

# Modelling of Hybrid Meta Heuristic Based Parameter Optimizers with Deep Convolutional Neural Network for Mammogram Cancer Detection

P. Ashwini<sup>1</sup>, Dr N. Suguna<sup>2</sup>, Dr. N. Vadivelan<sup>3</sup>

<sup>1</sup>Research Scholar, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Annamalai University, Chidambaram, India

ashwinireddy90@gmail.com

<sup>2</sup>Associate Professor, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Annamalai University, Chidambaram, India

rajusuguna81@gmail.com

<sup>3</sup>Professor, Department of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, Hyderabad, India. velancse@gmail.com

**Abstract**— Breast cancer (BC) is the common type of cancer among females. Mortality from BC could be decreased by identifying and diagnosing it at an earlier phase. Different imaging modalities are used to detect BC, like mammography. Even with proven records as a BC screening tool, mammography is time-consuming and has constraints, namely lower sensitivity in women with dense breast tissue. Computer-Aided Diagnosis or Detection (CAD) system assists a proficient radiologist to identify BC at an earlier stage. Recently, the advancement in deep learning (DL) methods are employed to mammography assist radiologists to increase accuracy and efficiency. Therefore, this study presents a metaheuristic-based hyperparameter optimization with deep learning-based breast cancer detection on mammogram images (MHODL-BCDMI) technique. The presented MHODL-BCDMI technique mainly focused on the recognition and classification of breast cancer on digital mammograms. To achieve this, the MHODL-BCDMI technique employs pre-processing in two stages: Wiener Filter (WF) based noise elimination and contrast enhancement. Besides, the MHODL-BCDMI technique exploits densely connected networks (DenseNet201) model for feature extraction purposes. For BC classification and detection, a hybrid convolutional neural network with a gated recurrent unit (HCNN-GRU) model is used. Furthermore, three hyperparameter optimizers are employed namely cat swarm optimization (CSO), harmony search algorithm (HSA), and hybrid grey wolf whale optimization algorithm (HGWWOA). Finally, the U2Net segmentation approach is used for the classification of benign and malignant types of cancer. The experimental analysis of the MHODL-BCDMI method is tested on a digital mammogram image dataset and the outcomes are assessed in terms of diverse metrics. The simulation results highlighted the enhanced cancer detection performance of the MHODL-BCDMI technique over other recent algorithms.

**Keywords**- Mammogram images; Breast cancer; DenseNet; Hyperparameter optimizer; Computer aided diagnoses.

## I. INTRODUCTION

The most deadly cancer among women in world is Breast cancer (BC) [1]. The early identification of malignancy helps in the disease diagnosis and it can help strongly to enhance the survival rate. For detecting BC, mammography is the most promising method among other methods and radiologists frequently preferred this method [2]. Mammogram images are generally noisy and of low contrast. In breast mammography, bright regions denote tumours. In few mammogram images, normal dense tissues and malignant tissues both may be present [3]. Only by applying thresholding, differentiating normal dense and malignant tissues is impossible [4]. Understanding the data of mass region of tumorous lesions in a mammogram is indispensable and helps to detect the cancer and its

segmentation. Consequently, detection of tumorous lesions in mammogram images is an active area of research [5].

Computer-Aided Diagnosis (CAD) is devised using classification and feature extraction that is a useful method for doctors in detecting and diagnosing abnormalities. The primary goal of CAD technique is to solve the task of interpreting DMs [6]. The objective of the mechanism is to correctly interpret DMs and effectively diagnose cancer. The CAD structures have been framed to resolve the dependence of the operator concerning diagnosis and cut the costs of clinical complementary technology [7]. CAD evaluated the knowledge that a computer or person collects and provides an outcome to determine type of lesion and whether that is tumorous or not. Medical imaging technology that adopts CAD-related Machine Learning Technique (MLT) is commonly used for cancer

detection and diagnosis [8]. To ameliorate the efficiency and solve the deficiency of the CAD methods, the representation learning value was emphasized [9]. One of representation learning methods is Deep Learning (DL) that uses the hierarchical representations of image datasets as attributes. The key characteristic of DL is that it could encode it in higher level of function representation and consider the content without requiring post-processing [10].

This study presents a metaheuristic based hyperparameter optimization with deep learning-based breast cancer detection on mammogram images (MHODL-BCDMI) technique. Primarily, the MHODL-BCDMI technique employs pre-processing in two stages: Wiener Filter (WF) based noise elimination and contrast enhancement. Next, the MHODL-BCDMI technique exploits densely connected networks (DenseNet201) model for feature extraction purposes. For BC detection and classification, hybrid convolutional neural network with gated recurrent unit (HCNN-GRU) model is used. Meanwhile, three hyperparameter optimizers are employed namely cat swarm optimization (CSO), harmony search algorithm (HSA), and hybrid grey wolf whale optimization algorithm (HGWWOA). Atlas, U2Net segmentation approach is used for the classification of benign and malignant types of cancer. The experimental analysis of the MHODL-BCDMI method is tested on digital mammogram image dataset.

## II. RELATED WORK

Zahoor et al. [11] intend to inspect ways for preventing the disease in addition to offer novel approaches of classification for minimalizing the risk of BC. With the reduction of the false positive rate, the accuracy of CAD method is enhanced. For deep feature extraction, The Modified Entropy Whale Optimization Algorithm (MEWOA) was devised. Salama and Aly [12] introduced an innovative structure for classification breast and cancer image segmentation. For classifying Mammographic Image Analysis Society (MIAS), various methods namely MobileNetV2, ResNet50, VGG16, DenseNet121 and InceptionV3 approaches have been implemented, Digital Database for Screening Mammography (DDSM) as well as Curated Breast Imaging Subset of DDSM (CBIS-DDSM) into malignant and benign. In [13], the authors presented a novel method, implemented on the Mini-MIAS dataset they are an inbuilt feature extraction and pre-processing approach utilizing K-mean clustering to select Speed-Up Robust Feature (SURF). A novel layer executes testing of the Multiclass SVM (MSVM) and DNN.

