

SEDDA-An Optimal Feature Selection Approach For Breast Cancer Diagnosis

S. Vani Kumari¹, K. Usha Rani²

¹ Research Scholar, Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam, & Lecturer in Department of Computer Applications, Government College for Women(A), Srikakulam, India.

² Professor, Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India.

Abstract:

Soft Computing is an area of Computer Science that aims at solving different types of problems ranging from Medical Image Classification, Content-Based Image Retrieval, Sentiment Analysis, etc. Medical Image Classification involves extracting the essential information from the images and develop models to classify the images. With the advent of the latest technologies in the medical field, there is an explosion of high dimensional data that is to be processed. Hence techniques are required to find the important features of data that are to be processed which can be achieved by feature selection. Feature selection is being applied in fields like Data Mining and Data Science. Feature Selection aims at the elimination of irrelevant and redundant features that adversely affect the performance of any machine learning technique. This paper aims to propose a novel metaheuristic method named Separated Enemy Driven Dragonfly Algorithm (SEDDA) which is an improvisation of the Dragonfly algorithm for selecting an optimal subset of features that are extracted from digital mammograms. The texture features of the segmented area of interest have been extracted and the optimal feature set has been obtained using SEDDA. An accuracy of 94.5 and sensitivity 87.3 was achieved by Multi-Layer Perceptron with Back Propagation which considered for classification purpose. The performance of the novel algorithm is compared with genetic algorithm, particle swarm optimization, ant colony with cuckoo search and dragonfly algorithm. The results show that the novel algorithm is more accurate than the other algorithms.

Keywords: Soft Computing, Metaheuristic, Dragonfly algorithm,

I. Introduction

Higher dimensionality is the main drawback that reduces the performance of a classifier. As the dimensionality of the data increases the amount of data necessary to provide reliable analysis grows exponentially and this phenomenon is termed as the curse of dimensionality [1]. Medical images are characterized by a very large number of distinct features. For classifying a patient into either cancerous or non-cancerous all the features need to be considered by the classifier which results in reduced accuracy. Since

some of the features that were considered for classification may be irrelevant, redundant and noisy. Feature selection aims to eliminate redundant and irrelevant features and hence improve the accuracy of the learning classifier. Reduction in the number of features also results in lesser computational time and memory [2]. A good feature selection method would indeed extract the most relevant features that can provide a better understanding of the process that generated the data. Hence feature selection is highly drawing attention in medical image analysis [3].

Feature selection is essential in machine learning and is been applied to solve a variety of problems and is considered to be an NP-hard problem [4]. Generally feature selection methods can be broadly categorized into filter, wrapper and embedded methods. In filter methods the subset of features are ranked based on statistical measures, few methods include variance threshold, correlation coefficient and chi-square test. The subset selection is independent of the learning algorithm being used, requires less computational time and avoids overfitting of data. The limitation of these methods is it may result in subsets that are non-optimal and redundant [5]. In wrapper methods, the subset selection is based on the results of the induction algorithm being applied. In wrapper methods repeated learning and cross-validations leads to higher computational time and expensive compared to filter methods. Sequential feature selection and recursive feature elimination methods are examples of wrapper methods. There is a chance of overfitting of data if the learning algorithm learns the data too well and gives rise to poor generalization [6]. Embedded methods on the other hand ensembles the advantages of both the filter and wrapper methods to find the feature subset. The performance of the embedded methods is better compared to that of the other two feature selection methods and is computationally better than the wrapper methods. An example for embedded method is the random forest method. This method suffers the drawbacks of the learning algorithm in specific[7-8].

One of the challenges faced by the feature selection methods is searching for an optimal subset of features. The main objective of feature selection is to find a set of Q features from an original set of P where $P < Q$ without any information loss. Hence generating all the possible subsets is practically impossible to solve such problems. If the dataset has P features then 2^P subsets are to be considered and evaluated which is considerably expensive[2].

Metaheuristics have gained prominence in recent years are being applied to solve the problem of feature selection. Metaheuristic algorithms explore the large search space efficiently and reduce the size of the space and hence are applied for solving a large-sized problem by delivering a satisfactory solution in a reasonable amount of time. Being efficient and effective in solving complex problems metaheuristics have been applied in many areas like machine learning, engineering design, planning in routing problems, image processing, etc[9]. Most of the metaheuristic methods are inspired by the natural metaphor eg evolution of species, bee colony, particle swarm, ant colony, etc.

Dragonfly algorithm is a new metaheuristic optimization technique based on swarm intelligence was developed by Miraliji[10]. The method has shown its ability to solve different optimization problems in the real world and hence was applied to solve the problem of feature selection. The algorithm is based on the static and dynamic swarming behaviors of the dragonfly in search of their food. Dragonfly can be applied for

solving single-objective, binary, and multi-objective problems and combined with other meta-heuristic methods to gain the combined advantage of the methods[11].

This paper intends to develop a novel breast cancer detection model for classifying benign and malignant patterns in mammograms. The entire process is done in four stages such as pre-processing, feature extraction, feature selection and classification. Textural features of the image are extracted using both Grey Level Co-occurrence Matrix (GLCM) and Grey Level Run-Length Matrix (GLRM) methods. Since the number of features is more, it is necessary to select the optimal features. Thus the optimal feature selection is carried out by adopting an improved Dragonfly Algorithm (DA) called Separated Enemy Driven Dragonfly Algorithm (SEDDA).

The paper is formulated as follows. Section II discusses the related works on this topic. Section III describes the proposed architecture and section IV describes various phases involved in the proposed method. Moreover, section V portrays the utilization of proposed SEDDA. In addition, section VI explains the results and section VII finalizes the paper.

