are used to find the optimal solution and resolve the before mentioned issues [14, 15]. Metaheuristic algorithms are approximation optimizers that have solutions to exit the local optimization and proper for a wide range of problems. Metaheuristic algorithms are an inspiration of various phenomena from the nature, behaviors of animals, breeding, to human societies and use these conceptions to simulate an approach for solving the optimization problem. Several kinds of metaheuristic algorithms have been proposed in recent years [16, 17], for example, biogeography-based optimization [18], elephant herding optimization [19], ant lion optimizer (ALO) algorithm [20], equilibrium optimizer [21], world cup optimizer [22], and Thermal Exchange Optimizer (TEO) [23].

Here, an enhanced design of the TEO algorithm is presented to provide more ability for this algorithm in terms of accuracy and consistency. The TEO algorithm is an inspiration of the temperature performance of the objects and their position which is exchanged between warm and cold portions to indicate the updated positions. In the TEO optimizer, the individual is split into two parts. One group is the candidates that are considered as cooling substances, and the other group is considered as the environment, and then, the reverse process has been made. Figure 4 shows the pairs of transfer objects.

The algorithm begins with a predefined number of random individuals as the initial solutions are as follows:

$$T_j^0 = \underline{T} + \theta \times (\bar{T} - \underline{T}), \quad (19)$$

$$j = 1, 2, \dots, n,$$

where θ denotes a random magnitude in the range $[0, 1]$, T_j^0 signifies the algorithm early population for the i^{th} object, and \underline{T} and \bar{T} stand for the minimum and the maximum limitations.

After achieving the cost value of the generated candidates, T number of the best cost individual positions is saved as *Thermal Memory* (TM) to provide higher efficiency with lower complexity to the algorithm. The *TM* individuals are

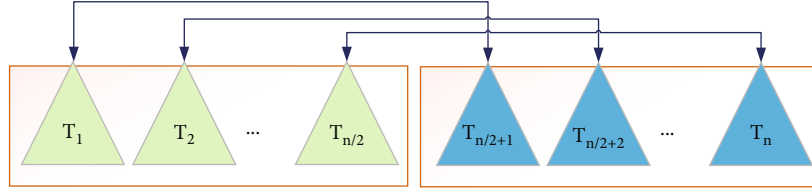


FIGURE 4: The pairs of transfer objects.

then combined to the individual, and the equal number of worst candidates is then taken out.

To provide more understanding, consider Figure 5. T_1 describes the environment object for $T_{(n/2)+1}$ cooling object, and contrariwise. If the object is less than ζ , the temperature exchanges gradually. ζ is formulated as follows:

$$\zeta = \frac{\text{Cos(object)}}{\text{Cos(worst object)}}. \quad (20)$$

Time is another term in the simulation of the optimizer that is related to the number of iteration. This term is obtained by the following equation:

$$t = \frac{\text{iteration}}{\text{Max.iteration}}. \quad (21)$$

To improve the global searching of the algorithm, the environmental temperature changing is considered that is formulated as follows:

$$T_i^e = (1 - (\alpha_1 + \alpha_2 \times (1 - t) \times \delta)) \times T_i^e, \quad (22)$$

where δ describes a random number between 0 and 1, T_i^e represents the preceding temperature of the object modified by T_i^e , and α_1 and α_2 represent the control variables, respectively.

Finally, the new position for the object temperature is achieved by the following:

$$T_i^N = T_i^e + (T_i^{\text{old}} - T_i^e) \exp(-\zeta t). \quad (23)$$

The algorithm also defines whether a component changes in the cooling objects or not. This has been stimulated by a term, called Pr. The Pr contains some individuals that are compared with $R(i)$ which is a random value in the range [0, 1]. If $R(i)$ is less than Pr, one dimension of the i^{th} candidate is randomly chosen and the magnitude is reformulated as given in the following:

$$T_{i,j} = \underline{T}_j + \delta(\bar{T}_j - \underline{T}_j) \exp(-\zeta t), \quad (24)$$

where $T_{i,j}$ describes the variable number j of the individual number i and \underline{T}_j and \bar{T}_j represent the lower and the higher limitations of the parameter number j , respectively. The