Malebary and Hashmi [14] presented a BMC system i.e. Breast Mass Classification. It has developed architecture depends on an amalgamation of random forest (RF), LSTM of RNN, CNN, k- mean clustering, boosting methods for classifying the breast

mass into malignant, normal, and benign. With existing classification systems, the BMC system is compared. In [15], introduced a computational structure to diagnose BC with the use of a ResNet50- CNN for categorizing mammogram images. To categorize and train the INbreast data into malignant or benign groups, the structure uses TL from the pretrained ResNet50 CNN on ImageNet. This innovative method facilitated initial classification and diagnosis of benign and malignant BC. Altameem et al. [16] applied mammography pictures for identifying BC. Four mammography imaging datasets with malignant, normal, and benign pictures utilizing different deep CNN (VGG-11, Inception V4, DenseNet121, and ResNet-164) methods as base classifiers are utilized. This method uses an ensemble method where Gompertz function has been exploited to frame fuzzy rankings of base classifier approaches. In stage one, from mammogram the breast area is mined and small square patches are made. In the next phase, breast mass was classified and identified into BI-RADS classes.

## III. THE PROPOSED MODEL

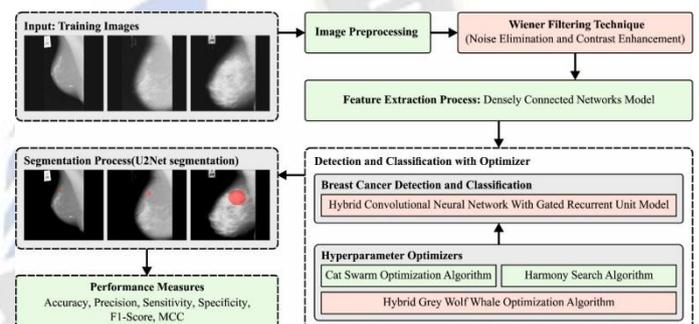


Fig 1: Overall flow of MHODL-BCDMI algorithm

In this study, we have focused on the design and development of the MHODL-BCDMI technique on the recognition and classification of BC on digital mammograms. The presented MHODL-BCDMI technique follows several stages of operations namely preprocessing, DenseNet201 based feature extraction, HCNN-GRU based classification, metaheuristics based hyperparameter tuning, and U2Net segmentation. Fig. 1 illustrates the overall flow of MHODL-BCDMI method.

### A. Image Pre-processing

Initially, the pre-processing of the digital mammograms takes place in two stages: WF based noise removal and contrast enhancement. The WF is a signal processing technique used for noise reduction in images [20]. It is based on the assumption that the noise and the image are statistically independent, and the filter tries to estimate the underlying signal by minimizing the mean square errors between the noisy image and the estimated signal. The WF is particularly effective in removing additive Gaussian noise from images. In addition, CLAHE is a widely used technique for contrast enhancement in digital

images. CLAHE improves the contrast of images by dividing the image into small, overlapping regions, and equalizing the histogram of each region separately. The contrast enhancement is limited to avoid over-amplification of noise or artefacts that might be present in the image.

**B. Feature Extraction**

To derive a relevant set of feature vectors, the DenseNet201 model is exploited in this study. The DenseNet201 exploits the condensed network to create model that is extremely parametrically effective and simple to train due to the potentiality of feature reuse by following layers [21], this improves the performance of the model and rises diversity of succeeding layer’s input. On numerous datasets, together with CIFAR-100 and ImageNet, the DenseNet201 demonstrated great outcomes. To improve connectivity in the DenseNet201, direct connection was made from every previous levels to each subsequent layer:

$$z^1 = H_1([Z^0, Z^1, \dots, z^{l-1}]) \tag{1}$$

Now,  $H_1$  denotes a non-linear conversion that might be expressed as a composite function comprising a convolution of linear unit function (3x3, B), and ReLU.  $[Z^0, Z^1, \dots, z^{l-1}]$  indicates the concatenation of feature map respective to layers 0 to  $-1$ . They can be fused into a single tensor for facilitating the application. Dense block is generated in the network structure and split by 1x1 convolutions, 2x2 average pooling and BN transition layer, for the downsampling purpose. The DenseNet201 growth rate, represented by the  $k$ -th hyperparameter, shows how the dense design generates new outcome. Because of its design that considers feature map as global state of network, DenseNet201 sufficiently function even with the low growth rate.

Every layer gives  $k$  feature map to the global state, where  $k$  is greater than the overall amount of input feature maps at  $l^{th}$  layer  $(FM)^l$  is evaluated by Eq. (2):

$$(FM)^l = k^0 + k(| - 1) \tag{2}$$

Now,  $k^0$  determines the input layer’s channel. Each 3x3 convolutional layer was preceded by the 1x1 convolutional layer that reduces the amount of input feature maps that is frequently greater than amount of output feature map,  $k$ . The added 1 x 1 convolutional layer produces 4k feature map and is represented as the bottleneck layer.

**C. Image Classification**

For BC detection and classification, the HCNN-GRU model is employed. CNN is a type of ANN that could well perform high-dimensional data [22]. It is widely employed in text

categorization, visual image, and video recognition. In the power system, there exist several smart devices and sensors. The spatiotemporal matrix was developed by using time sequence and the position of sensors to retain the spatial data in the power system.

This spatiotemporal matrix can be given by Eq. (3):

$$X = \begin{bmatrix} X_1(1) & X_1(2) & \dots & X_1(n) \\ X_2(1) & X_2(2) & \dots & X_2(n) \\ \vdots & \vdots & \ddots & \vdots \\ X_k(1) & X_k(2) & \dots & X_k(n) \end{bmatrix}, \tag{3}$$

Where  $X(n)$  characterizes the data recorded by  $k^{th}$  smart sensor at  $n$  time,  $k$  signifies the  $k^{th}$  smart sensor and  $n$  characterizes the  $n^{th}$  time sequence. CNN was used for processing the spatiotemporal matrix to extract the load features from the spatiotemporal matrix.

Initially, 2D spatiotemporal matrixes are stacked into 3D matrix blocks, later they are exploited with the convolution function. The objective of convolution operation is to obtain an extremely abstract feature, then the output of convolution function is applied to pooling function. The pooling function doesn’t change the input matrix depth, but it could decrease the number of nodes and the size of matrixes, to decrease the parameter in the NN. Afterwards repeated pooling and convolution operations, the extremely abstract feature was attained and flattened to 1D vector; hence it is interconnected with the FC layer. Next, the weight and bias vector parameters of the FC layer are iteratively calculated. Lastly, prediction outcomes are attained by the output of activation function.

