



An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators algorithm

Essam H. Houssein¹ · Marwa M. Emam¹ · Abdelmgeid A. Ali¹

Received: 26 August 2021 / Accepted: 14 May 2022 / Published online: 8 June 2022
© The Author(s) 2022

Abstract

Breast cancer is the second leading cause of death in women; therefore, effective early detection of this cancer can reduce its mortality rate. Breast cancer detection and classification in the early phases of development may allow for optimal therapy. Convolutional neural networks (CNNs) have enhanced tumor detection and classification efficiency in medical imaging compared to traditional approaches. This paper proposes a novel classification model for breast cancer diagnosis based on a hybridized CNN and an improved optimization algorithm, along with transfer learning, to help radiologists detect abnormalities efficiently. The marine predators algorithm (MPA) is the optimization algorithm we used, and we improve it using the opposition-based learning strategy to cope with the implied weaknesses of the original MPA. The improved marine predators algorithm (IMPA) is used to find the best values for the hyperparameters of the CNN architecture. The proposed method uses a pretrained CNN model called ResNet50 (residual network). This model is hybridized with the IMPA algorithm, resulting in an architecture called IMPA-ResNet50. Our evaluation is performed on two mammographic datasets, the mammographic image analysis society (MIAS) and curated breast imaging subset of DDSM (CBIS-DDSM) datasets. The proposed model was compared with other state-of-the-art approaches. The obtained results showed that the proposed model outperforms the compared state-of-the-art approaches, which are beneficial to classification performance, achieving 98.32% accuracy, 98.56% sensitivity, and 98.68% specificity on the CBIS-DDSM dataset and 98.88% accuracy, 97.61% sensitivity, and 98.40% specificity on the MIAS dataset. To evaluate the performance of IMPA in finding the optimal values for the hyperparameters of ResNet50 architecture, it compared to four other optimization algorithms including gravitational search algorithm (GSA), Harris hawks optimization (HHO), whale optimization algorithm (WOA), and the original MPA algorithm. The counterparts algorithms are also hybrid with the ResNet50 architecture produce models named GSA-ResNet50, HHO-ResNet50, WOA-ResNet50, and MPA-ResNet50, respectively. The results indicated that the proposed IMPA-ResNet50 is achieved a better performance than other counterparts.

Keywords Breast cancer classification · Deep learning · Transfer learning · Convolutional neural network · Marine predators algorithm · Opposition-based learning · Hyperparameters optimization

1 Introduction

Breast cancer is a familiar frequent malignancy in females and the second most common leading cause of death in women. The global occurrence of breast cancer has increased over time, and more cases have been reported every year. In comparison to other malignancies, it is more common in women. If this disease is not detected early, it might lead to death [1]. Early diagnosis improves the chances of successful therapy and survival, but its diagnosis is time-consuming and frequently results in an

✉ Essam H. Houssein
essam.halim@mu.edu.eg

Marwa M. Emam
marwa.khalef@mu.edu.eg

Abdelmgeid A. Ali
a.ali@mu.edu.eg

¹ Faculty of Computers and Information, Minia University, Minia, Egypt

agreement between pathologists. Computer-aided diagnosis (CAD) systems can improve diagnostic accuracy. Breast cancer can be classified as benign (non-hazardous) or malignant (threatening). Early detection of breast lesions and the distinction of malignant from benign lesions is critical for breast cancer prognosis [2]. Breast cancer affects approximately a million women globally each year, accounting for more than 25% of all female cancer occurrences. There is a tremendous need for early-stage development of new breast cancer methods with this exponential expansion. This motivates researchers to develop innovative methods for obtaining rapid and accurate diagnoses, ultimately extending patients' lives [1]. Reviewing preliminary diagnostic information and gathering relevant data from previous information is key to detecting this disease early and reliably. Medical imaging and deep learning (DL) techniques will aid this procedure. Medical imaging performs a significant part in clinical illness diagnosis, therapy evaluation, and the detection of anomalies in many body components, including eye [3], lungs [3], brain [4], breast [5], and stomach [6]. Medical image research aims to classify an organ's location, size, and characteristics in question, which is regarded as a viable way to obtain usable information from vast amounts of data. The most effective process to identify breast cancer is through medical imaging, including mammography images, histopathology images, magnetic resonance imaging (MRI), ultrasound, and thermography images [7].

Thermography images, also known as thermal imaging, have shown enormous promise in the early detection of breast cancer over the previous decade. Thermography images can help with other types of diagnosis by providing information about physiological changes [8].

Ultrasound imaging is another method for breast cancer detection that young women mostly use it. Noise levels might cause the process to fail when it attempts to detect microcalcifications and deeper breast tissue [9].

In addition to ultrasound and thermography screening, MRI is another approach for detecting cancer cells early. MRI uses magnetic instead of X-rays to produce extremely accurate three-dimensional (3D) transverse imaging [7].

Another imaging modality is called mammography images. Mammography is a very well and generally used approach for breast cancer screening, and it is the only image type that has been shown to decrease breast cancer mortality significantly [10]. It is an x-ray test image regarded as a reliable and accurate method for detecting breast cancer. [11]. In this paper, we used mammography images for breast cancer classification.

CAD systems have been used to aid clinicians in interpreting medical imaging to improve disease diagnosis. The critical role of a CAD system is feature extraction. Traditional feature extraction techniques have disadvantages

because they lack flexibility [10]. Recently, DL approaches have been presented for breast cancer diagnosis.

DL is a category of machine learning and artificial intelligence that focuses on a complicated structure of image features due to its capacity to learn autonomously. DL approaches use various recently developed models to improve feature extraction from data. These models have been used in various medical fields [7, 12]. DL is composed of multilayer neural networks (NNs) that use raw input images to generate a hierarchical feature structure. Stack autoencoders, deep-Boltzmann machines, and convolution NNs (CNNs) are examples of common DL algorithms [13].

CNNs have had significant success in biomedical imaging, such as mitosis cell detection from microscopic images, tumor detection, segmentation of neural laminae, skin disease, immune cell detection and classification, mass detection, COVID-19 prediction [14] and classification in mammograms [15]. CNNs are a popular technique for object detection and image classification that involves layer-wise automatic feature extraction. The preparation of CNNs for classification objectives depends on the knowledge of hyperparameter tuning. Hyperparameters for each layer are different. Few studies, such as [16], have recognized the importance of the hyperparameters in achieving high performance with CNN architectures and the need to consider them as an optimization problem. Because a CNN model's performance is determined by these hyperparameters [17], they need to be fine-tuned to achieve great results. The selection of hyperparameter values is frequently based on a mix of human expertise, trial and error, or a grid search method [18]. Training can take more days because of the computationally expensive core of CNN designs. Since the number of combinations grows exponentially with the increase in hyperparameters, the grid search method is usually not suited for CNN models. Tuning hyperparameters is a time-consuming effort for researchers; the number of layers of CNNs is increasing daily to cope with vast and complex datasets. It is not suitable to optimize hyperparameters manually at a reasonable cost. To improve them, different researchers have adopted different techniques. Some of them accepted their outcome, whereas others did not [17].

Thus, automatic hyperparameter optimization for CNN models is critical [19]. Meta-heuristic algorithms have significantly influenced hyperparameter optimization in several fields. They have been developed to address different real-world problems and have gained enormous interest in classification problems. Because meta-heuristic algorithms have great performance and are straightforward to implement, researchers have widely proved their experience to handle many types of challenging optimization issues in engineering, communications, industry, and social

sciences [20]. In addition, they have been employed in biological information [21], chemical information [22], feature selection [23], task scheduling in cloud computing [24], image segmentation [25], global optimization [26], and as well as cost-effective emission dispatch [27]. There are various meta-heuristic algorithms, of which the marine predators algorithm (MPA) [28]. MPA is one of the most recent algorithms, released in 2020 by Faramarzi et al., [28], showing more quality results than several classical and the latest counterparts on various mathematical and engineering benchmark problems. Faramarzi et al. applied twenty-nine test functions to evaluate the MPA performance, and it showed high performance in different optimization problems. MPA has many advantages, including requiring the fewest amount of adjustable parameters, being simple to implement with powerful search capability, and flexible in altering the basic MPA version [29].

However, all meta-heuristic algorithms should strike a balance between exploration and exploitation; other solutions can be stuck in optimal solutions or fail to converge [30, 31]. Premature convergence is a problem for the MPA. As it splits the optimization iterations into three parts, the first part is only for exploration. The second part serves as a means of transitioning from exploration to exploitation. The third part is specified for the exploitation phase; this could affect the search process by causing the population to get stuck in local optimal. Accordingly, like other meta-heuristic algorithms, the MPA has been improved. The no-free-lunch theorem states that no single method can solve all types of optimization problems. As a result of hybridization, various features can be combined into a single algorithm to address multiple challenges.

Thus, hybridization of many techniques from various scientific domains is required. Hybridization merges the benefits of various algorithms to create a more powerful, high-performing version with a promise of higher accuracy and performance. Notwithstanding MPA's success in search processes, it might be improved in various areas to demonstrate its usefulness on more complex optimization projects. To enhance the MPA, we hybrid it with the opposition-based learning (OBL) strategy [32] to produce solutions from probable regions to explore the search space more thoroughly. OBL is one of the most efficient approaches to improve meta-heuristic algorithms [33]. It is merged with meta-heuristic algorithms in multiple ways to improve explorative searchability. After that, the proposed Improved Marine Predators Algorithm (IMPA) is used to optimize the CNN's hyperparameters. As a result, automatic hyperparameter optimization for CNN architectures is critical. This paper proposes an improved meta-heuristic algorithm called IMPA using the OBL strategy to optimize the hyperparameters of the CNN architecture used for breast cancer classification. The proposed method is

IMPA-ResNet50, which depends on a pretrained CNN model named ResNet50 to diagnose breast cancer from two mammographic datasets, the mammographic image analysis society (MIAS) and curated breast imaging subset of DDSM (CBIS-DDSM) datasets, using transfer learning (TL). The main contributions of this paper are as follows:

- This paper proposes a diagnostic model for breast cancer.
- The proposed model is IMPA-ResNet50, which is based on TL and uses a pretrained CNN model called ResNet50, along with an enhanced optimization algorithm called IMPA.
- OBL has been used to improve the performance of the MPA.
- IMPA, an improved version of the original MPA based on OBL, is proposed to optimize the hyperparameters of CNN architecture.

The remainder of this paper is structured as follows: Sect. 2 presents some literature reviews. Section 3 explains the MPA, the OBL strategy, CNNs, and TL. Section 4 presents the improved version of the MPA (the IMPA method). Section 5 introduces details of the proposed model. The experimental results and performance analysis including the limitation of the proposed model are explained in Sect. 6. Finally, Sect. 7 concludes this paper and presents future directions.

