JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Evolutionary Optimization with Deep Convolutional Neural Network based Breast Cancer Detection and Classification

R. Gurumoorthy[1], Dr. M. Kamarasan[2]

[1]Research Scholar, Department of Computer and Information Science, Annamalai University, Annamalainagar, India.

[2]Assistant Professor, Department of Computer and Information Science, Annamalai University, Annamalainagar, India.

Abstract

Breast cancer (BC) is a common worldwide health issue, with its earlier detection performing an essential function to improve outcomes for patients. The crucial requirement for advanced BC detection and classification approaches stems from the disease's occurrence and possibility for early involvement for significantly enhancing survival rates. Standard diagnostic methods, but valuable, face limitations for efficiency and accuracy. Accordingly, there is a developing imperative to tackle cutting-edge technologies, like artificial intelligence (AI) and machine learning (ML), to refine and increase the diagnostic process of BC. This research offers a new moth flame optimization employing breast cancer detection and classification (MFO-BCDC) algorithm. This research work begins with the application of DCNN, a robustness deep learning (DL) model for image analysis. Convolutional layers in the network extract related features from mammographic images, offering a robust foundation for subsequent classification. Moth Flame Optimization (MFO) has been presented as a meta-heuristic algorithm for hyperparameter tuning of the DCNN. MFO, stimulated by the natural behavior of moths, efficiently analysis the hyperparameter space for enriching the network's performance. To complement the DL method, the study integrates the XGBoost method for classification. XGBoost, known for its ensemble learning abilities can be utilized to effectively handle the extracted features and determine prediction accuracy. The developed combined architecture was assessed employing benchmark BC datasets, and wide-ranging experiments were executed to analysis the effectiveness of all modules. Outcomes represent that the incorporation of DCNN with feature extraction, MFO for hyperparameter tuning, and XGBoost for classification noticeably increases the performances with different features.

Keywords: Breast cancer; Moth Flame Optimization; Deep Learning; Feature extraction; CAD; Histopathological Image

1. Introduction

In present scenario, entire world suffering from deadly illnesses. According to the report of WHO, cancer is one of the foremost reasons of death [1]. Particularly, in undeveloped countries female breast cancer (BC) is higher when compared to developed ones. For instance, rate of BC analysed annually is 1.38 million patients among them one-third of them die in Pakistan [2]. Generally, due to cancer the death rate is nearly 9.6 million by an existence rate of 1.7 million. Therefore, it is essential to identify tumor as well as begin the treatment at an initial stage; or else cancer might spread and affect the entire breast or other parts in body [3]. Depend up on precise and effectual analysis, a right primary treatment may enhance endurance rate up to 80 percentage in BC. The lesions affected by BC are classified as two kinds such as malignant and benign. Whereas, both sorts can be cancerous or noncancerous. For instance, in epithelial cells, irregularities are owing to benign lesions, but incapable of developing further and not lead to BC [4]. Next, malignant cells are considered as very risky due to their irregular development in body and viewed as tumorous cells. In microscopic images, it is very challenging task to analyse properly as well as categorize malignant and benign lesions [5].

By pattern detection and machine learning (ML), a worthy contract of hand-crafted or engineered features based on researchers have been developed for detecting BC histology images [6]. Feature extraction is basic procedure that can be employed to maximize classification accurateness by reducing the quantity of designated features in image classification [7]. Deep learning (DL) techniques comes will high power to mechanically remove features, recover data, and take up-to-date knowledgeable pictures of data. So, they can able to resolve issues of common feature removal models [8]. The automated detection of BC histopathological images is very significant challenge in CAD (Computer-Aided Detection/Diagnosis) systems [9]. DL techniques plays an extraordinary role by classifying, identifying and dividing prime BC histopathological images. Numerous researchers invested significant efforts in increasing strong computer-assisted utensils for the detection of BC histopathological images by employing DL [10]. In this study, most popular DL techniques projected in literature are mainly depends on CNNs.