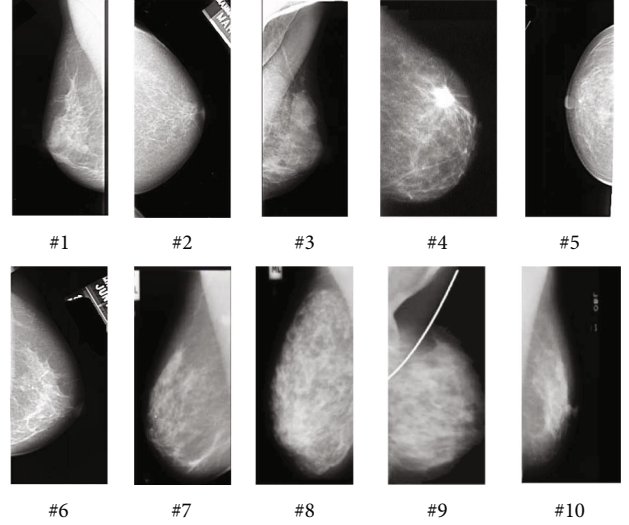


FIGURE 5: Some examples of the MIAS database mammography images.

algorithm is then terminated when the stopping criteria have been reached.

6.3. Advanced Thermal Exchange Optimizer. Although the Thermal Exchange Optimizer has a proper speed in solving the problems, it may be trapped in the local optimum point once solving complex and nonlinear optimization problems. Due to this problem, here, an advanced design of the TEO algorithm is designed and suggested to develop the search power of the original TEO algorithm and to resolve the mentioned issue. The movement of the worst individual in the groups (T_w) is improved in each iteration of the local search in the Advanced Thermal Exchange Optimization (ATEO) algorithm. First, an exchange vector is generated for the worst solution in each iteration:

$$T_i^a = T_b + \gamma \times (T_{r1} - T_{r2}), \quad (25)$$

where T_b signifies the best solution achieved by the current iteration, T_{r1} and T_{r2} represent two dissimilar agents that are randomly chosen from the population in each group, and γ describes the exchange coefficient to determine the differences range between T_{r1} and T_{r2} .

By considering the above condition, the value of the j^{th} parameter of the vector T_i^{New} in the following iteration is achieved as follows:

$$T_i^{\text{New}} = \begin{cases} T_i^a & \text{if rand} < \text{CG}, \\ T_i^N & \text{O.W.}, \end{cases} \quad (26)$$

where CG signifies the general intersection constant between 0 and 1 and rand signifies a random constant in the range of [0, 1]. If the cost value of the new solution has proper value in comparison to the preceding solution, the new individual substitutes the former one; else, it will be kept with no changes.

To give a proper result with the TEO algorithm, the population size should be selected wisely. Indeed, population size is a term to define the number of individuals (candidates) that are randomly generated and tested on the objective function to get the best solution. However, this case is one of the difficult parts of all metaheuristics [24]. Here, a self-adaptive mechanism is used for adjusting this case in each iteration. The main characteristic of the self-adaptive mechanism is that it regulates the population size automatically in each iteration with no user intervention [25]. Based on this mechanism, the initial population size before starting the algorithm main loop is considered as follows:

$$\text{PS} = 10 \times D, \quad (27)$$

where PS signifies the population size and D describes the problem dimensions. So, the new population size is achieved by the following:

$$\text{PS}^{\text{New}} = \text{round}(\text{PS} + \text{rnd} \times \text{PS}), \quad (28)$$

where rnd defines a random magnitude in the range [-0.5, 0.5].

The population size will increase or decrease by up to half the current population size. If the population size obtained for the next iteration increases compared to the population size in the former iteration ($\text{PopSize}_{\text{new}} > \text{PopSize}$), then all members of the present individual are kept unchanged.

Once the population size in the former iteration decreases compared to the population size ($\text{PopSize}_{\text{new}} < \text{PopSize}$), the best members of the present population are kept and the weak members have been discarded. If the size of the population does not change ($\text{PopSize}_{\text{new}} = \text{PopSize}$), so there will be no population changing. Finally, if the new population size reduces from the problem dimensions ($\text{PS}^{\text{New}} < D$), then the population size becomes equal to the problem dimension.