2 Literature review

This section provides a summary of previous work on breast cancer diagnosis. In [34], the authors presented a review of some works on breast cancer classification, and from this review, they conclude that CNNs achieve higher accuracy than multilayer perceptron (MLP) NNs. The authors in [15] proposed a DL technique based on TL. They used three pretrained models, namely GoogLeNet, VGGNet, and ResNet, to classify malignant and benign cells. They evaluate their approach on cytology images to see how well it works. In addition, in [35], the authors presented a deep-CNN model that incorporated TL to avoid overfitting occurring when dealing with small datasets. They evaluated the presented model's performance using four datasets: DDSM, INbreast, BCDR, and MIAS datasets. The DDSM dataset showed 97.35% accuracy and 0.98 area under the curve (AUC). The INbreast dataset yielded 95.5% accuracy and 0.97 AUC, whereas the BCDR database yielded 96.67% accuracy and 0.96 AUC.

The researchers in [36] presented an approach called BDR-CNN-GCN that combines a graph-convolutional network (GCN) with a CNN. A basic eight-layer CNN was used that was integrated with batch normalization and

dropout layer. The final BDR-CNN-GCN model was developed by combining this model with a two-layer GCN. The performance of this model was tested on the MIAS dataset, achieving 96.10% accuracy. Furthermore, in [37], the researchers presented a CNN Inception-v3 model that was trained on 316 images, yielding 0.946 AUC, 0.88 sensitivity, and 0.87 specificity. In addition, in [38], the authors proposed a classification method using CNN and TL. The purpose of the paper was to assess how eight fine-tuned pretrained models performed. In [39], the authors proposed a hybrid classification model: Alexnet, Mobilenet, and Resnet50. They achieved 95.6% accuracy from these hybridized models. In [40], the researchers used four different CNN architectures, namely InceptionV3, VGG16, ResNet50, and VGG19, including 5000 images, for model training (benign: 2500 and malignant: 2500). In addition, the prediction models were tested by 1007 images (benign: 788 and malignant: 219).

In the same context, in [41], the researchers proposed a TL model for classifying histopathological breast cancer images. The authors used the ResNet-18 model as a backbone model with a block-wise fine-tuning method. The performance of the model was improved using data augmentation techniques and global contrast normalization. In addition, in [42], the authors proposed a classification method for thermal images, which combines thermal images of different views using a CNN model. The method achieved 97% accuracy and 0.99 AUC, with a 100% specificity and 83% sensitivity. In [2], the authors presented a DL framework (DenseNet) that extracts image features and feeds them into a fully connected (FC) layer to classify cancerous and benign cells. The efficiency of the technique was evaluated through hyperparameter adjustment. In [43], the authors proposed a deep classification algorithm for mammography imaging named CNNI-BCC (CNN improvement for breast cancer). The CNNI-BCC model classifies breast images into malignant, benign, and healthy. They achieved 90.50% accuracy and 90.71% specificity. In [44], the authors demonstrate that TL provides improved performance. They employed Inception-v4 [45] pretrained with the ImageNet and DDSM datasets.

Furthermore, in [46], the authors used AlexNet, along with the support vector machine (SVM), to improve classification accuracy and using data augmentation techniques to increase input images. Performance is evaluated on two datasets: DDSM and CBIS-DDSM. The accuracy of the method is 71.01%, and when using SVM, the accuracy becomes 87.2%. In addition, in another work [47], a deep-CNN was proposed in which an MLP was employed in the FC layer to classify mammography images into benign, malignant, and normal. The authors employed a bilateral filter with vector grid computing to maintain edge information in the preprocessing step. They also use

hyperparameter tuning to evaluate performance. The results demonstrate that hyperparameter tuning of the final layers produces 96.23% overall accuracy and 97.46% average accuracy. In addition, in [48] used the CBIS-DDSM dataset to develop an automatic mammogram classification method based on TL and data augmentation. ResNet was fine-tuned to produce good results, and achieved accuracy was 93.15%.

To be specific, despite the encouraging results achieved by the CNN architectures in detecting breast cancer, the large number of hyperparameters is a barrier to attaining improved results. Thus, the use of hyperparameter optimization for CNN architecture is essential to enhance the performance of CNNs. In this paper, an optimized CNN model based on an improved MPA (IMPA) algorithm was proposed for breast cancer classification, which may aid health professionals in breast cancer diagnosis.

3 Preliminaries

This section explained the concept, the mathematical representation of the marine predators algorithm (MPA), along with the opposition-based learning (OBL) strategy, the convolutional neural networks (CNNs) architecture, and transfer learning (TL).

3.1 Marine predators algorithm

The motivation of MPA appears from the general foraging operation in ocean predators as well as predator-prey communications. In this scenario, a predator optimizes encounter rates to increase the chances of surviving in natural surroundings. MPA uses Lévy flight and Brownian motion to do a search using two simple random walk methods. The Lévy flight is usually performed in meta-heuristic algorithms and is most effective to avoid solution stagnation by executing a constructive search in local area [49]. Also, Brownian motion is a well-known global search instrument. The designers of MPA merged the search effectiveness of Lévy and Brownian motion to increase the trade-off scale through exploration and exploitation [50].

MPA initializes the search by randomly determining N_n search agents using Eq. (1):

$$x_i^0 = lb_i + \dot{q} \times (ub_i - lb_i), i \in \{1, 2, \dots, N_n\} \quad (1)$$

where \dot{q} is a random number in $[0,1]$, lb_i and ub_i are lower and upper bounds. During initialization and the basic population matrix, another $N_n \times \text{Dim}$ matrix is generated, including search agents with best fitness values, N_n and Dim indicate population size and dimensions. MPA refers to it as Elite:

$$\text{Elite} = \begin{bmatrix} X'_{1,1} & X'_{1,2} & \dots & X'_{1,\text{Dim}} \\ X'_{2,1} & X'_{2,2} & \dots & X'_{2,\text{Dim}} \\ \vdots & \vdots & \vdots & \vdots \\ X'_{N_n,1} & X'_{N_n,2} & \dots & X'_{N_n,\text{Dim}} \end{bmatrix}_{N_n \times \text{Dim}} \tag{2}$$

where X^I denotes a vector with the highest fitness.

Prey is similar to Elite. Predators utilize it to update their positions. The initialization generates the initial prey, from which the fittest one creates the Elite in a single term. The Prey is depicted in the following way:

$$\text{Prey} = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,\text{Dim}} \\ X_{2,1} & X_{2,2} & \dots & X_{2,\text{Dim}} \\ \vdots & \vdots & \vdots & \vdots \\ X_{N_n,1} & X_{N_n,2} & \dots & X_{N_n,\text{Dim}} \end{bmatrix}_{N_n \times \text{Dim}} \tag{3}$$

$X_{i,j}$ indicates the j th dimension of the i th prey. These two matrices are crucial in the optimization process.

Following initialization, the primary iterative search method begins, which is divided into 3 stages that simulate various situations among predator and the prey while devising different search techniques. The three stages are based on iterations $it \in \{1, 2, 3, \dots, it_{\max}\}$ where it_{\max} denotes the maximum iterations. MPA refreshes potential solutions during these stages.

Stage 1: High velocity ($it < \frac{it_{\max}}{3}$) This simulates the situation in which the prey is outrunning the predator. This approach reinforces exploration and consumes more of the prior iterations. The mathematical representation of this rule is carried out by Eq. (4):

$$\begin{aligned} \overrightarrow{\text{stepsize}}_j &= \mathbf{R}_r \times (\overrightarrow{\text{Elite}}_j - \mathbf{R}_r \times \overrightarrow{\text{prey}}_j), j \in \{1, 2, \dots, N_n\} \\ \overrightarrow{\text{Prey}}_j &= \overrightarrow{\text{Prey}}_j + P \cdot \mathbf{R} \times \overrightarrow{\text{stepsize}}_j \end{aligned} \tag{4}$$

where R_r is a random number between [0,1] relies on normal distribution identify the Brownian motion. P is a constant equal to 0.5, and R is a vector of uniform random numbers in [0, 1]. The speed of predators and prey is great during this stage, which aids in the exploration of far-flung areas of the search space.

Stage 2: Unit velocity ($\frac{1}{3}it_{\max} < it < \frac{2}{3}it_{\max}$) Prey and predator move at equal speed. The transmission step is transitioned from exploration to exploitation in this phase. As a result, the population is separated into prey as exploration utilizing Lévy motion and predator exploiting Brownian motion. The first half of population identified by Eq. (5) and the second half using Eq. (6).

$$\begin{aligned} \overrightarrow{\text{stepsize}}_j &= \mathbf{R}_{\text{levy}} \times (\overrightarrow{\text{Elite}}_j - \mathbf{R}_{\text{levy}} \times \overrightarrow{\text{prey}}_j), \\ j &\in \{1, 2, \dots, N_n/2\} \end{aligned} \tag{5}$$

$$\overrightarrow{\text{prey}}_j = \overrightarrow{\text{prey}}_j + P \cdot \mathbf{R} \times \overrightarrow{\text{stepsize}}_j$$

where \mathbf{R}_{levy} is a random number relies on the Lévy. The multiplication of \mathbf{R}_{levy} , $\overrightarrow{\text{prey}}_j$ imitates the predator motion in Lévy, whereas adding the step size to prey position affects prey movement.

$$\overrightarrow{\text{stepsize}}_j = \mathbf{R}_r \times (\mathbf{R}_r \times \overrightarrow{\text{Elite}}_j - \overrightarrow{\text{Prey}}_j), j \in \{N_n/2, \dots, N_n\}$$

$$\overrightarrow{\text{prey}}_j = \overrightarrow{\text{Elite}}_j + P * \hat{\text{CF}} \times \overrightarrow{\text{stepsize}}_j \tag{6}$$

where $\hat{\text{CF}}$ is a parameter that control step size and is calculated using Eq. (7):

$$\hat{\text{CF}} = \left(1 - \frac{it}{it_{\max}}\right)^{\left(\frac{2-it}{it_{\max}}\right)} \tag{7}$$

Stage 3: Low velocity ($it > \frac{2}{3}it_{\max}$): The population can be modified by the Lévy flight using Eq. (8):

$$\begin{aligned} \overrightarrow{\text{stepsize}}_j &= \mathbf{R}_{\text{levy}} \times (\mathbf{R}_{\text{levy}} \times \overrightarrow{\text{Elite}}_j - \overrightarrow{\text{Prey}}_j), j \in \{1, \dots, N_n\} \\ \overrightarrow{\text{prey}}_j &= \overrightarrow{\text{Elite}}_j + P * \hat{\text{CF}} \times \overrightarrow{\text{stepsize}}_j \end{aligned} \tag{8}$$

MPA also employs a theory in marine predators known as eddy formation or Fish Aggregating Devices (FADs), where the predators contemplate longer jumps in different positions in quest of more food to infuse variation into possible solutions using Eq. (9):

$$\overrightarrow{\text{prey}}_j = \begin{cases} \overrightarrow{\text{prey}}_j + \text{CF} [lb_i + r \times (ub_i - lb_i)] \times \ell & r \leq \text{FADs} \\ \overrightarrow{\text{prey}}_j + [\text{FADs}(1 - r) + r](\overrightarrow{\text{prey}}_{r1} - \overrightarrow{\text{prey}}_{r2}) & \text{else} \end{cases} \tag{9}$$

where $\overrightarrow{\text{prey}}_j$, $\overrightarrow{\text{prey}}_{r1}$, and $\overrightarrow{\text{prey}}_{r2}$ denotes vectors for j th candidate solution, a random finding solution, and another random finding solution, respectively; where, r indicates a random number in [0,1], FADs is a constant equal to 0.2, ℓ a binary vector includes zero and one.