Umer et al. [11] projected a DL-based method that employed a huge openly accessible breast HI datasets. Primarily, a DL technique has been developed for feature extraction. Secondarily, the removed feature vector can be accepted to this developed new feature selection (FS) method for better FS. Lastly, for the classification of BC into invasive ductal carcinoma (IDC) as well as normal categorical various ML methods could be utilized. Elmannai et al. [12] introduced an incorporated a 2 deep-CNNs (DCNNs) for extracting distinguished image features employing TL. The pre-trained Xceptions and Inception the techniques have been exploited in parallel. Next, the feature maps must be integrated and decreased by dropout previously provided to the final fully connected (FC) layers for categorization. This analysis gives a sub-image classification after that, a whole image classification depends upon maximum probability and majority vote rules. A 4 tissue malignancy levels have been deliberated.

In [13], a CAD technique dependent upon DL was metered. To this aim, a 5 pre-trained CNN methods have been examined and tested—DenseNet201, InceptionResNetV2, Xception, VGG19, and ResNet152 by employing data augmentation approaches, and a novel method was presented for TL. These techniques could be examined and trained with HI acquired from the BreakHis database. Sahu et al. [14] developed a computer aided ensemble technique for analysing of CE employing a ReNet18 and SVM but, pre-trained ReNet18 framework was for extracting the features from the X-ray images as well as SVM was utilized to identifying the cancer. For enriched the effectiveness, haze reduction was implemented to boost the quality of images and then cancerous segmentation for separating the cancerous regions from the image by utilizing histogram-related K-means approach.

Attallah et al. [15] designed an innovative CADx technique called as Histo-CADx including a 2 phases. The primary combination phases includes the analysis of the effect of combining numerous DL methods with hand-crafted feature extraction techniques employing the AE-DL approach. The secondary combination phase builds a multiple classifier system (MCS) to integrate outputs in classifiers, for additional increase the accuracy of this developed Histo-CADx. In [16], a DL method was introduced employed an "end-to-end" training approach. Initially, this developed method applies the adapted contrast enhancement approach. Then, the moveable texture TTCNN was introduced as well as energy layer both are incorporated. This developed algorithm includes only a one energy layer and 3 layers of convolution. Lastly, the effectiveness has been examined dependent upon deep features of CNN methods.

This research offers a new moth flame optimization employing breast cancer detection and classification (MFO-BCDC) algorithm. This research work begins with the application of DCNN, a robustness DL model for image analysis. Convolutional layers in the network extract related features from mammographic images, offering a robust foundation for subsequent classification. Moth Flame Optimization (MFO) has been presented as a meta-heuristic algorithm for hyperparameter tuning of the DCNN. To complement the DL method, the study integrates the XGBoost method for classification. XGBoost, known for its ensemble learning abilities can be utilized to effectively handle the extracted features and determine prediction accuracy. The developed combined architecture was assessed employing benchmark BC datasets, and wide-ranging experiments were executed to analysis the effectiveness of all modules.

2. The proposed model

In the article, we have introduced a novel MFO-BCDC methodology. The main purposes of MFO-BCDC algorithm comprise three essential processes namely DCNN-based feature extraction, MFO-based hyperparameter tuning, and XGBoost-based classification. Fig. 1 represents the entire process of MFO-BCDC methodology.