6.4. Algorithm Verification. After designing and introducing the proposed Advanced Thermal Exchange Optimization (ATEO) algorithm, the performance of the method should be analyzed to ensure its ability in use in our purpose.

The present work uses the Single Objective Bound Constrained Numerical Optimization (CEC2020) benchmark standard benchmark in 20 dimensions to analyze the effectiveness of the method; in other words, each faction has 20 decision variables that should be optimally selected. The CEC2020 is known as one of the latterly introduced bench-

TABLE 2: The variable setting of the compared optimizers.

Algorithm	Parameter	Value
BH [27]	a	[0, 1]
	Number of stars	100
MVO [28]	Traveling distance rate	[0.6, 1]
	Wormhole existence prob.	[0.2, 1]
SHO [29]	\vec{M}	[0.5, 1]
	\vec{h}	[5, 0]

mark functions for analyzing optimization problems. The termination criteria, including the maximum number of the calculation of the fitness function and the minimum error value, are set $1e7$ and $1e-8$, respectively. The constraint of the decision variables is in the range [-100, 100], and 35 independent runs have been established for giving a reliable result. The formulation for each equation can be found in [26]. To provide an appropriate analysis for the suggested algorithm, a comparison of its results with some latest algorithms including blackhole (BH) [27], multiverse optimize (MVO) [28], spotted hyena optimize (SHO) [29], and original Thermal Exchange Optimization (TEO) [23] algorithm has been performed. Table 2 indicates the parameter setting of the compared algorithms.

During the simulations, for all of the compared algorithms, the population size is equal to 150. To analyze the algorithms' behavior, the mean magnitude and the (mean) and the standard deviation value (Std) have been extracted from the results. Table 3 discusses the achievements of the algorithms applied to the CEC2020 benchmark sets.

As seen in Table 3, the scores of the analyzed algorithms have been shown. It is clear from the results that in some test functions, the suggested ATEO algorithm escapes the local optimum and found the optimal value. Here, the mean value is employed to consider all of the runs for the algorithm, although the minimum value of the algorithms gives too better or even incomparable results. The minimum value of "Mean" for the proposed ATEO algorithm against the other compared algorithms indicates its higher accuracy to find the minimum value. On the other hand, the minimum value of "Std" for the proposed ATEO algorithm state's better reliability of the proposed algorithm than the comparative methods for the studied CEC2020 benchmark function.

7. Classification

With the advances made in the field of imaging and production of high-resolution digital images, the need for accurate image classification is felt so that one of the most basic parts of image processing is image classification. The important point in image classification is providing a method with high accuracy. Taking into account the above-mentioned reasons, backpropagation is in the form of preparing in the CNN mostly. We also clarified why different methods have been proposed to overcome it because of certain major drawbacks of the backpropagation method. Here, the proposed

TABLE 3: The comparison achievements between the suggested ATEO algorithm and the other compared algorithms on the CEC2020.

		ATEO	TEO [23]	BH [27]	MVO [28]	SHO [29]
F1	Mean	7.38e8	5.83e11	4.22e15	8.37e13	9.07e15
	Std	1.29e8	6.19e10	5.13e13	4.38e11	5.46e11
F2	Mean	5.79e1	9.67e2	4.67e3	1.76e4	4.46e6
	Std	4.31e1	2.84e2	3.82e2	6.37e2	2.08e5
F3	Mean	2.08e1	6.92e2	9.37e2	5.17e2	4.83e5
	Std	1.46e0	3.27e0	4.28e1	8.09e1	6.17e4
F4	Mean	0.00	6.15e-10	5.80e-6	4.96e-7	7.67e-6
	Std	0.00	3.48e-11	9.37e-7	4.18e-8	4.08e-8
F5	Mean	1.76e2	4.53e2	6.37e4	6.55e3	9.86e7
	Std	3.82e1	1.27e2	5.19e3	2.41e2	8.19e3
F6	Mean	3.29e-1	6.12e-1	8.09e0	7.18e0	2.96e1
	Std	4.13e-1	2.73e-1	3.46e-1	4.82e-2	4.63e0
F7	Mean	3.18e0	4.16e0	8.09e1	5.33e2	4.29e3
	Std	1.24e0	1.08e0	6.17e1	6.81e1	2.82e2
F8	Mean	7.19e1	8.35e2	2.19e4	5.24e5	2.56e4
	Std	2.76e0	4.37e0	3.77e1	4.65e1	4.07e3
F9	Mean	1.96e2	3.17e2	6.51e3	2.85e3	5.11e4
	Std	1.07e1	2.03e1	8.09e2	6.19e1	6.97e1
F10	Mean	5.76e2	9.83e2	9.23e3	5.37e3	5.17e3
	Std	4.27e-1	5.94e-1	2.60e0	1.93e1	6.93e1