3.2 Opposition-based learning

OBL [32] is a helpful approach in avoiding staleness in competitor solutions [51]. It is an essential concept, which enhances search mechanism exploitation. In meta-heuristic algorithms, convergence occurs almost rapidly when the primary solutions are near the ideal position; otherwise, late convergence is expected. By exploring opposite search zones close to the global optimal, the OBL strategy

produces better results. The OBL strategy operates by searching the search space in two directions. These two directions are defined by one solution, whereas the other direction is defined by the opposite solution. The OBL strategy then chooses the best direction from all solutions [52].

- *Opposition number* The concept of opposite numbers may be explained by describing OBL. It is defined by considering Y_0 as a real number $\in [u, p]$. The opposite number of Y_0 is denoted using Eq. (10) [52]:

$$\bar{Y}_0 = u + p - Y_0. \quad (10)$$

The opposite number in dimensional (M) space is calculated using Eq. 11 and Eq. 12:

$$Y = Y_1, Y_2, Y_3, \dots, Y_M \quad (11)$$

$$\bar{Y} = [\bar{Y}_1, \bar{Y}_2, \bar{Y}_3, \dots, \bar{Y}_M] \quad (12)$$

The items in \bar{Y} are represented by Eq. (13):

$$\bar{Y}_k = u_q + p_q - Y_q \text{ where } q = 1, 2, 3, \dots, M \quad (13)$$

- *Opposition-based optimizations* The opposite item \bar{Y}_0 is changed with the corresponding solution Y_0 throughout the fitness function. If $f_i(Y_0)$ is better than $f_i(\bar{Y}_0)$, Y_0 is constant; oppositely, $Y_0 = \bar{Y}_0$. As a result, the solutions have been modified in terms of the best value of Y and \bar{Y} [53].

3.3 Convolutional neural networks

The main structure of every CNN architecture will be discussed in this subsection. They are a type of deep NNs that is used to recognize and classify images. Recently, CNNs have become an essential method in image analysis, especially when recognizing or detecting faces, text, and medical imaging [13]. CNNs have been successfully used for image classification and segmentation since their initial development in 1989 [54], and they are designed to work as the human brain does in visual perception: they include layers of “neurons” that only react to neurons in their immediate environment. CNNs can extract the topological aspects of an image by concatenating three types of layers[55]: *convolutional layers*, *pooling layers*, and *and FC layers*. Figure 1, shows an example of CNN architecture [7].

- *Convolutional layers* These layers are structured into feature maps according to the local connection concept and weight distribution concept. A weight group known as a filter bank connects all neurons in a feature map to local patches at the previous stage. All units share a filter row on a feature map. For various feature maps,

different filter banks are used. The reason for local connection and weight distribution is to reduce the number of parameters by exploiting the highly linked local pixel neighborhood, and local image characteristics are location-independent. Then, the summation of weights is passed on to an **activation function** such as Sigmoid [56] and rectified linear unit (ReLU) [57]. The activation function facilitates the nonlinear transformation of transmitted data to the following processing phases [58].

- *Pooling layer* As shown in Fig. 1, the pooling layer comes after the convolution layer. It employs a subsampling technique to integrate convolutional layer features comparable to a single layer (semantically). The primary goal of this layer is to reduce an image’s dimension (by close grouping pixels in a specific portion of the image into a single value) while highlighting its features. Some of the common popular kinds of operation performed on this layer are max pooling and main pooling [59].
- *Fully connected layer* The final layer of the CNN is a classifier, which decide the class of the input data based on CNN’s features discovered and extracted. The number of units in the FC layer is equal to the number of classes or classifications. [59].

A CNN model depends on its hyperparameters, so to improve its accuracy, some researchers have proposed that these hyperparameters must be fine-tuned to achieve great results. Table 1 shows hyperparameters and their description related to CNN architecture. As mentioned before in Sect. 1, the meta-heuristic algorithms are widely regarded as excellent techniques to optimize the hyperparameters of the CNN architecture to increase its performance. Figure 2 presents the process of using an optimization algorithm to optimize a CNN’s hyperparameters.

As shown in Fig. 2, optimizing hyperparameters of a CNN starts by initializing the meta-heuristic algorithm’s population, and the number of hyperparameters determines the number of dimensions to optimize. Following that, images must be normalized before being fed to the CNN. The suggested technique puts the CNN architecture in a function so that it may be called later when the fitness function is assessed; at this point, hyperparameters will be optimized. Once the fitness function has been established, the iterations of the meta-heuristic algorithm’s positions are updated according to the algorithm employed. New solutions are assessed, and the best one is selected. Finally, the stop criterion is assessed; if not, the operation is continued to obtain new solutions.

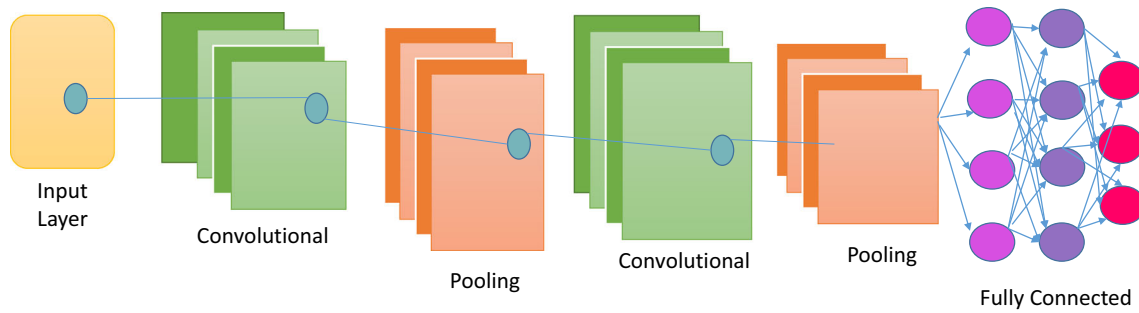
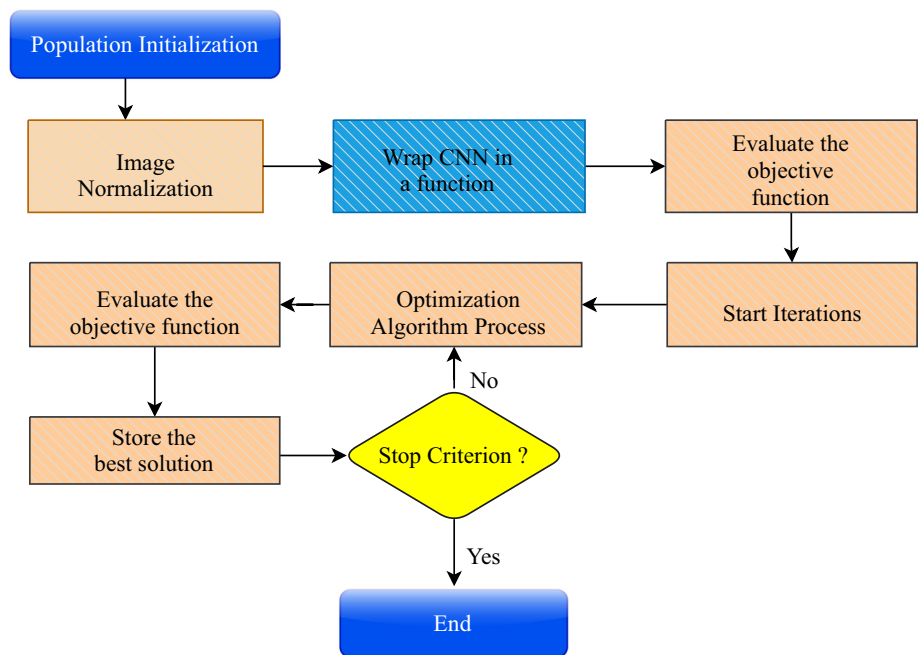


Fig. 1 CNN standard architecture [7]

Table 1 Hyperparameter description of CNN

| Hyperparameter-name | Description |
|------------------------------|--|
| Learning rate | The initial learning rate for the CNN architecture is one of the significant hyperparameters that affect output performance. When the learning rate is low, the model requires more iterations |
| Number of hidden layer units | Expanding the number of hidden layer units enhances the model and reduces computational efficiency |
| Batch size | It refers to the number of sub-samples sent to the network for parameter updates |
| Dropout rate | A dropout is a regularization approach that reduces overfitting by enhancing validation accuracy and consequently generalizing power |
| Activation Function | Activation functions allow DL techniques to learn nonlinear prediction limits |
| Number of epochs | It is the number of times the entire training data is taken through the training process |

Fig. 2 Block diagram of the standard process for hyperparameter optimization in a CNN using meta-heuristic algorithms



3.4 Transfer learning

TL is a key to improving the performance of a DL model on a small dataset, such as medical images. DL models need a lot of data, computing power, and time to train from

scratch. To tackle these issues, pretrained models and fine-tuning (FT) are used. TL enables a DL architecture to learn efficiently from a dataset with fewer samples by transferring learned features from other DL architectures that have earlier learned from large datasets [60]. Using a pretrained

network is one of the most well-known and frequently used strategies to deal with small datasets, such as ImageNet [61], used in the assigned tasks [62]. Several models have been pretrained on the ImageNet, such as AlexNet [63], VGG [64], ResNet [65], Inception [66], and DenseNet [67]. Feature extraction and Fine tuning (FT) are different approaches that TL can use. The feature extraction approach removes FL layers from an ImageNet-trained network while maintaining the remaining network layers, which comprise a sequence of convolution and pooling layers and are known as the convolutional part, a fixed feature extractor. Then, on top of the fixed feature extractor, any classifier can be added.