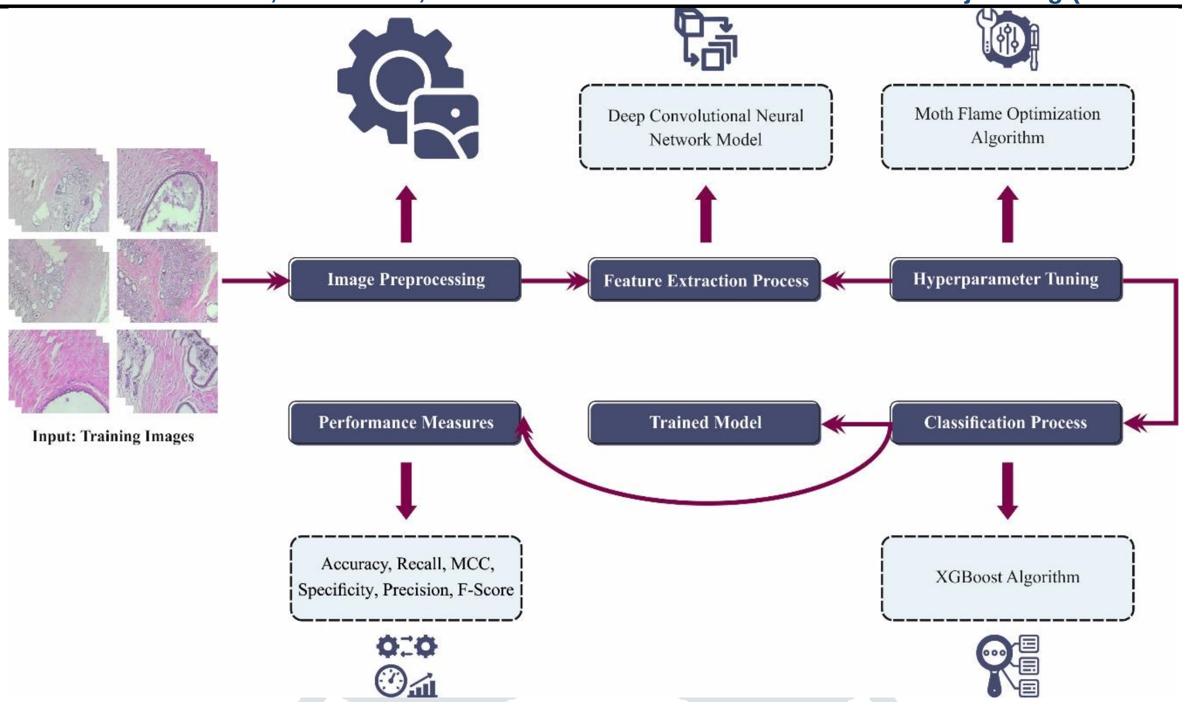

Fig. 1. Overall process of MFO-BCDC methodology

2.1. Feature extraction using DCNN

Initially, the MFO-BCDC technique is applied DCNN model for extracting features. CNN is a DL approach that takes an input image, provides values (learnable weight and bias) to various objects in the image, and differentiates one from the other [17]. Typically, CNN is intended for image analysis tasks including medical image analysis, image classification, image and video recognition. This NN model transform input image into a subset known as filters. The matrix function between the input image and the filters section helps extract significant features. A CNN comprises of multiple layers namely fully connected, convolution, and pooling layers. Multiple pooling and convolution functions are repeated, followed by multiple FC layers. Convolution is a linear function which generates the mapping features by multiplying a sequence of weights with the input images (n-dimensional matrices). It extracts high level features like edges, from input image. The mapping feature can be produced by the sum of dot product between all the elements of filter and the input tensor matrix to generate a convoluted mapping feature. Then, these convoluted features are inputted to the next layer of CNN.

After each convolution layer, the ReLU activation function provides nonlinearity to the network. The activation function implements an entry-wise non-linear conversion and set the negative pixel as 0. This function tries to resolve the gradient vanishing problems faced while other activations including $tanh$, or sigmoid are utilized. The equation for ReLu function is formulated as follows:

$$f(x) = \begin{cases} 0, & for\ x < 0, \\ x, & for\ x \geq 0 \end{cases} \tag{1}$$

The pooling process, otherwise known as downsampling, decreases the dimensionality of mapping feature. Downsampling subsample large size feature map to generate small feature map, which keeps the features. Pooling controls over-fitting by decreasing the parameter count and computation in the network. Pooling is performed by the different non-linear functions. The most popular way of pooling utilized in CNN is known as max pooling that splits the input images into rectangles and output the maximum for the subregion.

A FC layer generates the classification output of CNN after a sequence of pooling and convolution operations. Similar to the classical ANN, the FC layer has a similar working mechanism namely multi-layer perceptron (MLP). After the convolution and pooling operations, the high-level extracted features are fattened to 1D vector form and are inputted to the FC layer. The last FC layer is a classifier that categorizes the input image into multiple classes.