TABLE 4: The feature extraction for training data.

#	H	CR	E	CN	ER
1	0.816	0.173	0.794	0.257	0.298
2	0.757	0.038	0.996	0.047	0.264
3	0.869	0.046	0.957	0.032	0.317
4	0.806	0.042	0.896	0.031	0.376
5	0.794	0.041	0.987	0.135	0.395
6	0.843	0.010	0.917	0.009	0.219
7	0.585	0.057	0.967	0.037	0.293
8	0.810	0.007	0.968	0.011	0.417
9	0.594	0.068	0.979	0.028	0.294
10	0.704	0.041	0.968	0.046	0.407

TABLE 5: The feature extraction for the testing data.

#	H	CR	E	CN	ER
1	0.794	0.072	0.794	0.047	0.272
2	0.758	0.053	0.856	0.012	0.311
3	0.786	0.032	0.851	0.046	0.347
4	0.865	0.029	0.783	0.019	0.215
5	0.749	0.028	0.764	0.018	0.337
6	0.708	0.029	0.886	0.029	0.318
7	0.819	0.017	0.895	0.053	0.319
8	0.693	0.022	0.851	0.031	0.420
9	0.649	0.034	0.817	0.050	0.433
10	0.684	0.069	0.963	0.079	0.351

Advanced Thermal Exchange Optimizer (ATEO) is developed and used to reduce the proper and output magnitude by the selection of suitable network weights replacing backpropagation in CNN for mean square error (MSE). The MSE can be mathematically described by the following equation:

$$\text{MSE} = \frac{1}{T} \sum_{j=1}^N \sum_{i=1}^M (y_j^i - d_j^i)^2, \quad (29)$$

where M and N represent the value of the output layers

and the data, respectively, and y_j^i and d_j^i define the obtained and the proper magnitudes for j^{th} unit in the output layer of the CNN in time t , respectively.

8. Simulation Results

This study presents an efficient and automated method for brain tumor detection by combining deep neural networks and metaheuristics. The technique involves image preprocessing, image segmentation, extraction of features, and then classification. Based on digital mammogram images, the process is validated.