FT is to replace the FC layers of a pretrained model with new ones and retrain them on the input dataset and fine-tune the kernels in the pretrained model's convolutional part using backpropagation. It is the same as the feature extraction method, except that the final layers of the frozen convolutional part are unfrozen. These layers are then retrained, along with the new classifier obtained via the feature extraction procedure. FT seeks to make the most abstract elements of the pretrained model more relevant to the new goal [19]. The steps for implementing these methods are as follows:

- The classifier of the pretrained model is removed.
- The pretrained model's convolutional base is frozen.
- Add a new classifier and train it on the top of the convolutional part. Then, some layers of the convolutional part are unfreezing.
- Finally, the new classifier and the unfrozen layers are jointly trained.

4 Improved marine predators algorithm

In this section, the proposed IMPA is explained. When examining the performance of the original MPA, it is clear that it does not adequately explore all search space solutions. Furthermore, because it divides the optimization phases into three discrete portions, it suffers from a convergence rate. Thus, the original version of the MPA is improved using the OBL strategy (IMPA) and used to optimize the hyperparameters of the pretrained CNN architecture. Algorithm 1 presents the pseudo-code for the IMPA. The MPA's diversity was improved in the search process using the OBL strategy during the initialization phase to improve the search operation as follows:

$$\text{Opp}_s = lb_a + ub_a - y_b, b \in 1, 2, \dots, N_n \quad (14)$$

where Opp_s is a vector produced by applying OBL, lb_a , and ub_a are lower and upper bounds of the a th component of Y , respectively.

The phases of the proposed IMPA are described in the next subsections:

4.1 Initialization steps in IMPA

The IMPA starts by initializing its parameters: maximum iterations t_{\max} , population size N_n , FADs, P , and dimensions Dim . The MPA begins by initializing the first search agent y^0 and saving outputs. Then, the OBL strategy is applied to determine the Opp_s of the initial population by Eq. (14).

4.2 Optimization processes

The optimization operation is divided into three phases as presented in Sect. 3.1. After completing these phases, the OBL strategy is used to calculate the fitness function for each solution in y and \bar{y} , the proposed method updates the global best solution by calculating and comparing the fitness of y_b and Opp_s .

4.3 Final steps in IMPA

After finishing the optimization process, saving the memory, and updating Elite, and the FADs are calculated by Eq.(9). The proposed method selects the best solution.

4.4 IMPA computational complexity

The IMPA's time and space expenses are explained in this subsection as follows:

1. *Time complexity* The IMPA generates N_n search agents with size Dim , and the initialization time complexity was $O(N_n \times Dim)$. In addition, the IMPA computes the fitness of each search agent as $O(I_{t_{\max}} \times N_n \times Dim)$, where $I_{t_{\max}}$ determines the maximum iterations. Furthermore, the IMPA requires $O(T_t)$ to perform T_t number of its primary processes. So, the time complexity of the IMPA is represented by $O(I_{t_{\max}} \times T_t \times N_n \times Dim)$.
2. *Space complexity* The IMPA space complexity is $O(N_n \times Dim)$.

Algorithm 1 The pseudo-code for the IMPA algorithm.

Input: $It_{max} = 50$, $N_n = 30$, $P = 0.5$, $FADs = 0.2$, lower-bound: lb, upper-bound: ub, Dimension D_{im} .
Output: Best fitness and best positions in search space.
 Randomly initialize the population y_i^0 with dimensions D_{im} .
 Apply OBL on the initial population y_i^0 by Eq. (14) and save result in Opp_s .
while $t \leq It_{max}$ **do**
 for $i \leq N_n$ **do**
 Evaluate y_i using the fitness function and store results in fit_i .
 Compute the fitness value.
 if $Fit_i < FitOpp_i$ **then**
 $y_i = Opp_i$;
 end if
 end for
 Perform memory saving.
if $t < \frac{It_{max}}{3}$ **then**
 Update the prey by Eq. (4).
else if $\frac{It_{max}}{3} < t < 2 \times \frac{It_{max}}{3}$ **then**
 Update first half of the solutions by Eq. (5) and the second half by Eq. (6).
else if $t > 2 \times \frac{It_{max}}{3}$ **then**
 Update the prey by Eq. (8).
end if
 Perform memory save and Elite update.
 Apply FADs by Eq. (9).
 Save the results.
end while
 Return the best solution.

5 The proposed IMPA-ResNet50 classification model

This section introduces the proposed IMPA-ResNet50 model based on TL from a pretrained CNN architecture. The pretrained model employed in this proposed model is ResNet50. The IMPA algorithm is used to optimize the hyperparameters of the pretrained CNN model to get the best performance of this model. The ResNet50 model was trained using TL methods after determining the best values for the parameters. After the model has been trained, it is verified using a different test set. The test set is then used to validate the fully trained model. The proposed model has split into four main phases, as shown in Fig. 3. These phases operate in this order:

1. *Phase 1* Data preprocessing and data augmentation.
2. *Phase 2* Hyperparameters optimization.
3. *Phase 3* The learning phase.
4. *Phase 4* The performance evaluation.

In the first phase, the datasets were enhanced and divided into two training and test sets. Also, multiple data augmentation procedures have been applied to increase training sets. The proposed model has been applied to two datasets, namely CBIS-DDSM and MIAS. In the second phase, the IMPA is used to optimize the hyperparameters in

the pretrained CNN architecture (ResNet50). In the third phase, ResNet50 was completely trained using the values of the hyperparameters determined in the second phase that helped the architecture accurately diagnose the test set in the next phase. The phases of IMPA-ResNet50 will be presented in detail in the following sections.

5.1 Phase 1: data preprocessing and data augmentation

In this phase, data preprocessing and data augmentation were applied to the two mammographic datasets. First, the images were enhanced by removing noise and resizing them to 224×224 resolution before applying data augmentation, thereby minimizing the storage capacity and reducing computational time. Second, multiple data augmentation procedures [68] have been applied to increase training sets, decrease overfitting, speed up the convergence process, and improve generalization. Here, data augmentation was implemented using the Keras ImageDataGenerator to enlarge the images of the dataset's training set. Table 2 lists the used data augmentation approaches and their ranges.

5.2 Phase 2: hyperparameters optimization

The TL approach adopts the exact structure of the pretrained architecture after making small modifications. The most significant difference is that the classifier is replaced with a new one, requiring adjusting or adding several hyperparameter values. Tuning parameters of a CNN has a significant contribution to classification efficiency. As previously mentioned in Sect. 3.3, in this subsection, we determine hyperparameters that the proposed IMPA can optimize. In the proposed IMPA-ResNet50 model, eight hyperparameters are optimized: the learning rate, the batch size, the three dropout rates of the three dropout layers, and the number of units of the first three dense layers. As a result, the search space is eight-dimensional, with each point representing a combination of the eight hyperparameters.

5.3 Phase 3: learning phase

Feature extraction and FT are performed to adapt the ResNet50 model to learn from the used datasets (CBIS-DDSM and MIAS). The convolutional base is unmodified in the feature extraction process, but the basic classifier is replaced by the newest one, which fits the datasets. The new classifier has eight layers: a flatten layer, four dense layers, and three dropout layers separating the dense layers. The IMPA is used to calculate the learning rate for the convolutional layer, the number of neurons in the first three

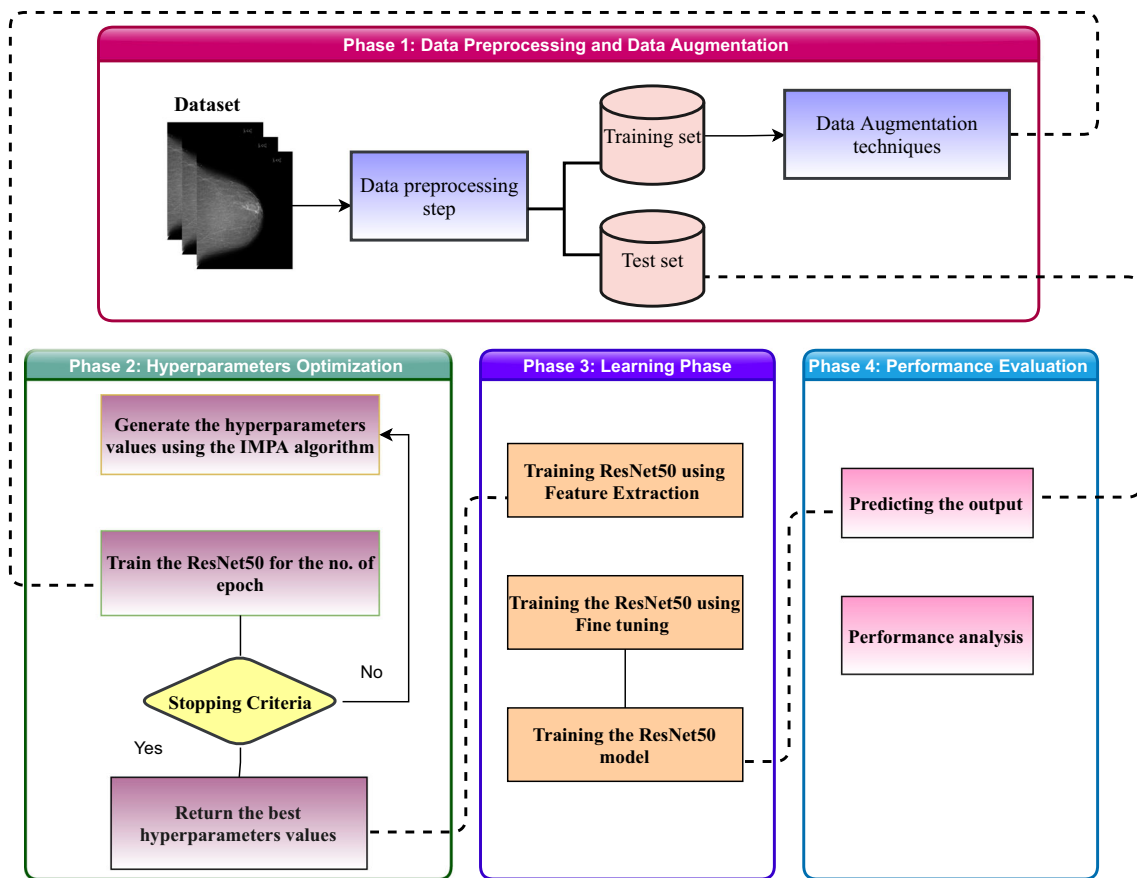


Fig. 3 The proposed IMPA-ResNet50 architecture block-diagram phases

Table 2 The data augmentation approaches and their ranges

| Data-augmentation technique | Range |
|--|---------|
| Shearing | 0.1 |
| Zooming | 0.1 |
| Width shift | 0.3 |
| Height shift | 0.3 |
| Rotation | 15 |
| Featurewise center | True |
| Featurewise standard deviation normalization | True |
| Fill mode | Reflect |
| Vertical flip | True |
| Horizontal flip | True |

dense layers that use the activation function (ReLU), and the rates of all dropout layers. The last dense layer has one neuron with a softmax function. After training the new classifier for some epochs, fine-tuning is performed by retraining the last two blocks of the convolutional part of ResNet50 integrating the new classifier.