2.2. MFO based hyperparameter tuning

At this stage, the MFO is introduced as a metaheuristic algorithm for hyperparameter tuning of the DCNN. The MFO technique is considered by drawing stimulus in the flight forms of moths [18]. During the MFO, moths elegantly orbit individual, various flames. All the moths diligently patrols its own flame, and upon determining a higher performance, the moth quickly creates the essential upgrade. The agent places of dealing with the objective endure changes using the application of Eq. (2).

$$M_i = S(M_i, F_j) \tag{2}$$

M_i refers the i^{th} moth, but F_j appears to the j^{th} objective. The variable S denotes the spiral function, ruled by a group of 3 certain situations. Initially, the spiral beginning supports with the position engaged by the moth. Then, the spiral's termination coincides with the specific flame place. At last, the spiral oscillation range can constantly endure confined from the borders of the determined searching space. After the success of all necessities, the spiral function, represented as S, is correctly written utilizing Eqs. (3) and (4).

$$S(M_i, F_j) = D_i \times e^{bt} \times \cos(2\pi t) + F_j \tag{3}$$

$$D_i = |F_j - M_i| \tag{4}$$

D_i implies the difference of i^{th} agent and j^{th} objective. b symbolizes the value ordering the spiral's method, but t signifies the arbitrary value falling from the range in [r-1]. It is important to note that r defines an adaptive convergence constant that endures a linear reduction from -1 to -2. This progressive decrease of r assists the drive of expediting the convergence method nearby the flame. Lesser t values equal to smaller distances among the j^{th} flame as well as i^{th} moth. But, once a single moth can enclose many different flames, the efficacy in developing the optimum performance are compromised. Therefore, to resolve this problem, the count of objectives are progressively reduced during the iterations.

$$flame\ number = round\left(N - l \times \frac{N-1}{T}\right) \tag{5}$$

l indicates the iteration. N stands for the flame quantity at maximal value, and T refers to the maximal iterations. This dynamic feature efficiently harmonizes the exploration and exploitation features from the searching space. Regarding power point tracking, the moths signify duty cycles. Once optimizer, the moths fine-tune their places compared with their connected flames. Later all the iterations, the objectives have been efficient by their higher objectives.

The MFO method derives a fitness function (FF) to achieve increased classification efficiency. It evaluates a positive integer to signify the better performance of the candidate solutions. In this study, the reduction of the classification error rate can be measured as the FF, as denoted in Eq. (6).

$$fitness(x_i) = ClassifierErrorRate(x_i)$$

$$= \frac{number\ of\ misclassified\ samples}{Total\ number\ of\ samples} * 100 \tag{6}$$

2.3. XGBoost classification

The XGBoost technique is employed for classification process. XGBoost (Extreme gradient boosting,) is a robust algorithm enhanced based on the gradient boosting iterative decision tree (GBDT) model [19]. By continuously adding new DTs, the objective function is improved. Based on the residual of the prior tree, each tree is sequentially trained. The value of loss function continued to reduce each time a tree is added. Thus, the XGBoost model is extensively applied in wildfire disaster risk prediction, stock prediction, wind power prediction, etc.

The constraint regularization term and the loss function are the two major parts of objective function $L(\phi)$. The constraint regularization term is a penalty mechanism to avoid the model from over-fitting and the loss function is the degree to which the model fits the data.

$$L(\phi) = \sum_{i=1}^{n} l\,(\hat{y}_i, y_i) + \sum_{k=1}^{k} \Omega\,(f_k) \tag{7}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\|\omega\|^2 \tag{8}$$

Where the forecasted value of sample x_i is y_i, T denotes the number of leaf nodes of the tree, y_i denotes the real value, the output value of the k^{th}subtree is f_k, the regular term of k^{th} tree is $\Omega(\mathrm{f}_k)$, the number of subtrees indicates $k,$ the leaf node value is ω, the hyperparameter is λ, and the training error of the sample x_i is $l(\hat{y}_i, y_i)$.

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k\,(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{9}$$

$$\sum_{k=1}^{k} \Omega\,(f_k) = \sum_{k=1}^{t-1} \Omega\,(f_k) + \Omega(f_t) \tag{10}$$

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{11}$$

The objective function $L^{(t)}$ is extended once the structure of optimum tree is defined, by the second-order Taylor to attain the optimum weight and objective function value on every leaf.

3. Result analysis

The experimental outcomes of the MFO-BCDC method can be studied on the BreakHis dataset. It contains 100x dataset comprising 2081 samples under two classes as defined in Table 1.

