BIOCYBERNETICS AND BIOMEDICAL ENGINEERING 40 (2020) 337–351

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

Simultaneous feature weighting and parameter determination of Neural Networks using Ant Lion Optimization for the classification of breast cancer

Dalwinder Singh[a], *Birmohan Singh*[a], *Manpreet Kaur*[b,*]

[a] *Department of Computer Science and Engineering, Sant Longowal Institute of Engineering and Technology, Longowal, Punjab, India*

[b] *Department of Electrical and Instrumentation Engineering, Sant Longowal Institute of Engineering and Technology, Longowal, Punjab, India*

ARTICLE INFO

Article history:
Received 1 August 2019
Received in revised form
1 November 2019
Accepted 13 December 2019
Available online 25 December 2019

Keywords:
Antlion optimization
Breast cancer
Feature weighting
Neural Networks

ABSTRACT

In this paper, feature weighting is used to develop an effective computer-aided diagnosis system for breast cancer. Feature weighting is employed because it boosts the classification performance more as compared to feature subset selection. Specifically, a wrapper method utilizing the Ant Lion Optimization algorithm is presented that searches for best feature weights and parametric values of Multilayer Neural Network simultaneously. The selection of hidden neurons and backpropagation training algorithms are used as parameters of neural networks. The performance of the proposed approach is evaluated on three breast cancer datasets. The data is initially normalized using *tanh* method to remove the effects of dominant features and outliers. The results show that the proposed wrapper method has a better ability to attain higher accuracy as compared to the existing techniques. The obtained high classification performance validates the work which has the potential for becoming an alternative to the other well-known techniques.

1. Introduction

Breast cancer is considered as a major health problem worldwide. It is the second most common type of cancer among women after lung cancer which causes more deaths [1]. The factors responsible for possibly developing breast cancer include family history, aging, genetic risk factors, menstrual periods, and obesity, but how and when these factors change the breast cells into cancerous is not known [2]. Although there is no procedure defined to prevent breast cancer, early detection is the only key for better prognosis and treatment [3]. Detection of early invasive cancer, before it reaches to the other vital organs, is essential to reduce breast cancer mortality. The survival chances for women from cancer depends upon the size of the tumor. The survival rate of

* *Corresponding author at*: Department of Electrical and Instrumentation Engineering, Sant Longowal Institute of Engineering and Technology, Longowal, Punjab, India.
E-mail addresses: dalwindercheema@outlook.com (D. Singh), birmohansingh@sliet.ac.in (B. Singh), aneja_mpk@yahoo.com (M. Kaur).
https://doi.org/10.1016/j.bbe.2019.12.004

women is 95 % for 1 cm, 85 % for 1–2 cm, and 60 % for 2–5 cm tumor size in the next ten years [4].

1.1. *Problem statement*

Worldwide, breast cancer is about 15 % of all types of cancers in women. In the United States, the estimated new cases of cancer will be 1,735,350, and the expected number of deaths will be 609,640 for 2018. The estimated cases of cancer in women are 878,980, of whom 266,120 are likely be breast cancer cases which will result in 40,920 deaths [5]. In India, total numbers of new cancer cases are expected to reach 1.73 million in 2020 from around 1.45 million in 2016, out of which more than 10 % cases will be related to breast cancer [6]. Although the cases of breast cancer are increasing, the trend also shows a reduction in mortality rate due to the widespread adoption of better diagnostic facilities and improvements made in breast cancer treatment [7]. However, it resulted in vast volumes of diagnostic data from the medical records, mammograms, ultrasounds and biopsies of screened women. Availability of the small number of the experts is also causing delay in the identification of cancer. Here, Computer Aided Diagnosis (CAD) systems can play an important role and increase the sensitivity of diagnosis. CAD systems for mammography are either used for a second opinion or as a visual tool to assist radiologists [8].

Several techniques have been developed for the classification of breast cancer. But still, there is a scope to design an appropriate method for developing and implementing a more effective diagnosis system for breast cancer. The motivation of this work is to make efficient use of information from the large volume of data generated so that that correct classification may enhance the treatment options. In this paper, feature weighting is used for obtaining better predictive models of Back-Propagation Neural Networks (BPNN) on breast cancer data. A wrapper-based approach is employed that uses Ant Lion optimization algorithm to search for the best weights of features along with the optimal values of neural network parameters. Initially, data is normalized to suppress the outliers and to deal with the numerically dominant features. Afterward, the proposed approach is applied to find the best predictive model by weighting features while selecting the optimal training algorithm and the number of hidden neurons required. To validate the effectiveness of proposed work, experiments are performed on Wisconsin original Breast Cancer (WBC), Wisconsin Diagnostic Breast Cancer (WDBC), and Breast Cancer Coimbra Dataset (BCCD) datasets that have been taken from the UCI Machine Learning Repository [9].

Related work is discussed in Section 2, and the dataset information is provided in Section 3. Section 4 covers feature weighting, neural networks, and explains the proposed wrapper method. Results, discussions, and comparisons with existing works are presented in Section 5, and Section 6 concludes the work.

2. Literature review

From literature, it has been observed that numerous methods have been proposed to classify the breast cancer dataset. Previous work on WBC, WDBC, and BCCD is reviewed in this section.