The architecture of GRU-CNN hybrid NN includes a CNN and GRU modules. The input is the data of time series and spatiotemporal matrix gathered from the power system; the output is the prediction of upcoming load value. With that regard, it is better at processing 2D data like spatiotemporal images and matrixes. The CNN architecture exploits local connection and shared weight to directly extract local features from the spatiotemporal matrix data and attain efficient representation via the pooling and convolution layers. The architecture of CNN module has a flatten operation and two convolution layers, and every convolution layer has a pooling and convolution operation. After, the high-dimensional data are flattened into 1D data, and the output of CNN model was interconnected to the FC layer. In contrast, the objective of the GRU model is to capture the long-term dependency and the GRU learns relevant data in the past data for a longer period, and the useless data would be forgotten by forget gate. The input of GRU model was time sequence data; the GRU model has

several gated recurrent units, and the output of this gated recurrent unit is interconnected to the FC layer.

#### D. Hyperparameter Tuning Process

In order to improve the BC detection results of the HCNN-GRU method, three metaheuristic optimizers are used namely CSO, HSA, and HGWWOA. The presented method not only derives a fitness function to obtain better performance of classification and also defines a positive integer to describe the superior outcome of the candidate solution. Here, the decline of the classification error rate is treated as a fitness function.

$$\begin{aligned} \text{fitness}(x_i) &= \text{ClassifierErrorRate}(x_i) \\ &= \frac{\text{number of misclassified samples}}{\text{Total number of samples}} * 100 \end{aligned} \quad (4)$$

##### i. CSO Algorithm

CSO algorithm is based on real time behaviour of cats at active mode and taking rest at certain point in time [23]. Usually, cats are in two modes namely seeking and tracing modes. In seeking mode, cats are in rest but they are alerted always, and other mode called tracing mode where cats chase the target. This procedure is constantly performed until each iteration was completed. Initially, cats were initialized randomly over multiple dimensions. In such cases, consider  $n$  dimensions because task was randomly generated and they are classified into seeking and tracing modes according to the population. For all the iterations, solution is evaluated for cats and later fitness was evaluated for the cat, till better solution is accomplished. In tracing mode, cat has initialized with certain velocity with dimension, hence new velocity is evaluated by Eq. (5).

$$Vel_i^d(t+1) = p * Vel_i^d(t) + c * r * (x_d^{best} - x_i^d) \quad (5)$$

Where  $x_i^d$  represent the location of the cat at  $ith$  iteration.  $Vel_i^d(t)$  denotes velocity of  $ith$  cat at  $tth$  iteration,  $c$  denotes a constant and  $r$  shows the random integer between zero and one.  $x_d^{best}$  indicates the cat's global better location.

$$x_i^d(t+1) = x_i^d + Vel_i^d(t) + 1 \quad (6)$$

This calculation should be performed until cats reach target with better value and after fitness calculation, update solution with non-dominated cat.

##### ii. HSA

The mathematical formulation of HSA is constructed in five different stages and is explained in the following [24].

In the first phase, the initialization process is performed where the initial value for the harmony vectors ( $X_i$ ) comprising related objective function amounts ( $F_i$ ) and decision variable ( $X_i = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ ) are defined. The decision variable demonstrates various musicians, and the objective function evaluation demonstrates harmony which the musician achieves. Here, the crucial parameters of the HSA include Harmony Memory Considering Rate (HMCR), Harmony Memory Size (HMS), the termination criteria, and Pitch Adjusting Rate (PAR), which is considered as the maximal iteration counts (MaxIter) are defined.

In the second phase, the Harmony Memory (HM) is defined involving the solution vector randomly generated with harmony memory (HMS) size, categorized based on the values of objective function. This can be mathematically expressed as follows:

$$HM = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^{HMS} \end{bmatrix} \quad (7)$$

In the third stage, a New Harmony vector ( $X'_i = \{x'_1, x'_2, \dots, x'_i, \dots, x'_n\}$ ) is improvised from the initial harmony or HM vectors according to the randomization, pitch adjustment, and memory consideration procedure. The decision variable is defined by selecting values from HM or selecting from initial harmony vector. With that regard, a distributed uniformly random integer within (0,1) is generated to decide between two choices. When the random integer was greater than the formerly defined HMCR, then new harmony vector is chosen from the HM, whereas for the randomly generated number lesser than HMCR, the new vector is defined for forming the initial harmony vector ( $X_i$ ). This aspect was mathematically given as follows:

$$x'_i \rightarrow \begin{cases} x'_i e_{HM} & \text{with the probability of } (HMCR) \\ x_i e_{X_i} & \text{with the probability of } (1 - HMCR) \end{cases} \quad (8)$$

PAR is applied to mathematically formulate the mutation process for the value attained from the HM by generating another random number equally spread in the interval (0,1). Once the

generated random value is greater than the formerly defined PAR, then New Harmony vector chosen from the HM chooses adjacent values with PAR probability; but, no PAR is made if the created random integer is lesser than the PAR. This consideration was mathematically formulated in the following:

$$\rightarrow \begin{cases} x'_i + (bw \times u) & \text{with the probability of (PAR)} \\ x_i & \text{with the probability of (1 - PAR)} \end{cases} \quad (9)$$

In Eq. (9),  $bw$  denotes the arbitrary distance bandwidth and  $u$  indicates the uniform distribution integer within  $(-1,1)$ .

In the fourth stage, upgrade the HM, and if the recently made harmony vector outperformed the worst harmony in the HM regarded objective function values, then the new harmony can be replaced by worst, also the HM is arranged by the value of objective function. The 3<sup>rd</sup> and 4<sup>th</sup> stages are reiterated in the 5<sup>th</sup> step until the ending condition is satisfied.

### iii. HGWWOA

GWO is stimulated by the leadership hierarchy and hunting strategy of gray wolves [25]. The mathematical model includes search for prey, social hierarchy, attacking prey, hunting, and enriching prey and defined below:

Social hierarchy: The leader wolf is named alpha ( $\alpha$ ). The 2<sup>nd</sup> and 3<sup>rd</sup> levels in the group are beta ( $\beta$ ) and delta ( $\delta$ ).

Encircling prey: the gray wolf encircles prey for chasing that is arithmetically expressed as follows:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(n) - \vec{X}(n)| \quad (10)$$

$$\vec{X}(n+1) = \vec{X}(n) - \vec{A} \cdot \vec{D} \quad (11)$$

Where  $n$  designates the existing iteration, and  $\vec{X}$  represents the location vector of gray wolf.  $\vec{A}$  and  $\vec{C}$  vector is defined by the following expression:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (12)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (13)$$

Where  $r_1$  and  $r_2$  denotes the random number within  $[0,1]$  and was decreased linearly from 2 to 0.