Table 1 Details on dataset

Classes	**Magnification : 100X**
Benign	644
Malignant	1437
Total	**2081**

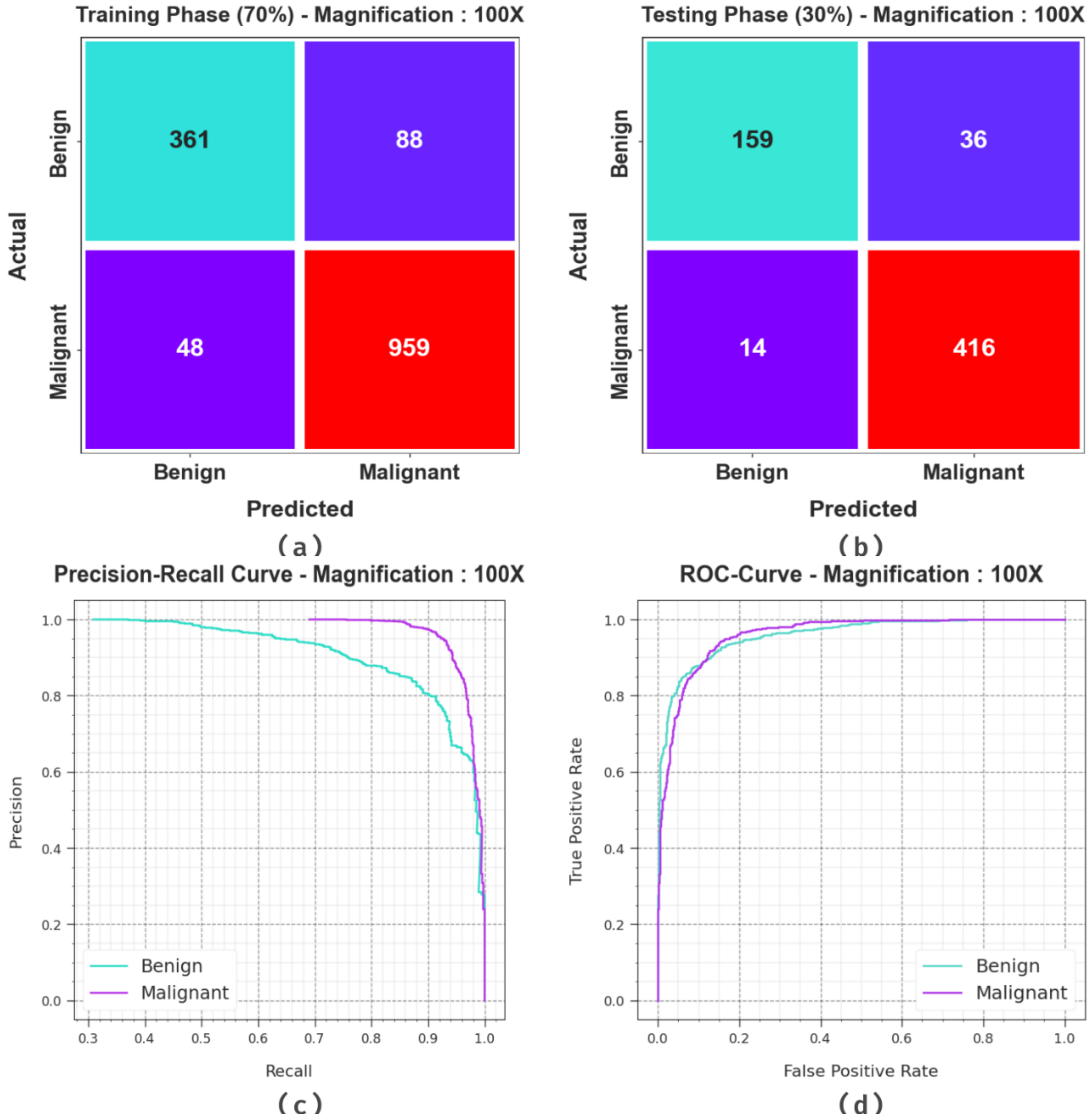

Fig. 2. 100x dataset (a-b) Confusion matrices and (c-d) PR and ROC curves

Fig. 2 shows the classifier analysis of the MFO-BCDC system with 100x dataset. Figs. 2a-2b exhibits the confusion matrix achieved by the MFO-BCDC technique on 70:30 of TRPH/TSPH. This figure represented that the MFO-BCDC method can be correctly identified and categorized with 2 class labels. Additionally, Fig. 2c reveals the PR analysis of the MFO-BCDC system. The figure described that the MFO-BCDC model attains higher PR performance on each class. Also, Fig. 2d shows the ROC analysis of the MFO-BCDC mthodology. The figure displaysed that the MFO-BCDC algorithm leads to better outcomes with greater ROC values on distinct classes.