II. Related Works

Several research works have been conducted to exhibit an automated detection system for the identification of breast cancer in its early stage. This study highlights only the recent and relevant works carried out in this area. Mammogram images are to be pre-processed for improving the quality of the images before extracting the features. Pre-processing involves different steps like background removal, filtering and segmentation. Active Contour method is a popular method for separation of background from foreground[12]. Speckle, salt and pepper noises can be removed from images using Wiener Filter[13] and rotation invariance is obtained using a popular technique, Discrete Fourier Transform[14]. The watershed method is applied to obtain a segmented area of interest[15]. Textural features are reported to be the most prominent type of features that are extracted from the segmented area of interest[15].

Biswas et.al[16] proposed a Computer-Aided Detection (CAD) system to classify mammograms into normal and abnormal by extracting textural features using GLCM. Support Vector Machine(SVM), k-Nearest Neighbor(k-NN) and Artificial Neural Network (ANN) are adopted for classification. María, et al [17] classified mammographic images obtained from Digital Database for Screening Mammography (DDSM) into normal, benign, and abnormal using backpropagation based ANN by extracting the features based on Harlick descriptors. Singh et.al [18] adopted a CAD system for classifying mammographic images obtained from the Mini-MIAS database by extracting the GLCM features in three directions. Most relevant features are selected using Ada-Boost feature selection and the Random Forest method is used for classification. Nawel, et. al[19] proposed a novel system for classifying breast images based on optimal feature selection by combining Mutual Information and Correlation-based feature selection.

Sudha et.al [20] proposed a novel meta-heuristic feature selection method based on the enhanced cuckoo search for selecting the minimal number of features for classifying mammograms. Textural features based on GLCM have been extracted from the mammograms and the feature selection algorithm selected the best subset that could classify the images efficiently. Kayode et.al [21] proposed a hybrid algorithm for selecting the relevant features from mammograms and thereby improving the classification accuracy. GLCM

features are extracted from the mammograms and the weighted average of all the features is considered. Sequential forward selection and Genetic algorithm methods are combined to select the optimal features. Figlu, et al. [22] developed a CAD model to classify mammographic images. Fusion-based feature extraction based on 2D -BDWT, and GLCM is used to extract features of mammograms. PCA was adopted to reduce the size of the feature vector. Later Forest Optimisation was adopted for both optimal feature selection and SVM, k-NN, and C4.5 are used for classification. Sudha et.al [23] implemented a hybrid model for selecting optimal features from breast cancer images based on Cuckoo search and harmony search. Texture, shape, intensity histogram, and radial distance features are extracted from the region of interest and the optimal features are selected using the proposed hybrid model.

Chaieb et.al [24] adopted a two-step feature subset selection to obtain the most relevant features. Various Texture and shape-based features are extracted and in the first step Tabu search, genetic algorithm, and relief-f are used to select relevant features. In the second step, the most relevant features are rated based on five measures derived from the Rodrigues approach. GLCM features are found to be the most relevant features of the experimental results. J.B.Jona et.al[25] applied a hybrid metaheuristic method, a combination of Ant Colony Optimization and Cuckoo Search for selecting optimal features of mammogram images obtained from Mini-MIAS database. The efficiency of the method is tested based on the SVM classifier based on the Radial basis function. Mafara et.al [26] proposed a wrapper feature selection method based on dragonfly to select optimal features and hence improve the classification accuracy. Eighteen benchmark datasets from UCI repository were considered for experimentation and the binary dragonfly method proved to be efficient which could search the feature space and select the most informative features.

III. Proposed Architecture

Let the input mammogram image is represented as I, which is subjected to pre-processing using three methods, namely active contour for background removal, Wiener filtering, and DFT. The resultant output from the pre-processed image is segmented using watershed algorithm, which segments the affected region of the image. Moreover, the segmented image is subjected to feature extraction by GLCM and GLRM features. GLCM includes four directions, such as 0^0 , 90^0 , 180^0 , and 270^0 . As the lengths of both the features are much extended, it is necessary to select the features. Hence, for feature selection, an optimization technique called SEDDA MLP is adopted to select the features in an effective manner. Finally, the optimized features are given to MLPBPN, where the number of hidden neurons are optimized using SEDDA. Also, it determines the nature of breast cancer, whether it is normal, benign or malignant. Thus the classified breast cancer image can be obtained with better classification accuracy.

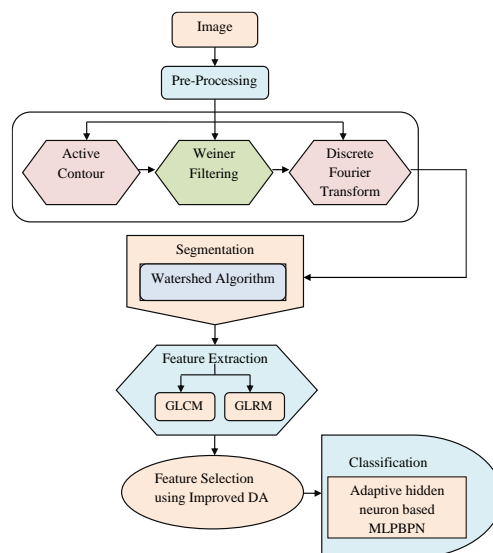


Fig. 1. Overall architecture of the proposed method

IV. Phases Embraced In Diagnosing Breast Cancer

A. Pre-processing

Three wellknown preprocessing techniques are applied in a sequence to preprocess the mammogram images .In the first step an Active Contour [27] is applied for background removal. In the second step a wiener filter[28] is used to enhance the quality of the images and finally in the third step Discrete Fourier Transform [29] is applied to obtain rotation invariance.

B. Segmentation

Watershed algorithm [30] is adopted for finding the segmented area of interest.

C. Feature Extraction

The segmented area of interest is subjected to feature extraction by using GLCM [31] and GLRM [32] schemes.