Hunting: The hunting strategy is mathematically expressed as follows:

$$\vec{D}_\alpha = |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}|, \vec{D}_\beta = |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}|, \vec{D}_\delta = |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}| \quad (14)$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_3 \cdot \vec{D}_\alpha, \vec{X}_2 = \vec{X}_\beta - \vec{A}_3 \cdot \vec{D}_\beta, \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta \quad (15)$$

$$\vec{X}(n+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (16)$$

Attacking prey: If  $|A| < 1$  the wolf moves to prey to attack.

Search for prey: If  $|A| > 1$  wolf move away from prey to search for best prey.

WOA is derived from social behaviors of humpback whales because they have special hunting strategy named bubble-net foraging approach. This technique can be accomplished by generating distinct bubbles along a 9<sup>th</sup> shaped or circle path. The arithmetical expression of finding the prey, prey encircling, and spiral bubble-net feeding maneuver are defined in the following:

- Encircling prey: Humpback whale encloses prey and changes the location towards the better solution during iterations. These behaviors are mathematically formulated by (10) and (11).
- -Bubble-net attacking approach

The humpback whale attacks the prey using bubble-net approach that is given in the following:

1. Shrinking encircling system: This procedure could be performed by minimizing the value.
2. Spiral updating location: The helix-shaped movement can be arithmetically defined below equation:

$$\vec{X}(n+1) = \vec{D}' \cdot e^{bv} \cdot \cos(2\pi v) + \vec{X}^*(n) \quad (17)$$

$$\vec{D}' = |\vec{X}^*(n) - \vec{X}(n)| \quad (18)$$

Where  $\vec{X}$  denotes the location vector.  $v$  shows the random integer within  $[-1]$ ,  $b$  denotes constant to define the shape of logarithmic spiral,  $\vec{X}^*$  represent the location vector of prey, and The humpback whales swim round the prey within the spiral-shaped path and shrinking circle:

$$\vec{X}(n+1) = \begin{cases} \vec{X}(n) - \vec{A} \cdot \vec{D} & \text{if } p < 0.5 \\ \vec{D}' \cdot e^{bv} \cdot \cos(2\pi v) + \vec{X}^*(n) & \text{if } p > 0.5 \end{cases} \quad (19)$$

In Eq. (19),  $p$  denotes the randomly generated value within  $[0,1]$ .

Search for prey: The humpback whales search arbitrarily for prey. The vector  $A$  can utilize for searching for prey when  $|A| > 1$ . This method is mathematically expressed as:

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \quad (20)$$

$$\vec{X}(n+1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \quad (21)$$

In this study, a novel hybrid is recommended to enhance the performance. The suggested technique uses the leadership hierarchy to adopt the bubble-net attacking approach. In the exploitation stage, the presented method chooses three optimal candidate solutions (alpha ( $\alpha$ ), beta ( $\beta$ ) and delta ( $\delta$ )) from entire searching agents and other search agent would alter their locations as per the location of the optimal search agents.

As discussed above, humpback whales swim round prey within two systems. The presented mathematical model to upgrade whales location with the use of the leadership hierarchy. The upgrading location of humpback whales with a spiral-shaped path and be expressed below:

$$\vec{D}'_{\alpha} = |\vec{X}_{\alpha}(n) - \vec{X}|, \vec{D}'_{\beta} = |\vec{X}_{\beta}(n) - \vec{X}(n)|, \vec{D}'_{\delta} = |\vec{X}_{\delta}(n) - \vec{X}(n)| \quad (22)$$

$$\vec{X}_1(n) = \vec{X}_{\alpha}(n) + \vec{D}'_{\alpha} \cdot e^{bv} \cdot \cos(2\pi v),$$

$$\vec{X}_2(n) = \vec{X}_{\beta}(n) + \vec{D}'_{\beta} \cdot e^{bv} \cdot \cos(2\pi v),$$

$$\vec{X}_3 = \vec{X}_{\delta}(n) + \vec{D}'_{\delta} \cdot e^{bv} \cdot \cos(2\pi v) \quad (23)$$

$$\vec{X}(n+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (24)$$

#### E. U2Net Segmentation Process

Finally, the U2Net model is applied to segment the BC. The U2-net is a two-stage nested U-structure [26]. The outer layer is a larger U-structure including 11 phase. All the phases are populated by the residual U-block (RSU). In theory, nested U-structure enabled the abstraction of multi-level and multi-scale features. It includes 3 major parts: (1) map fusions, (2) encoding, and (3) decoding stages, as given as follows below.

(1) There are six stages in the encoder. Every stage was encompassed by RSU. In the RSU of the first 4 phases, the feature map is minimized for attaining more large-scale data and rising the receptive field. Then, dilated convolution is used for the replacement of the pooling process. This stage is needed to avoid context data loss..

(2) The decoder phases have same structure as encoding phase. All the decoding stages concatenate the up-sampled feature maps from its prior phase.

(3) Feature map fusion with deep supervision method is the last stage used for generating a probability map. Then, the output can be up-sampled to the input image size and combined with the concatenation function.

The U2-net design has low computing and memory costs along with deep architecture with rich multiscale features. Furthermore, as the U2-net structure depends only on RSU blocks and doesn't exploit any pre-trained backbone, it becomes flexible and easier to adapt to working environment.

## IV. RESULTS AND DISCUSSION

### A. Simulation Environment

In this section, the BC detection results of the MHODL-BCDMI approach are tested on the MIAS-mammography dataset [27]. It holds 319 samples with three class labels as defined in Table 1. Fig. 2 illustrates the sample images. Fig. 3 represents the original and segmented images.