In Table 2 and Fig. 3, a BC detection results of MFO-BCDC algorithm on 70:30 of TRPH/TSPH. The results states that the MFO-BCDC method has efficient and recognizes benign and malignant samples correctly. With 70% of TRPH, the MFO-BCDC approach has offers average $accu_y$ of 87.82%, $prec_n$ of 89.93%, $reca_l$ of 87.82%, $spec_y$ of 87.82%, F_{score} of 88.76%, and MCC of 77.72%. In addition, on 30% of TSPH, the MFO-BCDC method gets average $accu_y$ of 89.14%, $prec_n$ of 91.97%, $reca_l$ of 89.14%, $spec_y$ of 89.14%, F_{score} of 90.37%, and MCC of 81.06%.

Table 2 BC detection outcome of MFO-BCDC algorithm on 70:30 of TRPH/TSPH

Classes	Accuracy	Precision	Recall	Specificity	F-Score	MCC
TRPH (70%)						
Benign	80.40	88.26	80.40	95.23	84.15	77.72
Malignant	95.23	91.60	95.23	80.40	93.38	77.72
Average	**87.82**	**89.93**	**87.82**	**87.82**	**88.76**	**77.72**
TSPH (30%)						
Benign	81.54	91.91	81.54	96.74	86.41	81.06
Malignant	96.74	92.04	96.74	81.54	94.33	81.06
Average	**89.14**	**91.97**	**89.14**	**89.14**	**90.37**	**81.06**