The features extracted using GLCM is denoted by $F = \{f_1, f_2 \dots f_n\}$, whereas, the features extracted using GLRM is indicated by $G = \{g_1, g_2 \dots g_n\}$. Both the features can be combined and represented as $F^* = F + G$.

D. Feature Selection

As the length of the extracted features is more, it is essential to adopt a feature selection process, by which significant features can be selected optimally.

E. Classification

The selected features are subsequently given to MLPBPN technique where the number of hidden neurons are optimized using SEDDA for classifying the type of breast cancer, whether it is normal, benign or malignant,

II. UTILIZATION OF PROPOSED SEDDA

A. Objective model

The features, which are selected optimally along with the hidden neurons on MLPBPN is given as the solution for encoding as shown by Fig. 2, where N_H is the count of hidden neurons and F_N^* is the count of selected features.

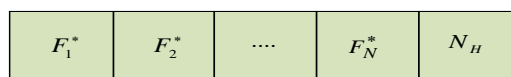


Fig. 2. Solution encoding of the proposed method

The chromosome features consists of values lying between 0 and 1. Here, if the solution variable is zero, it has to be omitted, and if the variable is one, the particular feature is considered. The length of chromosome can be evaluated by integrating the length of all the extracted features and the size of hidden neurons.

The fitness function of the proposed method is given by Eq. (19), where, R is the fitness value, h is the actual value and b is the predicted value. The objective model of the suggested method is shown by Eq. (20), where Z is the objective function to be achieved.

$$R = RMSE(h - b) \tag{19}$$

$$Z = Min\{RMSE(h - b)\} \tag{20}$$

B. Optimal Feature Selection

For choosing the features optimally, improved DA algorithm has been adopted. Generally, dragonflies are tiny predators that attack other tiny insects for prey. The chief motivation of the DA scheme [28] arises from dynamic and static swarming characteristics. These swarming characteristics are found to be related incredibly to the two most important stages of optimization by meta-heuristics function, called exploitation and exploration. Accordingly, in the static swarm, they fly in a particular direction in larger swarms that is performed in the exploitation phase. The entire behaviors are designed subsequently.

The separation formulation is measured as in Eq. (21), in which M_j reveals the position of neighboring individual, M is the position of the current individual and U is the number of neighboring individuals.

$$W_i = - \sum_{j=1}^U M - M_j \tag{21}$$

Alignment is estimated as specified by Eq. (22), in which Q_j denotes the velocity of j^{th} neighbouring individual. Moreover, the cohesion formulation is approximated by Eq. (23), in which M_j signifies the position neighboring individual, U indicates the neighborhood amount and M denotes the current individual's position.

$$B_i = \frac{\sum_{j=1}^U Q_j}{U} \tag{22}$$

$$A_i = \frac{\sum_{j=1}^U M_j}{U} - M \tag{23}$$

The attraction of a dragonfly to a food resource is evaluated by Eq. (24), in which M^+ reveals the position of the food source and M represents the current individual's position.

$$F_i = M^+ - M \tag{24}$$

Distraction outwards an enemy is specified by Eq. (25), in which M^- exposes the enemy's position and M is the position of the current individual.

$$E_i = M^- + M \tag{25}$$

The artificial dragonflies' position is required for updating an exploration phase and for carrying out their movements, two vectors are evaluated namely step vector and position.

The step vector shows the direction of the dragonfly movement as exposed in Eq. (26), in which W_i indicates the separation of i^{th} individual, ρ denotes the separation weight, a is the alignment weight and A_i represents the i^{th} individual cohesion, c signifies the cohesion weight, B exposes the alignment of i^{th} individual, f denotes the food factor, F_i signifies the food resource of the i^{th} individual, e indicates the enemy factor, w offers the inertia weight and E_i symbolizes the position of enemy of i^{th} individual.

$$\Delta M_{t+1} = (\rho W_i + a B_i + c A_i + f F_i + e E_i) + w \Delta M_t \tag{26}$$

Subsequent to the evaluation of step vector, the position vectors are formulated as in Eq. (27), in which t denotes the current iteration.

$$M_{t+1}^1 = M_t + \Delta M_{t+1} \tag{27}$$

For improving the stochastic nature, exploration, and exploitation, they are demanded to fly over the exploration space by an arbitrary walk, while there is nonexistence of neighboring solutions. Under such situations, the dragonfly's position is updated by Eq. (28), in which z denotes the size of the position vectors and t indicates the present iteration.

$$M_{t+1}^1 = M_t + Levy(z) \times M_t \tag{28}$$

The Levy flight is evaluated as in Eq. (29), in which, β is a constant, r_1 and r_2 are the arbitrary numbers that lies among $[0,1]$, and r can be validated as in Eq. (30), in which $\alpha(x) = (x-1)$

$$Levy(x) = 0.01 \times \frac{r_1 \times \delta}{|r_2|^\beta} \tag{29}$$

$$\delta = \left(\frac{\alpha(1+\beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\alpha \frac{(1+\beta)}{2} \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}} \right)^{\frac{1}{\beta}} \tag{30}$$

The traditional DA algorithm was capable to resolve only the continuous and single objective optimization issues. Hence, this paper adopts SEDDA MLP to prevail over the limitations existing in it. In the proposed improved DA, a new position update is also discovered using fitness threshold as given by Eq. (32). The final position is updated based on a variable D_α that is given in Eq. (31).

$$D_\alpha = |c \times F_i - M_t| \tag{31}$$

The newly updated positions with respect to the fitness threshold for food position and enemy position are given by Eq. (32) respectively.