Table 1 Details of database

Classes	No. of Instances
Normal	207
Benign	63
Malignant	49
<b>Total Number of Samples</b>	<b>319</b>

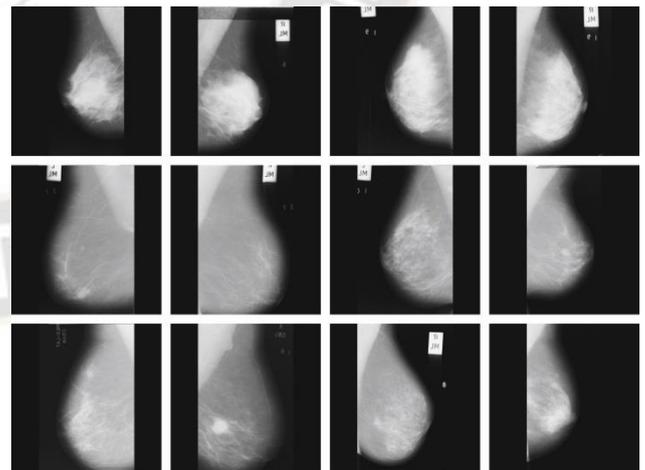


Fig. 2. Sample Images

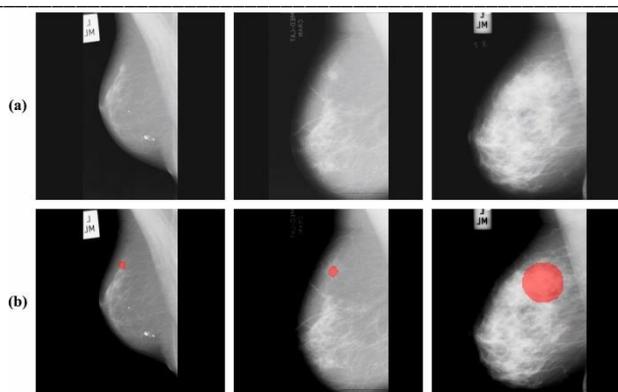


Fig. 3. a) Original Images b) Segmented Images

Fig. 4 portrayed the classifier results of the MHODL-BCDMI method under training set. Fig. 4a depicts the confusion matrix offered by the MHODL-BCDMI technique on 70% of TRS. The figure denoted that the MHODL-BCDMI method has identified 142 instances under normal, 41 under benign, and 35 instances under malignant. Similarly, Fig. 4b demonstrates the PR analysis of the MHODL-BCDMI approach. The figures specify that the MHODL-BCDMI method has obtained maximum PR performance under 3 classes. Finally, Fig. 4c illustrates the ROC investigation of the MHODL-BCDMI model. The figure depicted that the MHODL-BCDMI method has productive outcomes with higher ROC values under 3 class labels.

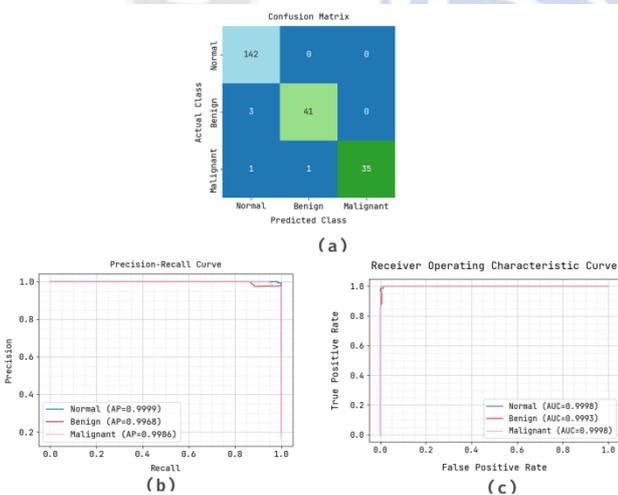


Fig. 4. Proposed Model Results on Training Set a) Confusion Matrix b) PR-Curve c) ROC Analysis

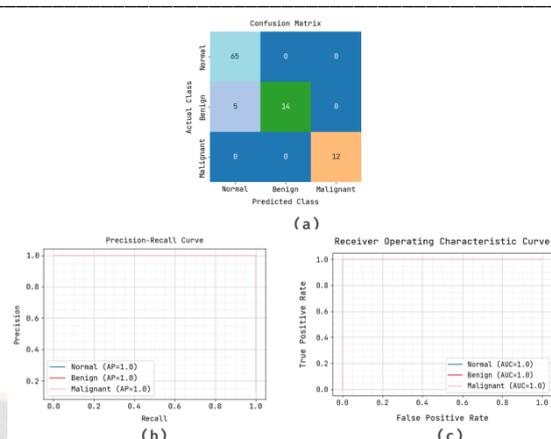


Fig. 5. Proposed Model Results on Testing Set a) Confusion Matrix b) PR-Curve c) ROC Analysis

Fig. 5 demonstrates the classifier results of the MHODL-BCDMI technique under testing set. Fig. 5a depicts the confusion matrix offered by the MHODL-BCDMI methodology on 30% of TSS. The figure denoted that the MHODL-BCDMI method has detected 65 instances under normal, 14 under benign, and 12 instances under malignant. Similarly, Fig. 5b demonstrates the PR analysis of the MHODL-BCDMI algorithm. The figures noted that the MHODL-BCDMI method has obtained highest PR performance under 3 classes. Finally, Fig. 5c illustrates the ROC investigation of the MHODL-BCDMI model. The figure represented that the MHODL-BCDMI technique has productive outcomes with greater ROC values under 3 class labels.

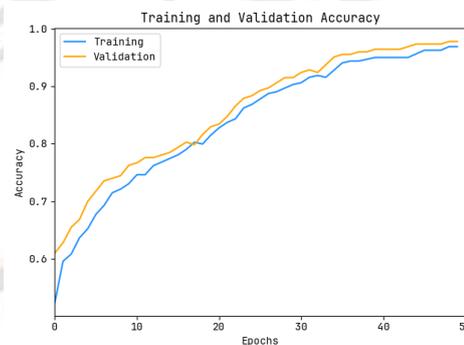


Fig. 6. Accuracy curve of the MHODL-BCDMI approach

Fig. 6 examines the accuracy of the MHODL-BCDMI technique during the training and validation process on test dataset. The figure notifies that the MHODL-BCDMI technique reaches increasing accuracy values over increasing epochs. In addition, the increasing validation accuracy over training accuracy exhibits that the MHODL-BCDMI technique learns efficiently on the test dataset.

The loss analysis of the MHODL-BCDMI method at the time of training and validation is demonstrated on the test dataset in Fig. 7. The outcomes indicate that the MHODL-BCDMI

approach reaches closer values of training and validation loss. The MHODL-BCDMI method learns efficiently on the test dataset.