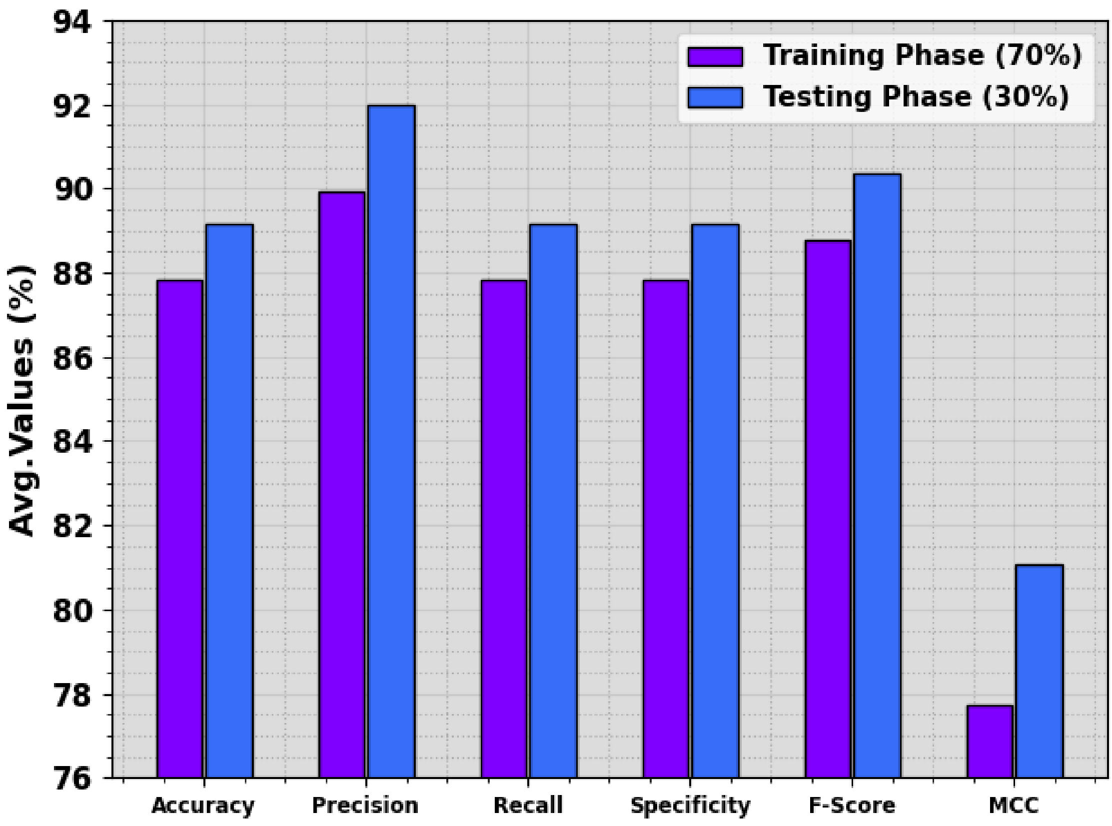

Fig. 3. Avergae of MFO-BCDC algorithm on 70:30 of TRPH/TSPH

Table 3 and Fig. 4 represents the comparative analysis of MFO-BCDC algorithm with recent methods on 100X dataset. The results implies that the MFO-BCDC system has outperformed higher result compared to other methods. Based on $accu_y$, the MFO-BCDC algorithm has obtained higher $accu_y$ of 89.14% while the PFTAS-QDA, InceptionV3, ResNet-50, InceptionResNetV2, and Xception approaches have gained lesser $accu_y$ of 82.64%, 76.96%, 71.26%, 70.45%, and 81.85% respectively. Accoding to $prec_n$, the MFO-BCDC system gets better $prec_n$ of 91.97% whereas the PFTAS-QDA, InceptionV3, ResNet-50, InceptionResNetV2, and Xception methodologies gains lower $prec_n$ of 82.64%, 92.03%, 72.97%, 91.31%, and 89.07% respectively. Besides, with $reca_l$, the MFO-BCDC system attains $reca_l$ of 89.14% but, the PFTAS-QDA, InceptionV3, ResNet-50, InceptionResNetV2, and Xception techniques get decreased $reca_l$ of 82.51%, 69.25%, 88.12%, 63.14%, and 84.87% correspondingly.

Table 3 Comparison outcome of MFO-BCDC algorithm with recent methods on 100X dataset

Magnification : 100X				
Methods	**Accuracy**	**Precision**	**Recall**	**F-Score**
MFO-BCDC	89.14	91.97	89.14	90.37
PFTAS-QDA	82.64	82.64	82.51	81.58
InceptionV3	76.96	92.03	69.25	80.51
ResNet-50	71.26	72.97	88.12	81.57
InceptionResNetV2	70.45	91.31	63.14	74.45
Xception	81.85	89.07	84.87	86.91