$$Fitness\ threshold = food\ fitness + (food\ fitness * 0.2) \quad (32)$$

Algorithm 1 :Proposed SEDDA

```

Initialize the population of dragonfly  $M_i(i = 1,2,...n)$ 
Initialize step vectors  $\Delta M_i(i = 1,2,...n)$ 
while the end condition is not fulfilled
    Compute the objective value of entire fireflies
    Update enemy and food source
    Update  $w, \rho, a, c, f$  and  $e$ 
    Compute  $W_B, A, F$  and  $E$  using Eq. (21-25)
    Neighboring radius is updated
    if a dragonfly involves one neighbor dragonfly,
        Update velocity vector by means of Eq. (26)
        Update position vector by means of Eq. (27)
    else
        update position vector by means of Eq. (28)
    For  $i=1:U$ 
        Find  $D_\alpha$  using Eq. (31)
        if  $fitness(i) < fitness\ threshold$ 
            Update the position using
             $M_{t+1}^2 = F_i - (B_i \times D_\alpha)$ 
        else
            Update the position using
             $M_{t+1}^2 = E_i - (B_i \times D_\alpha)$ 
        end
        Updated position =  $M^* = \left( \frac{M_{t+1}^1 + M_{t+1}^2}{2} \right)$ 
    end if
    
```


Verify and approve the novel positions
 depending on variable boundaries
 end while

The resultant position based on the average to two positions is taken as the optimal position.

C. Optimal Classification

For classifying the breast cancer image, MLPBPN is adopted, which classifies the type of cancer, whether it is normal, benign or malignant. Here, the number of hidden neurons is optimized using proposed SEDDA method. The adopted MLPBPN is a machine learning technique which is inspired by the biological nervous systems. The optimized features $F^* = F_1^*, F_2^*, \dots, F_N^*$ is provided to the network for classification purpose. The output of the hidden layer is indicated by Eq. (33), in which N_I signifies the number of input neurons, O_1 indicates the activation function, w_{bj}^H denotes the bias weight to the j^{th} hidden layer and w_{ij}^H denotes the weight from i^{th} input neuron to j^{th} hidden neuron.

$$H_j = O_1 \left(w_{bj}^H + \sum_{i=1}^{N_I} w_{ij}^H F_i^* \right) \tag{33}$$

Consequently, the output of MLPBPN model is given by Eq. (34), where N^H specifies the number of hidden neurons, O_2 indicates the activation function, w_{ij}^H indicates the weight from j^{th} hidden neuron to i^{th} output neuron w_{bi}^O indicates the bias weight to the i^{th} output neuron. In addition, the optimal selection of weight by the minimization of the objective function is revealed by Eq. (35), where h signifies the actual value and b denotes the predicted value.

$$b = O_2 \left(w_{bi}^O + \sum_{j=1}^{N^H} w_{ji}^O H_j \right) \tag{34}$$

$$e^R = \arg \min_{\{w_{bj}^H, w_{ij}^H, w_{bi}^O, w_{ji}^O\}} \|b - h\| \tag{35}$$

The classified image obtained from the MLPBPN scheme offers better classification about the nature of the breast cancer image, whether it is normal, benign or malignant.

III. RESULTS AND DISCUSSIONS

A. Experimental Setup

The proposed breast cancer diagnosis model using SEDDA-MLP was simulated using MATLAB, and the corresponding results were obtained. The simulation was carried out using MiniMIAS Database. Here, total of 322 images were taken, in which 207 images were normal, 63 images were benign and 52 images were malignant. In addition, the proposed SEDDA-MLP was compared with SVM [25], GA [26], PSO [27] and DA [28] and the results were obtained. The experimentation was done based on the convergence analysis and performance analysis, and the outcomes were validated.

B. Convergence analysis

From Fig. 3, the convergence analysis of the proposed method regarding the cost function with respect to various iterations for breast cancer diagnosis can be attained. From the Fig.3, the proposed scheme for 20th iteration is 23.64% better than GA, 22.86% better than PSO, 40.7% better than ACOCC and 22.09% better than DA techniques. Similarly, for 25th iteration, the suggested method is 23.64% superior to GA, 22.86% superior to PSO, 40.7% better than ACOCC and 22.09% superior to DA techniques. Thus the capability of the implemented scheme in detecting the breast cancer with minimized error has been validated on concerning the convergence analysis.