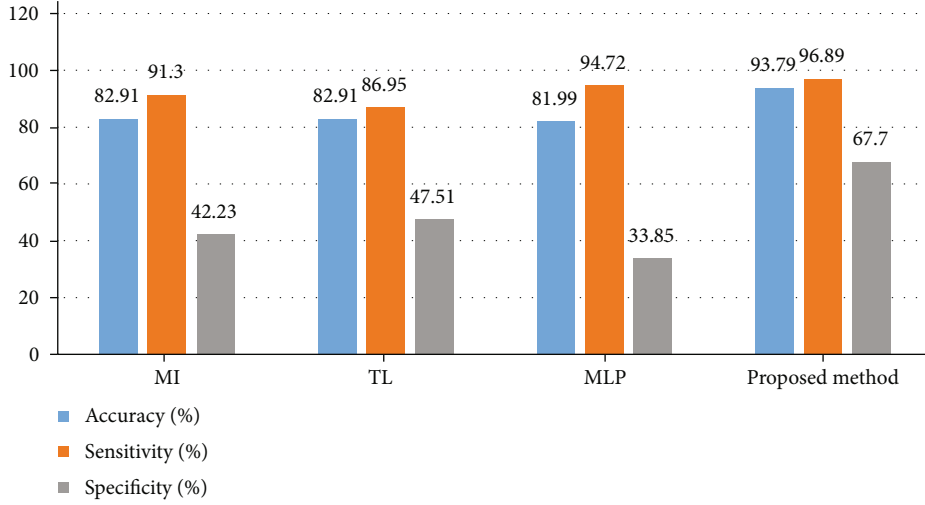


FIGURE 6: The comparison results between the suggested pipeline ATEO-based methodology and the mentioned methods applied to the MIAS database.

8.1. Dataset Description. To verify the accuracy and the ability of the suggested method, it is performed to a standard mammographic benchmark database, known Image xAnalysis Society Digital Mammogram Database (MIAS) [30]. The MIAS database is compiled by UK researchers to support researchers involved in working on mammogram images. The database includes 322 number of 1024×1024 digital mammography images that are taken from the UK National Breast Screening Program. The MIAS database also includes correct labels which are obtained with the help of experts. This database has been gotten available by the Pilot European Image Processing Archive (PEIPA) at the University of Essex. Figure 5 shows some examples of the MIAS database mammography images.

8.2. Simulations. By performing the discrete wavelet transform to the image and using decomposition of LL and HL characteristics from it, the GLCM characteristics are achieved by the wavelet decomposition's extracted levels. Then, the features are combined with the optimized CNN-based classifier that is arranged for the final detections. As mentioned before, five features including homogeneity (H), correlation (Cr), contrast (CN), energy (E), and entropy (ER) are employed to the LL and HL subband levels on the image. Table 4 illustrates the feature extraction for preparing data.

Also, Table 5 illustrates the feature extraction for the testing data.

For more clarification of the proposed automatic system, it is validated by three measurement indicators, precision, sensitivity, and specificity that are formulated in the following:

$$\begin{aligned}
 \text{Accuracy}(\%) &= \frac{TP + TN}{TP + FP + FN + TN}, \\
 \text{Sensitivity}(\%) &= \frac{TP}{TP + FN}, \\
 \text{Specificity}(\%) &= \frac{TN}{FP + TN},
 \end{aligned} \tag{30}$$

where TN, TP, FN, and FP represent Truly Negative, Truly Positive, False Negative, and False Positive, respectively.

To proper validation of the suggested technique, it was compared with three latest techniques including Multilayer Perceptron (MLP) [31], Multiple Instance (MI) [32], and Transfer Learning (TL) [33]. Figure 6 shows the comparison results between the suggested pipeline Advanced Thermal Exchange Optimization algorithm-based methodology and the mentioned methods applied to the MIAS database.

Based on Figure 6, the suggested ATEO-based methodology with a 93.79% accuracy rate has the highest precision, and the method of MI, TL, and MLP with 82.91%, 82.91%, and 81.99% is placed in the later ranks. Furthermore, the specificity of the suggested method with 67.7% provides the best achievements than the other compared methods. Finally, the total achievements display optimal results for the suggested technique to automatic breast cancer diagnosis.

9. Conclusions

Breast cancer has been a cause of death in women in the last decade; the rate of breast cancer is increasing worldwide. This cancer is common cancer detected in women, and death from breast cancer is more common in women between the ages of 15 and 54. In recent years, much research was performed on mammographic images to be able to diagnose cancerous tumors ignoring the intervention of a person by image processing methods and computer programming. The present study presented a computer-aided diagnosis system for automatic detections of breast cancers. The mammogram images were first preprocessed based on image contrast enhancement and noise reduction to improve and prepare the image for the next steps. Afterward, a method based on color space was used for image segmentation that is followed by mathematical morphology. To achieve the main characteristics of the mammogram images, a combined gray-level cooccurrence matrix (GLCM) and discrete wavelet transform (DWT) was applied to the processed images. Finally, a new optimized version of convolutional

neural network (CNN) and a new improved metaheuristic, called Advanced Thermal Exchange Optimization algorithm, was applied for features' categorization. Simulation achievements of the suggested technique were finally compared with three other techniques including Multilayer Perceptron (MLP), Multiple Instances (MI), and Transfer Learning (TL) applied on the MIAS mammogram database to show its superiority.