5.4 Phase 4: performance evaluation

Accuracy, sensitivity, specificity, precision, F-score, and AUC are the metrics used in this paper to assess the quality of the proposed method. The following is a summary of these metrics.

Accuracy (Acc) This determines how many cases have been correctly categorized. It is expressed by Eq. 15 [7]:

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{15}$$

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

Sensitivity (S_n) This analysis just displays how many of the total positive cases are only approximated correctly. This can be measured using Eq. 16 [7]:

$$S_n = \frac{TP}{(TP + FN)} \tag{16}$$

Specificity metric (S_p) This metric shows how correct the overall pessimistic predictions are and how accurate the normal prediction is. It is expressed using Eq. 17 [7]:

$$S_p = \frac{TN}{(TN + FP)} \quad (17)$$

Precision metric (P_r) This metric indicates how accurate the abnormal breast cancer prediction is. It is denoted using Eq. 18 [7]:

$$P_r = \frac{TP}{(TP + FP)} \quad (18)$$

Average F-score F-score is a metric of the test accuracy, as expressed in Eq. (19) [69]:

$$F1_j = \left\{ \frac{TP_j}{TP_j + FP_j} \right\} \quad (19)$$

$$F1 = \frac{1}{q} \sum_{j=1}^q F1_j$$

AUC The AUC indicates how well a model will perform in various scenarios, expressed using Eq. 20 [7]:

$$AUC = \frac{\sum R_i(I_i) - I_i(I_i + 1)/2}{I_t + I_f} \quad (20)$$

where I_t and I_f represent the number of positive and negative images, and R_i denotes the rate of the i_{th} positive image.

6 Experimental results and performance analysis

This section describes and analyzes the results to validate the proposed IMPA-ResNet50 model's performance for classifying mammography of breast cancer. This section is structured as follows: Sect. 6.1 presents the datasets used in this paper. Section 6.2 presents the platform used in this paper. Section 6.3 presents the experimental parameter settings for the IMPA. Section 6.4 presents the results of the CBIS-DDSM dataset, and Sect. 6.5 describes the results of the MIAS dataset. Section 6.6 presents the comparative results of the proposed IMPA-Resnet50 with four other meta-heuristic algorithms paring with the ResNet50 architecture: GSA-ResNet50, HHO-ResNet50, WOA-ResNet50, and MPA-ResNet50. Eventually, the limitations of the proposed model are illustrated in Sect. 6.7

6.1 Dataset description

In this paper, two datasets were used to evaluate the proposed model including MIAS [70] and CBIS-DDSM [71] described as follows:

1. **CBIS-DDSM dataset**The CBIS-DDSM dataset is an upgraded and standard version of the DDSM

mammogram dataset. Its images were decompressed and changed to the DICOM form for download. To process this dataset, we applied the guidance of CBIS-DDSM and transformed the DICOM format into PNG files to train the method to categorize images as benign or Cancer. The total number of images are 5283. Table 3 presents the specification of the dataset and its number of samples for the training and test sets.

2. **MIAS dataset** The MIAS dataset consists of 322 mammography images with a size 1024×1024 . This dataset has two (abnormal: 113; normal: 209) classes. The abnormal category was divided into two classes: benign contains 65 images, and malignant contains 48 images. Each class in the MIAS dataset contains useful information. The abnormality type, such as calcifications, masses, and asymmetries, is identified in the data. This dataset is categorized into six categories shown in Fig. 1. After applying data augmentation techniques as mentioned in Sect. 5.1 on the MIAS dataset, the number of images increases, as reported in Table 4 (Fig. 4).

6.2 Experimental platform

The IMPA and the IMPA-ResNet50 model are coded on Google Colaboratory [72] platform and implemented by Python 3 with Keras [73]. Keras is a high-level NN API that may be used with TensorFlow, CNTK, or Theano. It is the most popular DL framework. It was created for ease of use, and it has been able to conduct multiple tests and obtain findings as rapidly as possible, with the least amount of delay, allowing for adequate paper.

6.3 Parameter settings

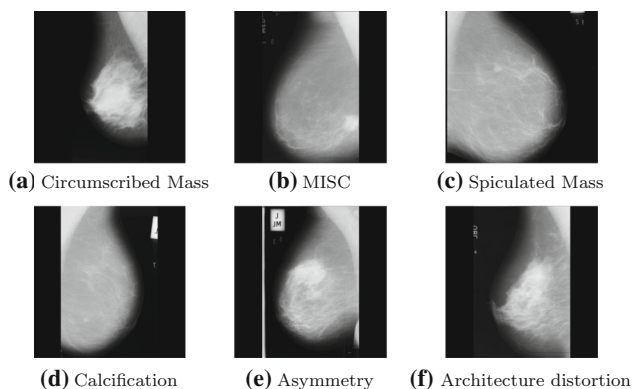
The parameters employed in the proposed model are listed in Table 5. The maximum number of iterations is 50, and the population size is 30, which are almost relative to our number of dimensions. The IMPA algorithm aims to optimize the initial learning rate parameter to fit in the optimal area. The range of the learning rate is between $1e^{-7}$ and $1e^{-3}$; it should be a small value in the fine-tuning method since the number of changes that will occur in the model must be quite small so that the features learned from the feature extraction method are not lost. The batch size value is between [1, 64]; the searching range of batch size is bounded by a lower bound of 1 and the upper bound of 64, and the dropout rates are in the range of [0.1, 0.9]; the searching range of dropout rates is bounded by a lower bound of 0.1 and the upper bound of 0.9 as a good value for

Table 3 Specifications of CBIS-DDSM dataset

| Dataset | Class | No. of training samples | No. of test samples |
|-----------|--------|-------------------------|---------------------|
| CBIS-DDSM | Benign | 1824 | 783 |
| | Cancer | 1873 | 803 |

Table 4 Specifications of MIAS dataset

| Dataset | No. of training samples | No. of test samples | Total images |
|-------------------|-------------------------|---------------------|--------------|
| MIAS | 904 | 386 | 1290 |
| Normal-category | 830 | | |
| Abnormal-category | 460 | | |

**Fig. 4** MIAS mammography Categories [70]**Table 5** Parameter settings for IMPA

| Parameter | Value |
|--|--------------|
| Maximum iteration numbers | 50 |
| Population size | 30 |
| Dimension | 8 |
| Learning rate | [1e-7, 1e-3] |
| Batch size | [1,64] |
| Dropout rate | [0.1,0.9] |
| Number of neurons | [50,500] |
| Maximum number of ResNet50 training epochs | 30 |

dropout in a hidden layer is between 0.1 and 0.9, and the number of neurons is in the range of [50, 500]. The epochs training number of ResNet50 was chosen by experimenting with more than one value. When employing a number of epochs more than 30, the experiment determined that each IMPA's training process takes exponential time. Also, when using less than 30 epochs, the results of the ResNet50 were not adequately accurate. As a result, the ResNet50 was trained using 30 epochs. The Dimension parameter of the IMPA denotes the number of hyperparameters that the

Table 6 Results of the proposed IMPA-ResNet50 on the CBIS-DDSM dataset

| Metrics | IMPA-ResNet50 (%) |
|-------------|-------------------|
| Accuracy | 98.32 |
| Sensitivity | 96.61 |
| Specificity | 98.56 |
| Precision | 98.68 |
| F-score | 97.65 |
| AUC | 97.88 |

proposed IMPA can optimize. It is set to 8 as they are 8 hyperparameters: the learning rate, the batch size, the three dropout rates of the three dropout layers, and the number of units of the first three dense layers.

6.4 Analysis of IMPA-ResNet50 model for CBIS-DDSM dataset

This subsection presents the results of the proposed IMPA-ResNet50 model using hyperparameters determined by the IMPA according to the CBIS-DDSM dataset. In addition, it presents the comparison among other studies and related work. Furthermore, to demonstrate the effectiveness of the IMPA in determining the best values for the hyperparameters of the ResNet50 model that can achieve the most significant accuracy, we compare it with the ResNet50 model based on setting the hyperparameters of the ResNet50 manually, i.e., randomly chosen.

Table 6 presents the results of the IMPA-ResNet50 model on the CBIS-DDSM dataset evaluated in terms of accuracy, sensitivity, specificity, precision, F1-score, and AUC. The proposed method produced 98.32% accuracy on the test set. The average sensitivity, specificity, precision, F1-score, and AUC were 96.61%, 98.56%, 98.68%, 97.65%, and 97.88%, respectively. Table 7 compares the

Table 7 Comparison between the proposed IMPA-ResNet50 model and ResNet50 model on the CBIS-DDSM dataset

| Metrics | ResNet50 (%) | IMPA-ResNet50 (%) | Improvement (%) |
|-------------|--------------|-------------------|-----------------|
| Accuracy | 90.11 | 98.32 | 8.21 |
| Sensitivity | 89.80 | 96.61 | 6.81 |
| Specificity | 90.33 | 98.56 | 8.23 |
| Precision | 89.01 | 98.68 | 9.67 |
| F1-score | 90.00 | 97.65 | 7.65 |
| AUC | 91.88 | 97.88 | 6.00 |

proposed IMPA-ResNet50 model with the MPA-ResNet50 method based on manual search on the CBIS-DDSM dataset. According to the results presented in Table 7, the proposed IMPA-ResNet50 model outperforms ResNet50 model without the hyperparameters optimization. Where, the accuracy of the ResNet50 architecture is 90.11%, the sensitivity is 89.80%, specificity is 90.33%, the precision, F1 score, and AUC of this architecture are 89.01%, 90.00%, 91.88%, respectively. The improvement of the proposed IMPA-ResNet50 model relative to the ResNet50 architecture according to the accuracy is 8.21%, and the improvement according to the sensitivity, specificity, precision, F1 score, and AUC are 6.81%, 8.23%, 9.67%, 7.65%, 6.00%, respectively.