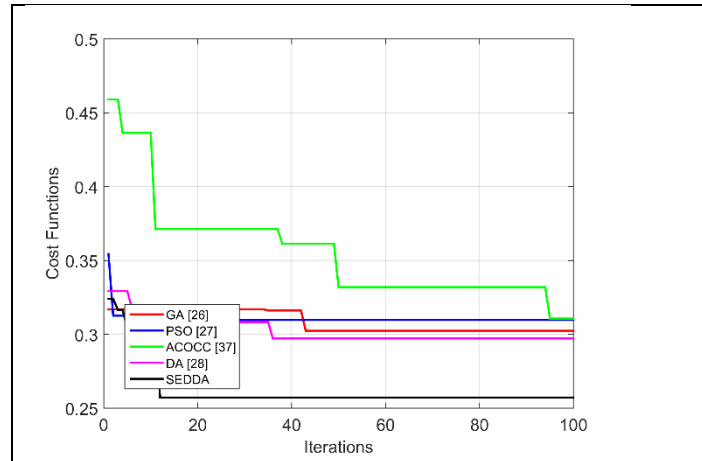
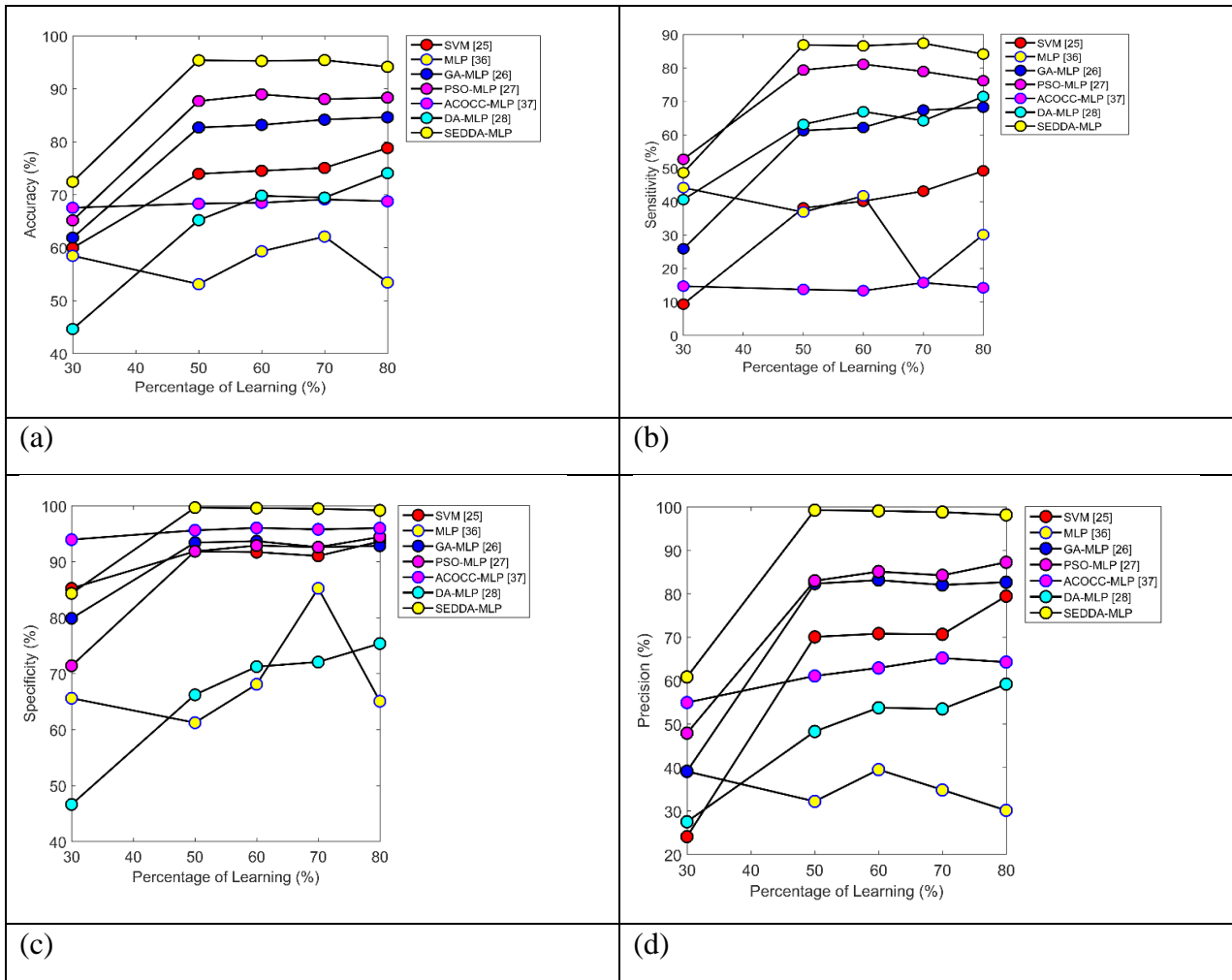


Fig. 3. Convergence analysis of the proposed and conventional schemes

C. Performance analysis

The performance analysis for breast cancer diagnosis for the suggested method is given by Fig. 4 for various measures. From Fig. 4(a), the proposed method regarding accuracy is 25% better than SVM, 8.75% better than GA-MLP, 28.42% better than ACOCC-MLP, 7.5% better than PSO-MLP and 6.25% better than DA-MLP techniques. Similarly, from Fig. 4(b), the suggested scheme regarding sensitivity is 80% superior to SVM, 84.7% superior to ACOCC-MLP, 22% superior to GA-MLP, 20% superior to PSO-MLP and 12% superior to DA-MLP techniques. Moreover, from Fig. 4(c), the implemented method in terms of specificity is 37.33% better than SVM, 12% better than GA-MLP, 6.66% better than PSO-MLP, 3% better than ACOCC-MLP and 4% better than DA-MLP methods. Also, from Fig. 4(d), the precision for the proposed scheme is 41.17% better than SVM, 38.77% better than ACOCC-MLP, 4.7% better than GA-MLP, 2.35% better than PSO-MLP and 3.52% better than DA-MLP techniques. In addition, from Fig. 4(e), the FPR of the presented method is 53.7% superior to SVM, 36% superior to GA-MLP, 60% superior to ACOCC-MLP, 16% superior to PSO-MLP and 12% superior to DA-MLP techniques. Furthermore, the suggested scheme regarding the FNR from Fig. 4(f) is 80% better than SVM, 82.35% better than ACOCC-MLP, 22% better than GA-MLP, 20% better than PSO-MLP and 16% better than DA-MLP algorithms. Similarly, from Fig. 4(g), the NPV of the suggested scheme is 37.33% superior to SVM, 46.15% superior to ACOCC-MLP, 12% superior to GA-MLP, 6.66% superior to PSO-MLP and 4% superior to DA-MLP methods. Moreover, from Fig. 4(h), the FDR measure of the implemented method is 72% better than SVM, 35.71% better than GA-MLP, 91.42% better than ACOCC-MLP, 21.42% better than PSO-MLP and 14.28% better than DA-MLP schemes. Also,

from Fig. 4(i), the F1-score of the proposed method is 80% superior to SVM, 75.55% superior to ACOC-MLP, 16.66% superior to GA-MLP, 15% superior to PSO-MLP and 11.66% superior to DA-MLP algorithms. Similarly, from Fig. 4(j), the MCC of the suggested scheme is 77.27% better than SVM, 80% better than ACOC-MLP, 31.81% better than GA-MLP, 13.63% better than PSO-MLP and 11.36% better than DA-MLP algorithms. Thus the performance analysis of the implemented scheme has been confirmed by means of the experimental analysis.