Data Availability

The database for analysis is based on MIAS (the mammographic image analysis society digital mammogram database) which can be obtained as follows: <http://peipa.essex.ac.uk/info/mias.html>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] N. Razmjoooy, F. R. Sheykhahmad, and N. Ghadimi, "A hybrid neural network-world cup optimization algorithm for melanoma detection," *Open Medicine*, vol. 13, no. 1, pp. 9–16, 2018.
- [2] 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.e10, 2018.
- [3] N. Razmjoooy, B. Somayeh Mousavi, F. Soleymani, and M. Hosseini Khotbesara, "A computer-aided diagnosis system for malignant melanomas," *Neural Computing and Applications*, vol. 23, no. 7–8, pp. 2059–2071, 2013.
- [4] Q. Liu, Z. Liu, S. Yong, K. Jia, and N. Razmjoooy, "Computer-aided breast cancer diagnosis based on image segmentation and interval analysis," *Automatika*, vol. 61, no. 3, pp. 496–506, 2020.
- [5] M. Toğaçar, K. B. Özkurt, B. Ergen, and Z. Cömert, "Breast-Net: a novel convolutional neural network model through histopathological images for the diagnosis of breast cancer," *Physica A: Statistical Mechanics and its Applications*, vol. 545, article 123592, 2020.
- [6] E. D. Carvalho, A. O. C. Filho, R. R. V. Silva et al., "Breast cancer diagnosis from histopathological images using textural features and CBIR," *Artificial Intelligence in Medicine*, vol. 105, article 101845, 2020.
- [7] V. Rajinikanth and S. C. Satapathy, "Segmentation of ischemic stroke lesion in brain MRI based on social group optimization and fuzzy-Tsallis entropy," *Arabian Journal for Science and Engineering*, vol. 43, no. 8, pp. 4365–4378, 2018.
- [8] N. Dey, V. Rajinikanth, A. Ashour, and J. M. Tavares, "Social group optimization supported segmentation and evaluation of skin melanoma images," *Symmetry*, vol. 10, no. 2, p. 51, 2018.
- [9] N. Razmjoooy, B. S. Mousavi, B. Sadeghi, and M. Khalilpour, "Image thresholding optimization based on imperialist competitive algorithm," in *3rd Iranian Conference on Electrical and Electronics Engineering (ICEEE2011)*, Iran: Islamic Azad University of Gonabad, 2011.
- [10] M. Tarafdar Hagh, H. Ebrahimian, and N. Ghadimi, "Hybrid intelligent water drop bundled wavelet neural network to solve the islanding detection by inverter-based DG," *Frontiers in Energy*, vol. 9, no. 1, pp. 75–90, 2015.
- [11] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.
- [12] R. Navid, V. V. Estrela, and H. J. Loschi, "A study on metaheuristic-based neural networks for image segmentation purposes," in *Data Science*, pp. 25–49, CRC Press, 2019.
- [13] N. Ghadimi, "A method for placement of distributed generation (DG) units using particle swarm optimization," *International Journal of Physical Sciences*, vol. 8, no. 27, pp. 1417–1423, 2013.
- [14] N. Ghadimi, A. Afkousi-Paqaleh, and A. Emamhosseini, "A PSO-based fuzzy long-term multi-objective optimization approach for placement and parameter setting of UPFC," *Arabian Journal for Science and Engineering*, vol. 39, no. 4, pp. 2953–2963, 2014.
- [15] M. Saeedi, M. Moradi, M. Hosseini, A. Emamifar, and N. Ghadimi, "Robust optimization based optimal chiller loading under cooling demand uncertainty," *Applied Thermal Engineering*, vol. 148, pp. 1081–1091, 2019.
- [16] Z. Yuan, W. Wang, H. Wang, and N. Ghadimi, "Probabilistic decomposition-based security constrained transmission expansion planning incorporating distributed series reactor," *IET Generation, Transmission & Distribution*, vol. 14, no. 17, pp. 3478–3487, 2020.
- [17] Q. Meng, T. Liu, C. Su, H. Niu, Z. Hou, and N. Ghadimi, "A single-phase transformer-less grid-tied inverter based on switched capacitor for PV application," *Journal of Control, Automation and Electrical Systems*, vol. 31, no. 1, pp. 257–270, 2020.
- [18] D. Simon, "Biogeography-based optimization," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 702–713, 2008.
- [19] G.-G. Wang, S. Deb, and L. D. S. Coelho, "Elephant herding optimization," in *2015 3rd international symposium on computational and business intelligence (ISCBI)*, Bali, Indonesia, 2015.
- [20] M. Mani, O. Bozorg-Haddad, and X. Chu, "Ant lion optimizer (ALO) algorithm," in *Advanced Optimization by Nature-Inspired Algorithms*, O. Bozorg-Haddad, Ed., vol. 720 of Studies in Computational Intelligence, pp. 105–116, Springer, Singapore, 2018.
- [21] A. Faramarzi, M. Heidarinejad, B. Stephens, and S. Mirjalili, "Equilibrium optimizer: a novel optimization algorithm," *Knowledge-Based Systems*, vol. 191, article 105190, 2020.
- [22] R. Navid, V. V. Estrela, R. Padilha, and A. C. B. Monteiro, "World Cup Optimization Algorithm: An Application for Optimal Control of Pitch Angle in Hybrid Renewable PV/Wind Energy System," in *Metaheuristics and Optimization in Computer and Electrical Engineering*, pp. 25–47, Springer, Cham, 2021.
- [23] A. Kaveh and A. Dardas, "A novel meta-heuristic optimization algorithm: thermal exchange optimization," *Advances in Engineering Software*, vol. 110, pp. 69–84, 2017.
- [24] Y. Guo, X. Dai, K. Jermisittiparsert, and N. Razmjoooy, "An optimal configuration for a battery and PEM fuel cell-based hybrid energy system using developed krill herd optimization algorithm for locomotive application," *Energy Reports*, vol. 6, pp. 885–894, 2020.
- [25] R. Rao and K. More, "Conception optimale et analyse d'une tour de refroidissement a convection mecanique grace a un