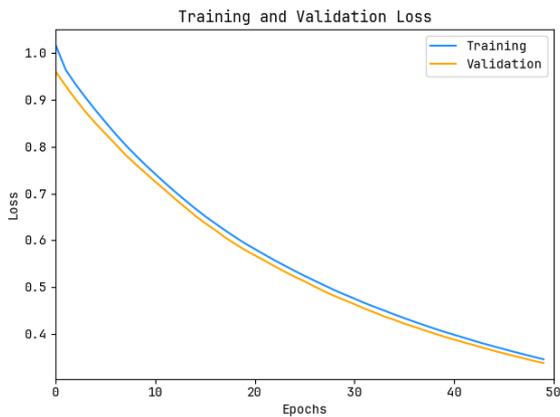


Fig. 7. Loss curve of the MHODL-BCDMI approach

In Table 2 and Fig. 8, the overall outcomes of the proposed method with different metaheuristic optimizers were given. Using HSA model on TRS, the HCNN-GRU classifier obtains  $accu_y$ ,  $prec_n$ ,  $sens_y$ ,  $spec_y$ ,  $F_{score}$ , and MCC of 73.54%, 78.09%, 53.10%, 77.11%, 54.64%, and 43.74% respectively. Concurrently, using HSA method on TSS, the HCNN-GRU classifier gains  $accu_y$ ,  $prec_n$ ,  $sens_y$ ,  $spec_y$ ,  $F_{score}$ , and MCC of 67.71%, 56.30%, 39.10%, 68.67%, 38.17%, and 14.83% respectively. Meanwhile, using CSO method on TRS, the HCNN-GRU classifier acquires  $accu_y$ ,  $prec_n$ ,  $sens_y$ ,  $spec_y$ ,  $F_{score}$ , and MCC of 83.86%, 92.17%, 71.15%, 85.42%, 76.87%, and 69.52% correspondingly. Similarly, using CSO method on TSS, the HCNN-GRU classifier obtains  $accu_y$ ,  $prec_n$ ,  $sens_y$ ,  $spec_y$ ,  $F_{score}$ , and MCC of 84.38%, 93.75%, 72.66%, 83.87%, 75.66%, and 69.04% correspondingly. Likewise, using HGWWOA model on TRS, the HCNN-GRU classifier obtains  $accu_y$ ,  $prec_n$ ,  $sens_y$ ,  $spec_y$ ,  $F_{score}$ , and MCC of 97.76%, 98.29%, 95.93%, 98.17%, 97.06%, and 95.72% respectively. Finally, using HGWWOA model on TSS, the HCNN-GRU classifier obtains  $accu_y$ ,  $prec_n$ ,  $sens_y$ ,  $spec_y$ ,  $F_{score}$ , and MCC of 94.79%, 97.62%, 91.23%, 94.62%, 93.71%, and 90.48% respectively.

Table 2 Classifier outcome of proposed method with different metaheuristic optimizers

Harmony Search Algorithm		
Metrics	Training Set	Testing Set
Accuracy	73.54	67.71
Precision	78.09	56.30
Sensitivity	53.10	39.10
Specificity	77.11	68.67

F1-Score	54.64	38.17
MCC	43.74	14.83
Cat Swarm Optimization		
Metrics	Training Set	Testing Set
Accuracy	83.86	84.38
Precision	92.17	93.75
Sensitivity	71.15	72.66
Specificity	85.42	83.87
F1-Score	76.87	75.66
MCC	69.52	69.04
Hybrid Grey Wolf whale Optimization Algorithm		
Metrics	Training Set	Testing Set
Accuracy	97.76	94.79
Precision	98.29	97.62
Sensitivity	95.93	91.23
Specificity	98.17	94.62
F1-Score	97.06	93.71
MCC	95.72	90.48

In Table 3, a detailed comparative study of the MHODL-BCDMI technique is clearly reported [28-30]. Fig. 9 investigates the results of the MHODL-BCDMI technique with recent method in terms of  $accu_y$ . The outcomes show that the MHODL-BCDMI technique reaches better performance.

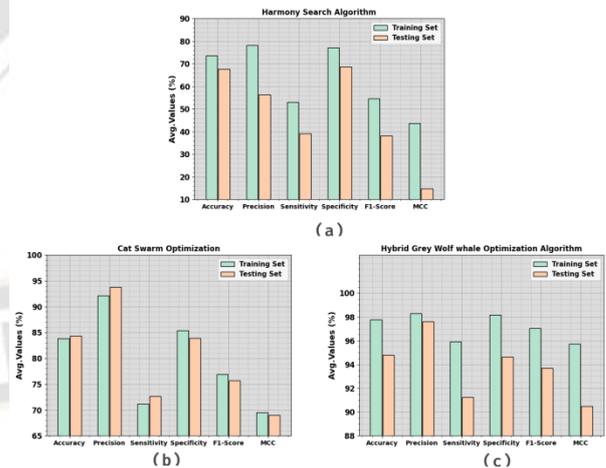


Fig. 8. Classifier outcome of proposed method (a) HAS, (b) CSO, and (c) HGWWOA

Based on  $accu_y$ , the MHODL-BCDMI technique attains improving  $accu_y$  of 97.76% while the CSO-HCNN-GRU, HSA-HCNN-GRU, KNN-XGBoost, ML-ANN, CNN-GRU, and CNN-LSTM models obtain decreasing  $accu_y$  of 84.38%, 73.54%, 95.06%, 95.49%, 95.62%, and 94.96% respectively.

Table 3 Comparative outcome of MHODL-BCDMI method with recent techniques

Methods	$Sens_y$	$Spec_y$	$Accu_y$
MHODL-BCDMI	95.93	98.17	97.76
CSO-HCNN-GRU	72.66	83.87	84.38
HSA-HCNN-GRU	39.10	68.67	73.54
KNN-XGBoost	95.60	96.15	95.06
ML-ANN	97.26	94.57	95.49
CNN-GRU	95.12	95.81	95.62
CNN-LSTM	95.49	94.71	94.96