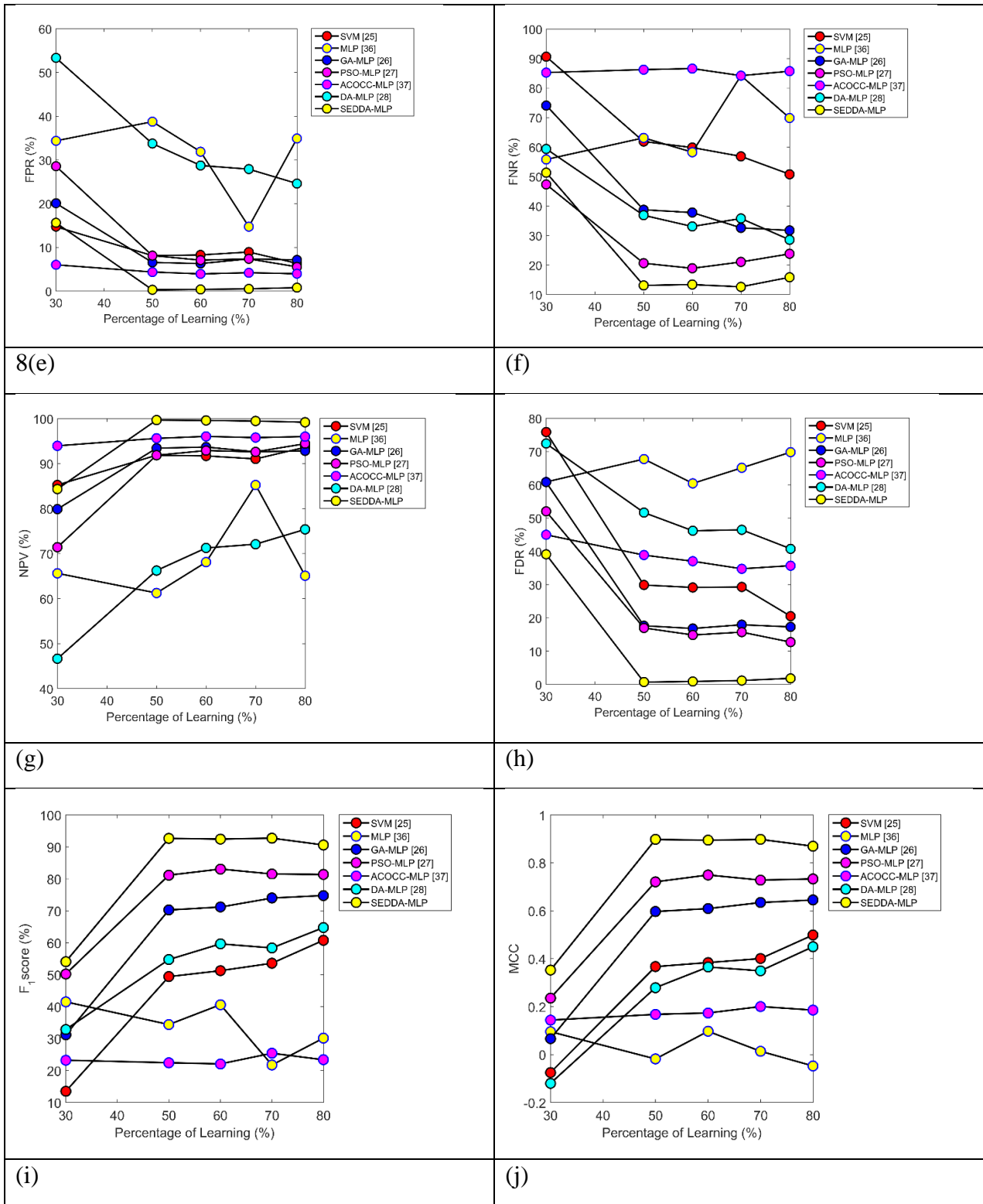


Fig. 4. Performance analysis for the proposed over conventional methods (a) Accuracy (b) Sensitivity (c) Specificity (d) Precision (e) FPR (f) FNR (g) NPV (h) FDR (i) F1-score (j) MCC

The overall performance analysis for breast cancer diagnosis for the implemented technique is specified by Table. II. From Table II, it can be noted that the proposed method in terms of accuracy is 27.7% better than SVM, 21.4% better than GA-MLP, 11.9% better than PSO-MLP, 34.93% better than MLP, 27.59% better

than ACOCC-MLP and 7.79% better than DA-MLP techniques. Likewise, the proposed idea regarding sensitivity is 26.74% superior to SVM, 50.78% superior to GA-MLP, 23.31% superior to PSO-MLP, 81.93% superior to MLP, 81.93% superior to ACOCC-MLP and 9.69% superior to DA-MLP methods. In addition, the implemented technique regarding specificity is 27.61% better than SVM, 89.94% better than GA-MLP, 6.91% better than PSO-MLP, 14.28% better than MLP, 3.7% better than ACOCC-MLP and 6.91% better than DA-MLP methods. Moreover, the precision for the implemented design is 45.85% better than SVM, 28.45% better than GA-MLP, 17% better than PSO-MLP, 64.69% better than MLP, 33.99% better than ACOCC-MLP and 14.98% better than DA-MLP techniques. Additionally, the FPR of the suggested process is 98.05% superior to SVM, 94.08% superior to GA-MLP, 92.79% superior to PSO-MLP, 96.42% superior to MLP, 88.12% superior to ACOCC-MLP and 92.79% superior to DA-MLP techniques. In addition, the proposed scheme in terms of FNR is 64.61% better than SVM, 85.03% better than NN, 3.7% better than ACOCC-MLP 77.76% better than GA-MLP, 61.25% better than PSO-MLP and 39.84% better than DA-MLP algorithms. In the same way, the NPV of the presented scheme is 27.51% superior to SVM, 8.52% superior to GA-MLP, 7.51% superior to PSO-MLP, 16.58% superior to MLP, 3.7% superior to ACOCC-MLP and 7.51% superior to DA-MLP methods. Moreover, the FDR evaluation of the implemented technique is 97.43% better than SVM, 95.89% better than GA-MLP, 93.34% better than PSO-MLP, 98.31% better than MLP, 96.84% better than ACOCC-MLP and 92.41% better than DA-MLP schemes. Furthermore, the F1-score of the proposed method is 37.45% superior to SVM, 76.56% superior to MLP, 72.58% superior to ACOCC-MLP 42.85% superior to GA-MLP, 20.31% superior to PSO-MLP and 12.65% superior to DA-MLP schemes. Correspondingly, the MCC of the implemented method is 61.13% better than SVM, 98.45% better than MLP, 77.68% better than ACOCC-MLP, 55.45% better than GA-MLP, 29.83% better than PSO-MLP and 19.81% better than DA-MLP algorithms. Thus the overall performance analysis of the proposed technique has been validated.