- algorithm Jaya ameliorate,” *International Journal of Refrigeration*, vol. 82, pp. 312–324, 2017.
- [26] J. J. Liang, B. Y. Qu, and P. N. Suganthan, *Problem Definitions and Evaluation Criteria for the CEC 2014 Special Session and Competition on Single Objective Real-Parameter Numerical Optimization*, Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou China and Technical Report, Nanyang Technological University, Singapore, 2013.
- [27] A. Hatamlou, “Black hole: a new heuristic optimization approach for data clustering,” *Information Sciences*, vol. 222, pp. 175–184, 2013.
- [28] 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.
- [29] G. Dhiman and V. Kumar, “Spotted hyena optimizer: a novel bio-inspired based metaheuristic technique for engineering applications,” *Advances in Engineering Software*, vol. 114, pp. 48–70, 2017.
- [30] J. Suckling, “The mammographic image analysis society digital mammogram database,” *Digital Mammo*, pp. 375–386, 1994.
- [31] M. Mehdy, P. Y. Ng, E. F. Shair, N. I. M. Saleh, and C. Gomes, “Artificial neural networks in image processing for early detection of breast cancer,” *Computational and Mathematical Methods in Medicine*, vol. 2017, Article ID 2610628, 15 pages, 2017.
- [32] P. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, and P. Honeine, “Multiple instance learning for histopathological breast cancer image classification,” *Expert Systems with Applications*, vol. 117, pp. 103–111, 2019.
- [33] E. Deniz, A. Şengür, Z. Kadiroğlu, Y. Guo, V. Bajaj, and Ü. Budak, “Transfer learning based histopathologic image classification for breast cancer detection,” *Health Information Science and Systems*, vol. 6, no. 1, pp. 1–7, 2018.