Furthermore, the proposed IMPA-ResNet50 model’s performance was compared with other published studies on breast cancer diagnosis using the CBIS-DDSM dataset in Table 8. These studies [35, 46, 48, 74–79] were chosen for comparison because they are recent models and were trained on a various data samples. The proposed IMPA-

ResNet50 model was compared with other studies in terms of accuracy, sensitivity, specificity, precision, F1-score, and AUC, as shown in Table 8. The symbol – in the table means the comparison method does not have an equivalent metric. In [35], several CNN models were introduced to classify three datasets. The significant performance was on the DDSM dataset that contains 5316 images and the ResNet50 model. The ResNet50 model achieves 97.27% accuracy. In [76], multi-deep CNN models were evaluated on two datasets: CBIS-DDSM and MIAS. They used 5272 images from the first dataset and achieved accuracy, Sensitivity, and specificity of 87.2%, 86.04%, and 89.40%, respectively. The study in [48] proposed a fine-tuned technique to enhance the performance of the ResNet50 model. This method was evaluated on 2620 images from the CBIS-DDSM dataset and achieved accuracy, specificity, Sensitivity, AUC of 93.15%, 92.17%, 93.83%, 0.95, respectively. While in [46], the authors hybridized the SVM and the CNN on CBIS-DDSM dataset to enhance the performance. The obtained accuracy was

Table 8 Comparison between the proposed IMPA-ResNet50 model and other related studies on the CBIS-DDSM dataset

| References | No. of images | Dataset | Model | Accuracy (%) | Sensitivity (%) | Specificity | Precision (%) | F1-score (%) | AUC |
|------------|---------------|-----------|---------------------|--------------|-----------------|-------------|---------------|--------------|--------|
| [35] | 5316 | DDSM | ResNet50 | 97.35 | – | – | – | – | 0.97 |
| [35] | 5316 | DDSM | VGG16 | 97.12 | – | – | – | – | 0.98 |
| [74] | 600 | DDSM | CNN | 96.7 | – | – | – | – | – |
| [75] | 2400 | DDSM | CNN-YOLO | 97.0 | 93.20 | 94.00 | – | – | 96.45% |
| [48] | 2620 | CBIS-DDSM | Fine-tuned ResNet50 | 93.15 | 93.83 | 92.17% | – | – | 95.0% |
| [76] | 5272 | CBIS-DDSM | ResNet50 | 87.2 | 86.04 | 89.40% | – | – | 95.00% |
| [46] | 5272 | CBIS-DDSM | Fine-tuned AlexNet | 87.20 | 86.2 | 87.7% | 88.0 | 87.1 | 94.00% |
| [77] | 3568 | CBIS-DDSM | ResNet50 | 96.6 | 92.95 | 88.60% | – | – | 93.4% |
| [78] | 11,562 | DDSM | DCNN | 92.80 | – | – | – | – | – |
| [79] | 2781 | CBIS-DDSM | AdaBoost | 90.91 | 82.96 | 98.38% | 86.00 | – | 98.32% |
| Proposed | 5283 | CBIS-DDSM | IMPA-ResNet50 | 98.32 | 95.61 | 98.56% | 98.68 | 97.65 | 97.88% |

87.2% as shown in Table 8. In addition, in [74], 600 images from the DDSM dataset have been classified and achieved 96.7% accuracy. In [75], a YOLO-CNN method was proposed for classifying 2400 images from DDSM dataset as seen in row 5 of Table 8. Several CNN models used in [77] to classify three breast cancer datasets. A DCNN model proposed in [78] uses a large number of images; 11562 images. The gained accuracy is 92.80%. According to the table, the proposed IMPA-ResNet50 model's classification performance is superior to other approaches. It outperforms all comparison methods in terms of all evaluation matrices.

6.5 Analysis of IMPA-ResNet50 model for MIAS dataset

This subsection presents the results of the proposed method IMPA-ResNet50 using hyperparameters calculated by the IMPA algorithm according to the MIAS dataset. Furthermore, it presents the comparison between other related studies. In addition, to demonstrate the effectiveness of the IMPA in determining the best values for the hyperparameters of the ResNet50 model that can achieve the most significant accuracy, we compare it with the ResNet50 model based on setting the hyperparameters of the ResNet50 model manually, i.e., randomly chosen without optimization.

Table 9 presents the results of IMPA-ResNet50 for the MIAS dataset evaluated in terms of accuracy, sensitivity, specificity, precision, F1-score, and AUC. The proposed method achieved 98.88% accuracy. The average sensitivity, specificity, precision, F1-score, and AUC were 97.61%, 98.40%, 98.30%, 97.10%, and 99.24%, respectively. Table 10 compares IMPA-ResNet50 model with the MPA-ResNet50 method based on manual search for the MIAS dataset. According to the results in Table 10, the proposed ResNet50 model outperform the ResNet50 architecture that selected without the hyperparameters optimization. Where, the accuracy of the ResNet50 architecture is 87.50%, the

Table 9 Results of the proposed IMPA-ResNet50 on the MIAS dataset

| Metrics | IMPA-ResNet50 (%) |
|-------------|-------------------|
| Accuracy | 98.88 |
| Sensitivity | 97.61 |
| Specificity | 98.40 |
| Precision | 98.30 |
| F-score | 97.10 |
| AUC | 99.24 |

sensitivity is 88.10%, specificity is 86.12%, the precision, F1 score, and AUC of this architecture are 87.32%, 87.88%, 89.01%, respectively. The improvement of the proposed IMPA-ResNet50 model relative to the ResNet50 architecture according to the accuracy is 11.38%, and the improvement according to the sensitivity, specificity, precision, F1 score, and AUC are 9.51%, 12.28%, 10.98%, 9.22%, 10.23%, respectively.

In addition, the proposed IMPA-ResNet50 model's performance was compared with other published studies on breast cancer diagnosis using the MIAS dataset. The comparison studies, [35, 36, 43, 76, 80, 81], were chosen because they are based on CNN architectures and have used the same dataset. The methods were compared in terms of accuracy, sensitivity, specificity, precision, F1-score, and AUC, as shown in Table 11. The symbol - in the table means the comparison method does not have the equivalent metric. In [35], several CNN models were introduced to classify three datasets. The obtained results on the MIAS dataset are 98.23% accuracy and 0.99 AUC. In [36], the graph convolutional network is used with CNN to classify 322 images; the results of that method are reported in row 4 on Table 11. In [43], an improved CNN model was introduced to classify the MIAS breast cancer dataset that achieves 89.47%, 90.71%, and 90.50% for sensitivity, specificity, and accuracy, respectively. In [80], a DenseNet201 model evaluated on the MIAS dataset and obtain 92.73% accuracy. Based on the table, the proposed IMPA-ResNet50 model is superior to other approaches. It outperforms all comparison methods in terms of all evaluation matrices.

6.6 Comparison with other optimization algorithms

This subsection presents the comparative results between the IMPA algorithm and other well-known meta-heuristic algorithms to demonstrate that the IMPA algorithm effectively determines the best values for the ResNet50 architecture's hyperparameters to reach high accuracy. It was compared to four other meta-heuristic algorithms that have received much attention and highly cited meta-heuristics: the GSA algorithm [82], the HHO algorithm [83], the WOA algorithm [84], and the original MPA algorithm. These algorithms are also hybrid with the ResNet50 architecture produce models named GSA-ResNet50, HHO-ResNet50, WOA-ResNet50, and MPA-ResNet50, respectively. For a fair comparison between the IMPA and the other compared algorithms, all the compared algorithms have the same parameters as the IMPA, as stated in Table 5. Table 12 presents the comparison between the

Table 10 Comparison between the proposed IMPA-ResNet50 model and the ResNet50 model on the MIAS dataset

| Metrics | ResNet50 (%) | IMPA-ResNet50 (%) | Improvement (%) |
|-------------|--------------|-------------------|-----------------|
| Accuracy | 87.50 | 98.88 | 11.38 |
| Sensitivity | 88.10 | 97.61 | 9.51 |
| Specificity | 86.12 | 98.40 | 12.28 |
| Precision | 87.32 | 98.30 | 10.98 |
| F1-score | 87.88 | 97.10 | 9.22 |
| AUC | 89.01 | 99.24 | 10.23 |

Table 11 Comparison between the proposed IMPA-ResNet50 model and other related studies on the MIAS dataset

| References | No. of images | Dataset | Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) | F1-score (%) | AUC (%) |
|------------|---------------|---------|-----------------|--------------|-----------------|-----------------|---------------|--------------|---------|
| [35] | 322 | MIAS | ResNet50 | 98.23 | – | – | – | – | 99.0 |
| [76] | 1288 | MIAS | Fine-tuned DCNN | 95.4 | 96.60 | 92.10 | – | – | 99.00 |
| [36] | 322 | MIAS | CNN-GCN | 96.10 ± 1.60 | 96.20 ± 2.90 | 96.00 ± 2.30 | – | – | – |
| [80] | 330 | MIAS | DenseNet201 | 92.73 | 94.58 | 91.67 | – | – | – |
| [81] | 322 | MIAS | CNN | 82.68 | 82.73 | 82.71 | – | – | – |
| [43] | 322 | MIAS | CNN | 89.47 | 90.71 | 90.50 | – | – | – |
| Proposed | 1290 | MIAS | IMPA-ResNet50 | 98.88 | 97.61 | 98.40 | 98.30 | 97.10 | 99.24 |

Table 12 Comparison between the proposed IMPA-ResNet50 model with the MPA-ResNet50, GSA-ResNet50, HHO-ResNet50, and WOA-ResNet50 models

| Dataset | Classification Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) | F-score (%) |
|-----------|----------------------|--------------|-----------------|-----------------|---------------|-------------|
| CBIS-DDSM | IMPA-ResNet50 | 98.32 | 96.61 | 98.56 | 98.68 | 97.65 |
| | MPA-ResNet50 | 95.95 | 93.03 | 95.28 | 94.22 | 93.85 |
| | GSA-ResNet50 | 95.48 | 94.16 | 95.00 | 95.00 | 94.00 |
| | HHO-ResNet50 | 94.55 | 93.12 | 94.84 | 94.12 | 94.50 |
| | WOA-ResNet50 | 94.13 | 93.10 | 94.00 | 94.00 | 94.00 |
| MIAS | IMPA-ResNet50 | 98.88 | 97.61 | 98.40 | 98.30 | 97.10 |
| | MPA-ResNet50 | 94.95 | 94.03 | 94.28 | 94.22 | 94.85 |
| | GSA-ResNet50 | 94.38 | 94.00 | 93.38 | 94.16 | 94.00 |
| | HHO-ResNet50 | 94.30 | 93.50 | 94.18 | 93.69 | 94.00 |
| | WOA-ResNet50 | 93.38 | 93.00 | 93.00 | 93.00 | 93.00 |

proposed IMPA-ResNet50 model with the MPA-ResNet50, GSA-ResNet50, HHO-ResNet50, and WOA-ResNet50 models. According to the comparative results, the IMPA is more suitable for combining with the ResNet50 architecture to classify the mammography breast cancer datasets. It was able to select the optimal hyperparameter values for the ResNet50, resulting in a greater accuracy ratio for this architecture.