TABLE I. OVERALL PERFORMANCE ANALYSIS FOR THE PROPOSED OVER CONVENTIONAL METHODS

Measures	SVM [25]	MLP [36]	GA- MLP [26]	PSO- MLP [27]	DA- MLP [28]	ACOCC -MLP [37]	SEDDA- MLP
Accuracy	0.69473 7	0.62105	0.75087 7	0.84210 5	0.88070 2	0.69123	0.95438 6
Sensitivity	0.64210 5	0.15789	0.43157 9	0.67368 4	0.78947 4	0.15789	0.87368 4
Specificity	0.72105 3	0.85263	0.91052 6	0.92631 6	0.92631 6	0.95789	0.99473 7
Precision	0.53508 8	0.34884	0.70689 7	0.82051 3	0.84269 7	0.65217	0.98809 5
FPR	0.27894 7	0.14737	0.08947 4	0.07368 4	0.07368 4	0.042105	0.00526 3
FNR	0.35789 5	0.84211	0.56842 1	0.32631 6	0.21052 6	0.84211	0.12631 6
NPV	0.72105 3	0.85263	0.91052 6	0.92631 6	0.92631 6	0.95789	0.99473 7
FDR	0.46491 2	0.65116	0.29310 3	0.17948 7	0.15730 3	0.34783	0.01190 5
F1-score	0.58373 2	0.21739	0.53594 8	0.73988 4	0.81521 7	0.25424	0.92737 4
MCC	0.34944 9	0.01386 4	0.40056 4	0.63439 2	0.72811 6	0.2004	0.89790 7

D. ROC Analysis

ROC curve is a graphical plot that demonstrates the diagnostic potential of a binary classifier system as its threshold discrimination is altered. The ROC curve is generated by plotting the true positive rate (TPR) in opposition to the false positive rate (FPR) at a variety of threshold settings. The ROC analysis of the suggested SEDDA-MLP scheme is shown by Fig. 5, where the proposed method is 17.64% better than SVM, 23.52% better than GA-MLP, 23.52% better than ACOCC-MLP, 8.23% better than PSO-MLP and 8.23% better than DA-MLP techniques respectively. Thus from the ROC analysis, the enhancement of the suggested SEDDA-MLP method has been proved.

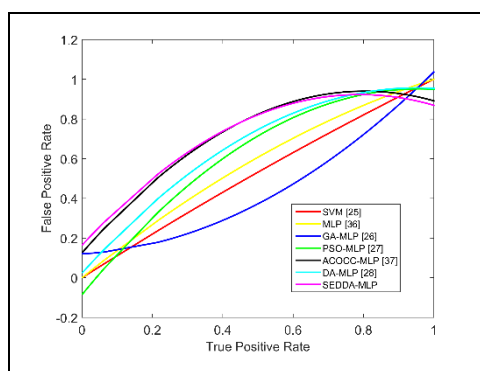


Fig. 5. ROC analysis of the proposed model.

IV. CONCLUSION

This paper has presented an optimal classification of benign and malignant patterns in mammogram. The diagnosis process was made depending on four stages like pre-processing, segmentation, feature extraction, and classification. In the pre-processing phase, the image was subjected to active contour, Weiner filtering, and DFT. Then the cyst in the image was segmented by means of WS scheme. Subsequent to the segmentation, features of the segmented image were extracted by means of both GLCM and GLRM techniques. As the number of features is more, it was essential to select the features in an optimal way. Therefore, the optimal feature selection was done by SEDDA MLP. The selected features were then classified using optimized MLPBPN. Moreover, the proposed SEDDA MLP was compared with traditional algorithms such as SVM, GA-MLP, PSO-MLP and DA-MLP and the results were obtained. From the analysis, the proposed method in terms of accuracy is 27.7% better than SVM, 21.4% better than GA-MLP, 11.9% better than PSO-MLP and 7.79% better than DA-MLP techniques. Likewise, the proposed idea regarding sensitivity is 26.74% superior to SVM, 50.78% superior to GA-MLP, 23.31% superior to PSO-MLP and 9.69% superior to DA-MLP methods. In addition, the implemented technique regarding specificity is 27.61% better than SVM, 89.94% better than GA-MLP, 6.91% better than PSO-MLP and 6.91% better than DA-MLP methods. Thus the capability of the suggested algorithm is verified in terms of its performance.

References

- [1]Bellman, Richard E. "Dynamic programming, ser." Rand Corporation research study. Princeton University Press, 1957.
- [2]Guyon, Isabelle, and André Elisseeff. "An introduction to variable and feature selection." *Journal of machine learning research* 3.Mar (2003): 1157-1182.
- [3]Wang, Ling, et al. "Feature selection based on meta-heuristics for biomedicine." *Optimization Methods and Software* 29.4 (2014): 703-719.
- [4]Chen, Bin, Jiarong Hong, and Yadong Wang. "The minimum feature subset selection problem." *Journal of Computer Science and Technology* 12.2 (1997): 145-153.
- [5]Cherrington, Marianne, et al. "Feature selection: filter methods performance challenges." 2019 International Conference on Computer and Information Sciences (ICCIS). IEEE, 2019.