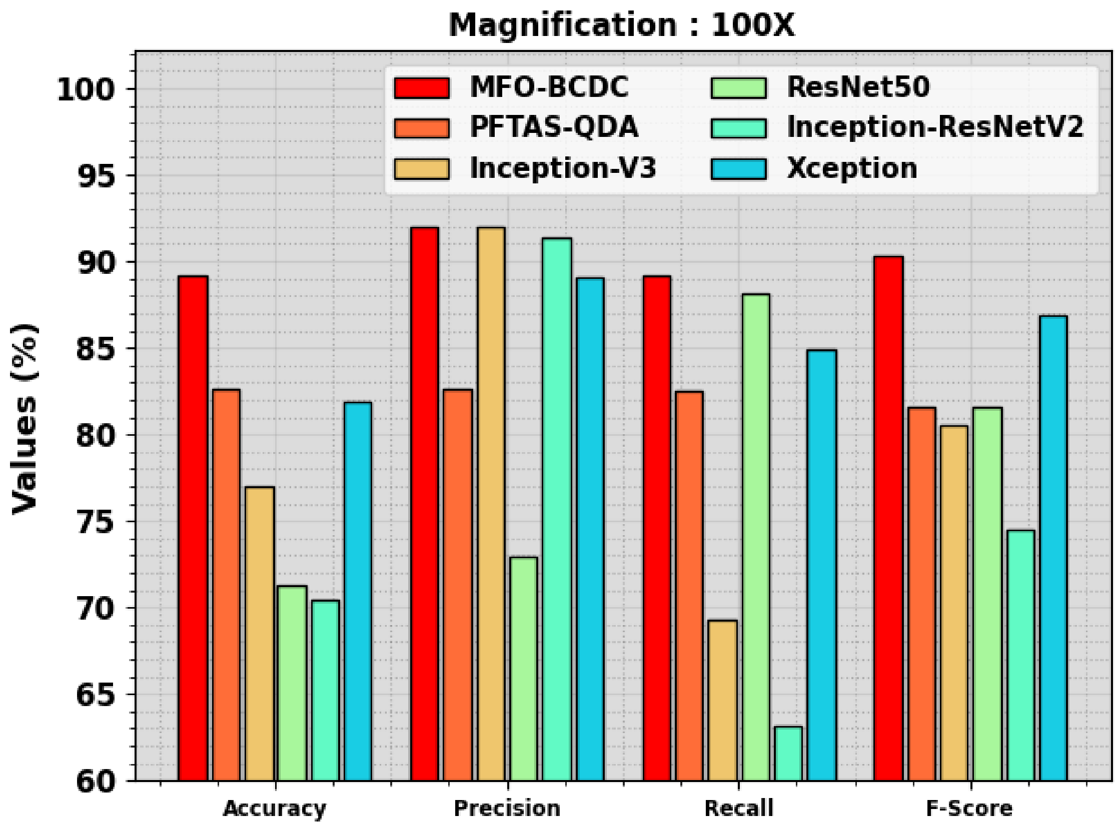

Fig. 4. Comparative outcome of MFO-BCDC algorithm on 100X dataset

4. Conclusion

In this study, we focuses on design and development of MFO-BCDC algorithm. The main purposes of MFO-BCDC algorithm comprise three essential processes namely DCNN-based feature extraction, MFO-based hyperparameter tuning, and XGBoost-based classification. This research work begins with the application of DCNN, a robustness DL model for image analysis. The MFO has been presented as a meta-heuristic algorithm for hyperparameter tuning of the DCNN. To complement the DL method, the study integrates the XGBoost method for classification. The developed combined architecture was assessed employing benchmark BC datasets, and wide-ranging experiments were executed to analysis the effectiveness of all modules.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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).

[9] 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.

[10] 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.

[11] Umer, M.J., Sharif, M., Alhaisoni, M., Tariq, U., Kim, Y.J. and Chang, B., 2023. A Framework of Deep Learning and Selection-Based Breast Cancer Detection from Histopathology Images. *Computer Systems Science & Engineering*, *45*(2).

[12] Elmannai, H., Hamdi, M. and AlGarni, A., 2021. Deep learning models combining for breast cancer histopathology image classification. *International Journal of Computational Intelligence Systems*, *14*(1), p.1003.

[13] Zaalouk, A.M., Ebrahim, G.A., Mohamed, H.K., Hassan, H.M. and Zaalouk, M.M., 2022. A deep learning computer-aided diagnosis approach for breast cancer. *Bioengineering*, *9*(8), p.391.

[14] Sahu, Y., Tripathi, A., Gupta, R.K., Gautam, P., Pateriya, R.K. and Gupta, A., 2023. A CNN-SVM based computer aided diagnosis of breast Cancer using histogram K-means segmentation technique. *Multimedia Tools and Applications*, *82*(9), pp.14055-14075.

[15] Attallah, O., Anwar, F., Ghanem, N.M. and Ismail, M.A., 2021. Histo-CADx: duo cascaded fusion stages for breast cancer diagnosis from histopathological images. *PeerJ Computer Science*, *7*, p.e493.

[16] Maqsood, S., Damaševičius, R. and Maskeliūnas, R., 2022. TTCNN: A breast cancer detection and classification towards computer-aided diagnosis using digital mammography in early stages. *Applied Sciences*, *12*(7), p.3273.

[17] Khatri, N., Lee, S. and Nam, S.Y., 2023. Transfer Learning-based Intrusion Detection System for a Controller Area Network. *IEEE Access*.

[18] Chang, C.C.W., Ding, T.J., Han, W., Chai, C.C., Yam, C.M., Yian, H.C. and Xin, L.H., 2023. Moth flame optimization for the maximum power point tracking scheme of photovoltaic system under partial shading conditions. *Energy Reports*, *9*, pp.374-379.

[19] Song, W., Han, X. and Qi, J., 2023. Prediction of Gas Emission in the Working Face Based on LASSO-WOA-XGBoost. *Atmosphere*, *14*(11), p.1628.