In summary, the following observations from the reported experiments are worth mentioning

- *In terms of accuracy* It is observed that the proposed IMPA-ResNet50 model outperforms the MPA-ResNet50 model, which means the improved version of the MPA achieves a significant performance when used to optimize the ResNet50 hyperparameters compared to the original MPA algorithm. The MPA-ResNet50 model achieves an accuracy of 95.59% on the CBIS-DDSM dataset and 94.95% on the MIAS dataset. Also, the IMPA-ResNet50 model outperforms all other compared models. In the case of using the

GSA algorithm to select the hyperparameters of the ResNet50, it achieves an accuracy of 95.48% on the CBIS-DDSM dataset and 94.38% on the MIAS dataset. In comparison, the HHO-ResNet50 model achieves an accuracy of 94.55% on the CBIS-DDSM dataset and 94.30% on the MIAS dataset. While the WOA-ResNet50 model achieves an accuracy of 94.13% on the CBIS-DDSM dataset and 93.38% on the MIAS dataset.

- *In terms of sensitivity, specificity, precision, and F-score* As listed in Table 12, the performance of the proposed IMPA-ResNet50 model outperforms all other compared models, which means the IMPA algorithm is very appropriate in conjunction with the ResNet50 architecture.

6.7 Limitations of the proposed model

This paper proposes an efficient breast cancer classification model that depends on the hybridization of pretrained CNN architecture and an improved meta-heuristic optimization algorithm. Although the proposed IMPA-ResNet50 model achieves high classification performance in breast cancer detection from mammography images, future studies need to address some limitations. The limitation of the IMPA algorithm and the limitation of the proposed IMPA-ResNet50 model are illustrated as follows:

- Because the IMPA algorithm is merged with the OBL strategy, it is comparatively computationally expensive than the original MPA algorithm.
- The running time may increase when adding the OBL, but the increase in the time here is for getting more performance compared to the original algorithm.
- The IMPA-ResNet50 was only implemented to classify mammography images. These results are limited to a specific dataset, MIAS dataset, and CBIS-DDSM dataset and may not be generalized to the other dataset.
- The performance of the IMPA-ResNet50 is closely related to medical imaging applications.
- The IMPA algorithm success in determining the values of the hyperparameters of the ResNet50 architecture only, and it may not be generalized to other pretrained CNN architecture.

7 Conclusion and future work

Deep Learning is one of the essential techniques in medical imaging classification. Convolution Neural Networks are examples of common DL techniques used for biomedical image classification that involves layer-wise automatic

feature extraction. The preparation of CNNs for classification objectives depends on the knowledge of hyperparameter tuning. Hyperparameters for each layer are different. Because these hyperparameters determine a CNN model's performance, they need to be fine-tuned to achieve great results. It is not suitable to optimize hyperparameters manually because manually selecting them is quite a complex and time-consuming task. Meta-heuristic algorithms have significantly influenced hyperparameter optimization in several fields. This paper proposes a novel classification model for breast cancer based on a hybridized pretrained CNN architecture (ResNet50) and an improved meta-heuristic optimization algorithm. The marine predators algorithm (MPA) is the used optimization algorithm, and we improve it using the OBL strategy to improve exploitation and avoid getting stuck in the local optimal. The Improved Marine Predators Algorithm (IMPA) is used to find the best values for the hyperparameters of the ResNet50 architecture, resulting in a model called IMPA-ResNet50. To the best of our knowledge, this is the first paper using the IMPA as an optimization algorithm to optimize the hyperparameters for the ResNet50 architecture for breast cancer classification. The proposed IMPA-ResNet50 model includes four phases: (1) the data preprocessing and augmentation phase, (2) hyperparameter optimization phase, (3) learning phase, and (4) performance evaluation phase. The proposed model is compared with other state-of-the-art methods and other CNN approaches. The comparison results showed the effectiveness of the proposed model in diagnosing breast cancer. The evaluation is performed on two mammography datasets: the curated breast imaging subset of DDSM (CBIS-DDSM) and the mammographic image analysis society (MIAS). To demonstrate the effectiveness of the IMPA in determining the best values for the hyperparameters of the ResNet50 model that can achieve the most significant accuracy, we compare it with the MPA-ResNet50 model, also compare it with the pretrained CNN ResNet50 model based on the manual search. The first experiment applied to the CBIS-DDSM consisted of 5283 images for two classes, normal and abnormal. The obtained accuracy was 98.32% in the testing phase according to the IMPA-ResNet50 model and 95.95% accuracy according to the MPA-ResNet50 model, which means the improved version for the MPA (i.e., IMPA) significantly improves performance when used to optimize the ResNet50 hyperparameters compared to the original MPA. The second experiment was performed using another dataset (MIAS) consisting of 1290 images after applying the data augmentation techniques. The obtained accuracy was 98.88% according to IMPA-ResNet50 95.95% accuracy according to MPA-ResNet50. Also, compared to four meta-heuristic algorithms: GSA, HHO, WOA, and the original MPA algorithm, the results

of the comparison showed the effectiveness of the proposed algorithm. The proposed method's performance has been assessed using five measures: accuracy, sensitivity, specificity, precision, and F-score. The results showed that the proposed model gets better results than the other competing algorithms, which means using the IMPA as a metaheuristic optimization algorithm to determine the ResNet50 architecture's hyperparameters boosts the classification models to achieve the best performance for a breast cancer diagnostic.

In future work, multiple datasets with more images will be used and evaluated in the proposed model. Also, various pre-trained models such as DensNet201, DensNet121, and Inception will be used to classify breast cancer in conjunction with the proposed IMPA algorithm. In addition, various metaheuristic algorithms will be used for hyperparameter tuning. Also, we will test the proposed method's performance in solving different medical image classification problems and different diagnosis applications, applying the IMPA algorithm to solving other medical and engineering issues, and using it as a feature selection method. Furthermore, various feature extraction methods will be used with CNN architecture to improve the classification accuracy further. Transfer learning with other models can enhance the performance.

Author contributions EHH contributed to supervision, software, methodology, conceptualization, formal analysis, investigation, visualization, and writing-review & editing. MM contributed to software, resources, data curation, conceptualization, formal analysis, investigation, visualization, and writing-review & editing. AAA contributed to supervision and writing-review & editing. All authors read and approved the final paper.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

Human and animal rights This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright

holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Hamidinekoo A, Denton E, Rampun A, Honnor K, Zwiggelaar R (2018) Deep learning in mammography and breast histology, an overview and future trends. *Med Image Anal* 47:45–67
2. Kousalya K, Saranya T (2021) Improved the detection and classification of breast cancer using hyper parameter tuning. *Materials Today: Proceedings*
3. Shahzad A, Usman AM, Muhammad S, Anam T, Khan Shoab A (2018) Decision support system for detection of hypertensive retinopathy using arteriovenous ratio. *Artif Intell Med* 90:15–24
4. Rajinikanth V, Satapathy SC, Fernandes SL, Nachiappan S (2017) Entropy based segmentation of tumor from brain mr images-a study with teaching learning based optimization. *Pattern Recogn Lett* 94:87–95
5. Fonseca P, Mendoza J, Wainer J, Ferrer J, Pinto J, Guerrero J, Castaneda B (2015) Automatic breast density classification using a convolutional neural network architecture search procedure. In: *Medical imaging 2015: computer-aided diagnosis*. International Society for Optics and Photonics, vol 9414, p 941428
6. Khan MA, Sharif M, Akram T, Yasmin M, Nayak RS (2019) Stomach deformities recognition using rank-based deep features selection. *J Med Syst* 43(12):329
7. Houssein EH, Emam MM, Ali AA, Suganthan PN (2020) Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. *Expert Syst Appl* 167:114161
8. Hossam A, Harb HM, AbdElKader HM (2018) Automatic image segmentation method for breast cancer analysis using thermography. *J Eng Sci* 46(1):12–32
9. Qi X, Zhang L, Chen Y, Pi Y, Chen Y, Lv Q, Yi Z (2019) Automated diagnosis of breast ultrasonography images using deep neural networks. *Med Image Anal* 52:185–198
10. Wang Y, Feng Y, Zhang L, Wang Z, Lv Q, Yi Z (2021) Deep adversarial domain adaptation for breast cancer screening from mammograms. *Med Image Anal* 73:102147
11. Dhawan AP (2011) *Medical image analysis*, vol 31. Wiley
12. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ (2017) Deep learning for brain mri segmentation: state of the art and future directions. *J Digital Imaging* 30(4):449–459
13. Yap MH, Pons G, Martí J, Ganau S, Sentís M, Zwiggelaar R, Davison AK, Robert Martí (2017) Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J Biomed Health Inform* 22(4):1218–1226
14. Houssein EH, Abohashima Z, Elhoseny M, Mohamed WM (2022) Hybrid quantum-classical convolutional neural network model for COVID-19 prediction using chest X-ray images. *J Comput Des Eng* 9(2):343–363
15. Khan S, Islam N, Jan Z, Din IU, Rodrigues JJPC (2019) A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recogn Lett* 125:1–6
16. Ucar F, Korkmaz D (2020) Covidiagnosis-net: deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (covid-19) from x-ray images. *Med Hypotheses* 140:109761
17. Mohakud R, Dash R (2021) Survey on hyperparameter optimization using nature-inspired algorithm of deep convolution neural network. In: *Intelligent and cloud computing*. Springer, pp 737–744
18. Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. *J Mach Learn Res* 13(2):281–305