- [6]Kohavi, Ron, and George H. John. "Wrappers for feature subset selection." *Artificial intelligence* 97.1-2 (1997): 273-324.
- [7]Venkatesh, B., and J. Anuradha. "A review of feature selection and its methods." *Cybernetics and Information Technologies* 19.1 (2019): 3-26.
- [8]Chandrashekar, Girish, and Ferat Sahin. "A survey on feature selection methods." *Computers & Electrical Engineering* 40.1 (2014): 16-28.
- [9]Talbi, El-Ghazali. *Metaheuristics: from design to implementation*. Vol. 74. John Wiley & Sons, 2009.
- [10] Mirjalili, Seyedali. "Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems." *Neural Computing and Applications* 27.4 (2016): 1053-1073.
- [11] Rahman, Chnoor M., and Tarik A. Rashid. "Dragonfly Algorithm and Its Applications in Applied Science Survey." *Computational Intelligence and Neuroscience* 2019 (2019).
- [12] Oliveira, Roberta B., et al. "Computational methods for the image segmentation of pigmented skin lesions: a review." *Computer methods and programs in biomedicine* 131 (2016): 12141.
- [13] Yoo, Jae-Chern, and Chang Wook Ahn. "Image restoration by blind-Wiener filter." *IET Image Processing* 8.12 (2014): 815-823.
- [14] Bianconi, Francesco, and Antonio Fernández. "Rotation invariant co-occurrence features based on digital circles and discrete Fourier transform." *Pattern Recognition Letters* 48 (2014): 34-41.
- [15] Avinash, S., K. Manjunath, and S. Senthil Kumar. "An improved image processing analysis for the detection of lung cancer using Gabor filters and watershed segmentation technique." *2016 International Conference on Inventive Computation Technologies (ICICT)*. Vol. 3. IEEE, 2016.
- [16] Biswas, Ranjit, Abhijit Nath, and Sudipta Roy. "Mammogram classification using gray-level co-occurrence matrix for diagnosis of breast cancer." *2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE)*. IEEE, 2016.
- [17] Pérez, María, et al. "Mammogram classification using back-propagation neural networks and texture feature descriptors." *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*. IEEE, 2017.
- [18] Singh, Vibhav Prakash, et al. "Mammogram classification using selected GLCM features and random forest classifier." *International Journal of Computer Science and Information Security* 14.6 (2016): 82.
- [19] Zemmal, Nawel, et al. "An Enhanced Feature Selection Approach based on Mutual Information for Breast Cancer Diagnosis." *2019 6th International Conference on Image and Signal Processing and their Applications (ISPA)*. IEEE, 2019.
- [20] Sudha, M. N., and S. Selvarajan. "Feature selection based on enhanced cuckoo search for breast cancer classification in mammogram image." *Circuits and Systems* 7.4 (2016): 327-338.
- [21] Kayode, A. A., N. O. Akande, and E. O. Asani. "Implementation of a Hybrid Feature Selection Algorithm for Improving Classification of Mammograms." *Journal of Engineering and Applied Sciences* 14.13 (2019): 4419-4429.

- [22] Mohanty, Figlu, et al. "Digital mammogram classification using 2D-BDWT and GLCM features with FOA-based feature selection approach." *Neural Computing and Applications* (2019): 1-15.
- [23] Sudha, M. N., and S. Selvarajan. "Hybrid approach towards feature selection for breast tumour classification from screening mammograms." *International Journal of Biomedical Engineering and Technology* 29.4 (2019): 309-326.
- [24] Chaieb, Ramzi, and Karim Kalti. "Feature subset selection for classification of malignant and benign breast masses in digital mammography." *Pattern Analysis and Applications* 22.3 (2019): 803-829.
- [25] Jona, J. B., and N. Nagaveni. "Ant-cuckoo colony optimization for feature selection in digital mammogram." *Pakistan journal of biological sciences* 17.2 (2014): 266.
- [26] Mafarja, Majdi M., et al. "Binary dragonfly algorithm for feature selection." 2017 *International Conference on New Trends in Computing Sciences (ICTCS)*. IEEE, 2017.
- [27] Hemalatha, R. J., et al. "Active Contour Based Segmentation Techniques for Medical Image Analysis." *Medical and Biological Image Analysis* (2018): 17
- [28] George, M. Jayesh, and D. Anto Sahaya Dhas. "Preprocessing filters for mammogram images: a review." 2017 *Conference on Emerging Devices and Smart Systems (ICEDSS)*. IEEE, 2017.
- [29] Bianconi, Francesco, and Antonio Fernández. "Rotation invariant co-occurrence features based on digital circles and discrete Fourier transform." *Pattern Recognition Letters* 48 (2014): 34-41.
- [30] Nayak, Tejas, et al. "Automatic segmentation and breast density estimation for cancer detection using an efficient watershed algorithm." *Data Analytics and Learning*. Springer, Singapore, 2019. 347-358.
- [31] Öztürk, Şaban, and Bayram Akdemir. "Application of feature extraction and classification methods for histopathological image using GLCM, LBP, LBGLCM, GLRLM and SFTA." *Procedia computer science* 132 (2018): 40-46.
- [32] Manavalan Radhakrishnan and Thangavel Kuttiannan, "Comparative Analysis of Feature Extraction Methods for the Classification of Prostate Cancer from TRUS Medical Images", *IJCSI International Journal of Computer Science Issues*, Vol. 9, no.1, January 2012.