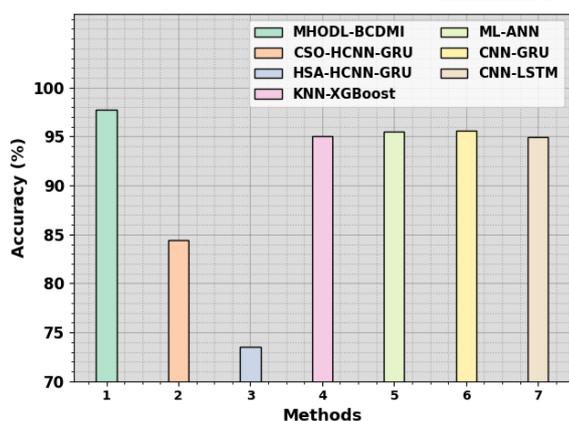


Fig. 9.  $Accu_y$  outcome of MHODL-BCDMI approach with recent algorithms

Fig. 10 inspects the results of the MHODL-BCDMI method with recent techniques in terms of  $sens_y$  and  $spec_y$ . The figure shows that the MHODL-BCDMI approach reaches better performance. Based on  $sens_y$ , the MHODL-BCDMI algorithm gains improving  $sens_y$  of 95.93% while the CSO-HCNN-GRU, HSA-HCNN-GRU, KNN-XGBoost, ML-ANN, CNN-GRU, and CNN-LSTM approaches obtain decreasing  $sens_y$  of 72.66%, 39.10%, 95.60%, 97.26%, 95.12%, and 95.49% correspondingly. Eventually, based on  $spec_y$ , the MHODL-BCDMI method attains improving  $spec_y$  of 98.17% while the CSO-HCNN-GRU, HSA-HCNN-GRU, KNN-XGBoost, ML-ANN, CNN-GRU, and CNN-LSTM methodology obtain decreasing  $spec_y$  of 83.87%, 68.67%, 96.15%, 94.57%, 95.81%, and 94.71% correspondingly.

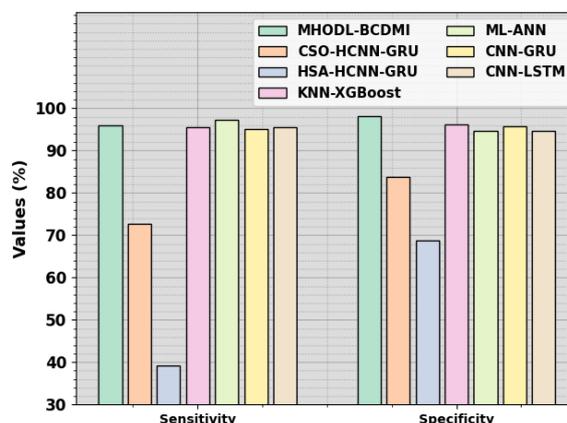


Fig. 10.  $Sens_y$  and  $spec_y$  outcome of MHODL-BCDMI approach with recent algorithms

In Fig. 11, the segmentation results of the MHODL-BCDMI approach is demonstrated. The results indicated that the MHODL-BCDMI technique reaches higher accuracy and IoU values with a rise in epochs. At the same time, the MHODL-BCDMI method attains decreasing loss values over an increase in epochs.

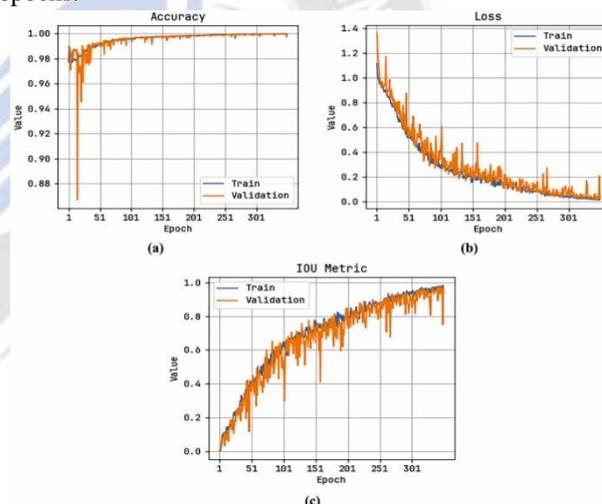


Fig. 11. Analysis of Segmentation a) Accuracy b) Loss c) IoU

These outcomes shown the better performance of the MHODL-BCDMI method over other compared methods.

## V. CONCLUSION

In this study, we have focused on the design and development of the MHODL-BCDMI technique on the recognition and classification of BC on the digital mammograms. The presented MHODL-BCDMI technique follows several stages of operations namely preprocessing, DenseNet201 based feature extraction, HCNN-GRU based classification, metaheuristics based hyperparameter tuning, and U2Net segmentation. Furthermore, three hyperparameter optimizers are employed namely CSO, HSA, and HGWWOA. Lastly, U2Net segmentation approach is used for the classification of benign and

malignant types of cancer. The experimental validation of the MHODL-BCDMI method is tested on digital mammogram image dataset and the outcomes are assessed in terms of diverse metrics. The simulation results highlighted the enhanced cancer detection performance of the MHODL-BCDMI technique over other recent approaches. Thus, the MHODL-BCDMI method can be exploited for automated BC detection process on digital mammograms. In future, the classification performance of MHODL-BCDMI technique can be enhanced by ensemble classifier models.