19. Ezzat D, Hassanien AE, Ella HA (2020) An optimized deep learning architecture for the diagnosis of covid-19 disease based on gravitational search optimization. *Appl Soft Comput* 98:106742
20. Cuevas E, Gálvez J, Avalos O (2020) Introduction to optimization and metaheuristic methods. In: *Recent metaheuristics algorithms for parameter identification*. Springer, pp 1–8
21. Hashim FA, Houssein EH, Hussain K, Mabrouk MS, Al-Atabany W (2020) A modified henry gas solubility optimization for solving motif discovery problem. *Neural Comput Appl* 32(14):10759–10771
22. Houssein EH, Neggaz N, Hosney ME, Mohamed WM, Hassaballah M (2021) Enhanced harris hawks optimization with genetic operators for selection chemical descriptors and compounds activities. *Neural Comput Appl* 33:1–18
23. Houssein EH, Saber E, Ali AA, Wazery YM (2021) Centroid mutation-based search and rescue optimization algorithm for feature selection and classification. *Expert Syst Appl* 191:116235
24. Houssein EH, Gad AG, Wazery YM, Suganthan PN (2021) Task scheduling in cloud computing based on meta-heuristics: review, taxonomy, open challenges, and future trends. *Swarm Evolut Comput* 62:100841
25. Houssein EH, Ibrahim IE, Neggaz N, Hassaballah M, Wazery YM (2021) An efficient ecg arrhythmia classification method based on manta ray foraging optimization. *Expert Syst Appl* 181:115131
26. Houssein EH, Mahdy MA, Blondin MJ, Shebl D, Mohamed WM (2021) Hybrid slime mould algorithm with adaptive guided differential evolution algorithm for combinatorial and global optimization problems. *Expert Syst Appl* 174:114689
27. Hassan MH, Houssein EH, Mahdy MA, Kamel S (2021) An improved manta ray foraging optimizer for cost-effective emission dispatch problems. *Eng Appl Artif Intell* 100:104155
28. Faramarzi A, Heidarnejad M, Mirjalili S, Gandomi AH (2020) Marine predators algorithm: a nature-inspired metaheuristic. *Expert Syst Appl* 152:113377
29. Elaziz MA, Ewees AA, Yousri D, Alwerfali HSN, Awad QA, Lu S, Al-Qaness MAA (2020) An improved marine predators algorithm with fuzzy entropy for multi-level thresholding: real world example of covid-19 ct image segmentation. *Ieee Access* 8:125306–125330
30. Črepinšek M, Liu S-H, Mernik M (2013) Exploration and exploitation in evolutionary algorithms: a survey. *ACM Comput Surv (CSUR)* 45(3):1–33
31. Morales-Castañeda B, Zaldivar D, Cuevas E, Fausto F, Rodríguez A (2020) A better balance in metaheuristic algorithms: Does it exist? *Swarm Evolut Comput* 54:100671
32. Tizhoosh HR (2005) Opposition-based learning: a new scheme for machine intelligence. In: *International conference on Computational intelligence for modelling, control and automation, 2005 and international conference on intelligent agents, web technologies and internet commerce*. IEEE, vol 1, pp 695–701
33. Rojas-Morales N, Rojas M-CR, Ureta EM (2017) A survey and classification of opposition-based metaheuristics. *Comput Ind Eng* 110:424–435
34. Desai M, Shah M (2020) An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (mlp) and convolutional neural network (cnn). *Clin eHealth* 4:1–11
35. Chougrad H, Zouaki H, Alheyane O (2018) Deep convolutional neural networks for breast cancer screening. *Comput Methods Prog Biomed* 157:19–30
36. Zhang Y-D, Satapathy SC, Guttery DS, Górriz JM, Wang S-H (2021) Improved breast cancer classification through combining graph convolutional network and convolutional neural network. *Inf Process Manag* 58(2):102439
37. Wang Y, Choi EJ, Choi Y, Zhang H, Jin GY, Ko S-B (2020) Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning. *Ultrasound Med Biol* 46(5):1119–1132
38. Masud M, Rashed AEE, Hossain MS (2020) Convolutional neural network-based models for diagnosis of breast cancer. *Neural Comput Appl*:1–12
39. Eroğlu Y, Yildirim M, Çinar A (2021) Convolutional neural networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mrmr. *Comput Biol Med* 133:104407
40. Zhang H, Han L, Chen K, Peng Y, Lin J (2020) Diagnostic efficiency of the breast ultrasound computer-aided prediction model based on convolutional neural network in breast cancer. *J Digital Imaging* 33:1218–1223
41. Boumaraf S, Liu X, Zheng Z, Ma X, Ferkous C (2021) A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images. *Biomed Signal Process Control* 63:102192
42. Sánchez-Cauce R, Pérez-Martín J, Luque M (2021) Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data. *Comput Methods Prog Biomed* 204:106045
43. Ting FF, Tan YJ, Sim KS (2019) Convolutional neural network improvement for breast cancer classification. *Expert Syst Appl* 120:103–115
44. Kim M, Zuallaert J, De Neve W (2017) Towards novel methods for effective transfer learning and unsupervised deep learning for medical image analysis. In: *Doctoral consortium (DCBIOSTEC 2017)*, pp 32–39
45. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-first AAAI conference on artificial intelligence*
46. Ragab Dina A, Sharkas Maha, Marshall Stephen, Ren Jinchang (2019) Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ*, 7:e6201,
47. Saranyaraj D, Manikandan M, Maheswari S (2020) A deep convolutional neural network for the early detection of breast carcinoma with respect to hyper-parameter tuning. *Multimed Tools Appl* 79(15):11013–11038
48. Chen Y, Zhang Q, Wu Y, Liu B, Wang M, Lin Y (2018) Fine-tuning resnet for breast cancer classification from mammography. In: *The international conference on healthcare science and engineering*. Springer, pp 83–96
49. Dokeroglu T, Sevinc E, Kucukyilmaz T, Cosar A (2019) A survey on new generation metaheuristic algorithms. *Comput Ind Eng* 137:106040
50. Houssein EH, Emam MM, Ali AA (2021) Improved manta ray foraging optimization for multi-level thresholding using covid-19 ct images. *Neural Comput Appl* 33(24):16899–16919
51. Aarts E, Aarts EHL, Lenstra JK (2003) *Local search in combinatorial optimization*. Princeton University Press
52. Tubishat M, Idris N, Shuib L, Abushariah MAM, Mirjalili S (2020) Improved salp swarm algorithm based on opposition based learning and novel local search algorithm for feature selection. *Expert Syst Appl* 145:113122
53. Elaziz MA, Oliva D, Xiong S (2017) An improved opposition-based sine cosine algorithm for global optimization. *Expert Syst Appl* 90:484–500
54. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Backpropagation applied to handwritten zip code recognition. *Neural Comput* 1(4):541–551
55. Lumini A, Nanni L (2019) Deep learning and transfer learning features for plankton classification. *Ecol Inform* 51:33–43

56. Wang Y, Li Y, Song Y, Rong X (2020) The influence of the activation function in a convolutional neural network model of facial expression recognition. *Appl Sci* 10(5):1897
57. Glorot X, Bordes A, Bengio Y (2011) Deep sparse rectifier neural networks. In: *Proceedings of the fourteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings*, pp 315–323
58. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS (2016) Deep learning for visual understanding: a review. *Neurocomputing* 187:27–48
59. Gaspar A, Oliva D, Cuevas E, Zaldívar D, Pérez M, Pajares G (2021) Hyperparameter optimization in a convolutional neural network using metaheuristic algorithms. In: *Metaheuristics in machine learning: theory and applications*. Springer, pp 37–59
60. Pardamean B, Cenggoro TW, Rahutomo R, Budiarto A, Karupiah EK (2018) Transfer learning from chest x-ray pre-trained convolutional neural network for learning mammogram data. *Procedia Comput Sci* 135:400–407
61. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25:1097–1105
62. Yamashita R, Nishio M, Do RKG, Togashi K (2018) Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9(4):611–629
63. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp 1097–1105
64. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. Preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
65. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 770–778
66. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016). Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826
67. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 4700–4708
68. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. *J Big Data* 6(1):1–48
69. Goutte C, Gaussier E (2005) A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In: *European conference on information retrieval*. Springer, pp 345–359
70. Suckling PJ (1994) The mammographic image analysis society digital mammogram database. *Digital Mammography*:375–386
71. Lee RS, Gimenez F, Hoogi A, Rubin D (2016) Curated breast imaging subset of dds. *Cancer Imaging Arch* 8:2016
72. Carneiro T, Da Nóbrega RVM, Nepomuceno T, Bian G-B, De Albuquerque VHC, Filho PPR (2018) Performance analysis of google colab as a tool for accelerating deep learning applications. *IEEE Access* 6:61677–61685
73. Chollet F et al. (2015) Keras
74. Jiao Z, Gao X, Wang Y, Li J (2016) A deep feature based framework for breast masses classification. *Neurocomputing* 197:221–231
75. Al-Masni MA, Al-Antari MA, Park J-M, Gi G, Kim T-Y, Rivera P, Valarezo E, Choi M-T, Han S-M, Kim T-S (2018) Simultaneous detection and classification of breast masses in digital mammograms via a deep learning yolo-based cad system. *Comput Methods Prog Biomed* 157:85–94
76. Ragab DA, Attallah O, Sharkas M, Ren J, Marshall S (2021) A framework for breast cancer classification using multi-dcnns. *Comput Biol Med* 131:104245
77. Khan HN, Shahid AR, Raza B, Dar AH, Alquhayz H (2019) Multi-view feature fusion based four views model for mammogram classification using convolutional neural network. *IEEE Access* 7:165724–165733
78. Song R, Li T, Wang Y (2020) Mammographic classification based on xgboost and dcnn with multi features. *IEEE Access* 8:75011–75021
79. Zhang H, Renzhong W, Yuan T, Jiang Z, Huang S, Jinpeng W, Hua J, Niu Z, Ji D (2020) De-ada*: a novel model for breast mass classification using cross-modal pathological semantic mining and organic integration of multi-feature fusions. *Inf Sci* 539:461–486
80. Xiang Yu, Zeng N, Liu S, Zhang Y-D (2019) Utilization of densenet201 for diagnosis of breast abnormality. *Mach Vis Appl* 30(7):1135–1144
81. Tan YJ, Sim KS, Ting FF (2017) Breast cancer detection using convolutional neural networks for mammogram imaging system. In: *2017 international conference on robotics, automation and sciences (ICORAS)*. IEEE, pp 1–5
82. Rashedi E, Nezamabadi-Pour H, Saryazdi S (2009) Gsa: a gravitational search algorithm. *Inf Sci* 179(13):2232–2248
83. Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H (2019) Harris hawks optimization: Algorithm and applications. *Fut Gen Comput Syst* 97:849–872
84. Mirjalili S, Lewis A (2016) The whale optimization algorithm. *Adv Eng Softw* 95:51–67

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.