## REFERENCES

- [1] Ragab, D.A., Attallah, O., Sharkas, M., Ren, J. and Marshall, S., 2021. A framework for breast cancer classification using multi-DCNNs. *Computers in Biology and Medicine*, 131, p.104245.
- [2] Chakravarthy, S.S. and Rajaguru, H., 2022. Automatic detection and classification of mammograms using improved extreme learning machine with deep learning. *Irbm*, 43(1), pp.49-61.
- [3] Saber, A., Sakr, M., Abo-Seida, O.M., Keshk, A. and Chen, H., 2021. A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. *IEEE Access*, 9, pp.71194-71209.
- [4] Fathy, W.E. and Ghoneim, A.S., 2019. A deep learning approach for breast cancer mass detection. *International Journal of Advanced Computer Science and Applications*, 10(1).
- [5] Singla, C., Sarangi, P.K., Sahoo, A.K. and Singh, P.K., 2022. Deep learning enhancement on mammogram images for breast cancer detection. *Materials Today: Proceedings*, 49, pp.3098-3104.
- [6] Suresh, R., Rao, A.N. and Reddy, B.E., 2019. Detection and classification of normal and abnormal patterns in mammograms using deep neural network. *Concurrency and Computation: Practice and Experience*, 31(14), p.e5293.
- [7] Darweesh, M.S., Adel, M., Anwar, A., Farag, O., Kotb, A., Adel, M., Tawfik, A. and Mostafa, H., 2021. Early breast cancer diagnostics based on hierarchical machine learning classification for mammography images. *Cogent Engineering*, 8(1), p.1968324.
- [8] ADEDIGBA, A.P., ADESHINAT, S.A. and AIBINU, A.M., 2019, December. Deep learning-based mammogram classification using small dataset. In 2019 15th International Conference on Electronics, Computer and Computation (ICECCO) (pp. 1-6). IEEE.
- [9] Shakeel, S. and Raja, G., 2021, January. Classification of Breast Cancer from Mammogram images using Deep Convolution Neural Networks. In 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST) (pp. 595-599). IEEE.
- [10] Sha, Z., Hu, L. and Rouyendegh, B.D., 2020. Deep learning and optimization algorithms for automatic breast cancer detection. *International Journal of Imaging Systems and Technology*, 30(2), pp.495-506.
- [11] Zahoor, S., Shoaib, U. and Lali, I.U., 2022. Breast cancer mammograms classification using deep neural network and entropy-controlled whale optimization algorithm. *Diagnostics*, 12(2), p.557.
- [12] Salama, W.M. and Aly, M.H., 2021. Deep learning in mammography images segmentation and classification: Automated CNN approach. *Alexandria Engineering Journal*, 60(5), pp.4701-4709.
- [13] Kaur, P., Singh, G. and Kaur, P., 2019. Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification. *Informatics in Medicine Unlocked*, 16, p.100151.
- [14] Malebary, S.J. and Hashmi, A., 2021. Automated breast mass classification system using deep learning and ensemble learning in digital mammogram. *IEEE Access*, 9, pp.55312-55328.
- [15] Rahman, H., Naik Bukht, T.F., Ahmad, R., Almadhor, A. and Javed, A.R., 2023. Efficient Breast Cancer Diagnosis from Complex Mammographic Images Using Deep Convolutional Neural Network. *Computational intelligence and neuroscience*, 2023.
- [16] Castro, J., González, J., Rodríguez, C., Garcia, J., & Jóhann, Þorvaldsson. Predicting Student Dropout Rates in Engineering Programs with Machine Learning. *Kuwait Journal of Machine Learning*, 1(1). Retrieved from <http://kuwaitjournals.com/index.php/kjml/article/view/105>
- [17] Altameem, A., Mahanty, C., Poonia, R.C., Saudagar, A.K.J. and Kumar, R., 2022. Breast cancer detection in mammography images using deep convolutional neural networks and fuzzy ensemble modeling techniques. *Diagnostics*, 12(8), p.1812.
- [18] Patel, J.J. and Hadia, S.K., 2021. An enhancement of mammogram images for breast cancer classification using artificial neural networks. *IAES International Journal of Artificial Intelligence*, 10(2), p.332.
- [19] Jabeen, K., Khan, M.A., Balili, J., Alhaisoni, M., Almujaally, N.A., Alrashidi, H., Tariq, U. and Cha, J.H., 2023. BC2NetRF: Breast Cancer Classification from Mammogram Images Using Enhanced Deep Learning Features and Equilibrium-Jaya Controlled Regula Falsi-Based Features Selection. *Diagnostics*, 13(7), p.1238.
- [20] Ms. Nora Zilam Runera. (2014). Performance Analysis On Knowledge Management System on Project Management. *International Journal of New Practices in Management and Engineering*, 3(02), 08 - 13. Retrieved from <http://ijnpme.org/index.php/IJNPME/article/view/28>
- [21] Ibrokhimov, B. and Kang, J.Y., 2022. Two-Stage Deep Learning Method for Breast Cancer Detection Using High-Resolution Mammogram Images. *Applied Sciences*, 12(9), p.4616.
- [22] Ferzo, B.M. and Mustafa, F.M., 2020, April. Image Denoising in Wavelet Domain Based on Thresholding with Applying Wiener Filter. In 2020 International Conference on Computer Science and Software Engineering (CSASE) (pp. 106-111). IEEE.

- [23] Qadir, A.M. and Abd, D.F., 2023. Kidney Diseases Classification using Hybrid Transfer-Learning DenseNet201-Based and Random Forest Classifier. *Kurdistan Journal of Applied Research*, pp.131-144.
- [24] Praveena, N., Juneja, K., Rashid, M., Almagrabi, A.O., Sekaran, K., Ramalingam, R. and Usman, M., 2023. Hybrid gated recurrent unit and convolutional neural network-based deep learning mechanism for efficient shilling attack detection in social networks. *Computers and Electrical Engineering*, 108, p.108673.
- [25] Mangalampalli, S., Swain, S.K. and Mangalampalli, V.K., 2022. Multi objective task scheduling in cloud computing using cat swarm optimization algorithm. *Arabian Journal for Science and Engineering*, 47(2), pp.1821-1830.
- [26] Himanshu Sharma, Vijay Kumar Joshi. (2023). An Efficient Load Balancing Approach For Resource Utilizations In Green Cloud Computing. *International Journal of Intelligent Systems and Applications in Engineering*, 11(2s), 395–404. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/2735>
- [27] Azizi, M., Talatahari, S., Basiri, M. and Shishehgharkhaneh, M.B., 2022. Optimal design of low-and high-rise building structures by Tribe-Harmony Search algorithm. *Decision Analytics Journal*, 3, p.100067.
- [28] Korashy, A., Kamel, S., Jurado, F. and Youssef, A.R., 2019. Hybrid whale optimization algorithm and grey wolf optimizer algorithm for optimal coordination of direction overcurrent relays. *Electric Power Components and Systems*, 47(6-7), pp.644-658
- [29] Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O.R. and Jagersand, M., 2020. U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern recognition*, 106, p.107404.
- [30] <https://www.kaggle.com/datasets/kmader/mias-mammography>
- [31] Ahmad, S., Ullah, T., Ahmad, I., Al-Sharabi, A., Ullah, K., Khan, R.A., Rasheed, S., Ullah, I., Uddin, M. and Ali, M., 2022. A novel hybrid deep learning model for metastatic cancer detection. *Computational Intelligence and Neuroscience*, 2022.
- [32] Han L and Yin Z (2022) A hybrid breast cancer classification algorithm based on meta-learning and artificial neural networks. *Front. Oncol.* 12:1042964. doi: 10.3389/fonc.2022.1042964
- [33] Raaj, R.S., 2023. Breast cancer detection and diagnosis using hybrid deep learning architecture. *Biomedical Signal Processing and Control*, 82, p.104558.