The work on the WBC includes: Quinlan [10] used C4.5 decision tree method, and Nauck and Kruse [11] used a neuro-fuzzy approach for the classification of data. Pena-Reyes and Sipper [12] proposed a fuzzy-genetic approach for distinguishing malignant and benign cancer samples. Chen and Hsu [13] developed genetic algorithm (GA) based approach for extracting the decision rules to build a decision-making model for NN. Übeyli [14] compared the performances of Support Vector Machines (SVM) and four different NNs. Akay [15] combined SVM with feature selection to achieve better accuracy. Karabatak and Ince [16] reduced data dimensions using the association rules and applied NN for classification with resultant features. Azar and El-Said [17] analysed the performance of six different SVMs and reported Linear programming SVM as the best classifier. Senapati et al. [18] initialized the centers and variances of radial basis function NN using K-PSO. Further, network weights were updated using the recursive least square method as a substitute for backpropagation. El-Baz [19] used rough set theory for selecting features and combined rules of the *k*-nearest neighbor classifier (KNN) for the diagnosis of breast cancer. Yang et al. [20] proposed the improved decision tree classifier, ID3, to enhance the disease prediction performance. Phan et al. [21] used the combination of feature weighting and SVM for improving the classification performance. The data was normalized with the min-max method, and GA was used to search the feature weights and parameters of SVM. Sudha [22] used the combination of rough set approach and GA to obtain higher performance using BPNN.

On the WDBC dataset, Guo and Nandi [23] used genetic programming for feature generation, modified Fisher linear discriminant analysis for feature extraction, and minimum distance for classification. Maglogiannis [24] applied five classifiers; SVM, Bayesian classifiers, and ANN to determine the classification performance where ANN outperformed the other four classifiers. Li et al. [25] proposed a Kernel Self-optimized Locality Preserving Discriminant Analysis for feature extraction and recognition. The data-dependent kernel was solved with the self-optimization method and obtained optimal projection matrix was used to reduce irrelevant features. Daoudi et al. [26] presented an Artificial Immune System for classification of benign and malignant breast cancer cases. Zheng et al. [27] proposed the combination of *k*-means clustering and SVM. The *k*-means clustering algorithm was used to determine the membership of each tumor separately by recognizing the hidden patterns of each tumor class individually. This membership are acted as new feature for learning the SVM model. Nilashi et al. [28] proposed a knowledge-based system using expectation maximization for clustering, Classification and Regression Trees (CART) for classification and principal component analysis to overcome the multicollinearity issue. Sayed et al. [29] proposed a chaotic crow search algorithm and Zhang et al. [30] employed the brain storm optimization algorithm for the feature subset selection while assessing performance with KNN classifier. Ghosh et al. [31] employed different machine learning algorithms for the selection of appropriate classifier. This comparative analysis includes SVM, KNN, Naïve Bayes (NB), Random Forests, Logistic Regression, and NN (Multilayer Perceptron (MLP)

and Deep Convolutional NN). Agarap [32] replaced the traditional classification function, *softmax*, with the rectified linear units in deep neural networks. The experiments were performed by pre-processing data with z-score normalization method.

The work which considered both datasets are: Krishnan et al. [33] proposed SVM based classifier for discriminating malignant and benign classes. Salama et al. [34] proposed an ensemble approach by fusing the decision tree (J48), MLP, NB, Sequential Minimal Optimization (SMO), and Instance-Based nearest neighbor classifiers. The ensemble method performed better on the original dataset whereas SMO outperformed ensemble approach on the diagnostic dataset. Lavanya [35] used CART classifier with feature selection and bagging technique to evaluate the performance. Ghosh et al. [36] used a neuro-fuzzy method that calculated a membership matrix considering the pattern-wise degree of memberships of breast cancer datasets. These matrix elements were input to a NN to learn the classification model. Wang et al. [37] proposed the SVM based ensemble learning algorithm to improve the classification performance. Their ensemble algorithm used six kernel functions with two SVM machines, thereby resulting in a total of twelve SVM classifiers. Liu et al. [38] proposed a hybrid feature subset selection approach in which information gain and a combination of simulated annealing (SA)-GA were utilized to identify the relevant features. The performance of selected features was measured with BPNN, KNN and SVM algorithms, and the highest accuracy was reported with the BPNN algorithm. Li et al. [39] proposed the smooth group regularization method to prune the nodes of Feedforward NN at the input layer. Additionally, pruning of input nodes also helps to reduce the dimensions of the input data.

On the BCCD dataset, Silva Araújo et al. [40] used fuzzy neural networks for the prediction of breast cancer. Three types of membership functions, unineuron with Gaussian, andneuron with triangular, and orneuron with Gaussian, are used to assess the performance using four features only. Andneuron with triangular membership function performed better than rest of the functions and five other classifiers. Singh [41] investigated the impact of various feature subset selection methods and machine learning algorithms on the prediction of breast cancer. Both filter and wrapper methods were used to rank the features whose performance was measured on six types of SVM. After selecting the best performing features, the performance was further measured on five other machine learning algorithms. Abdel-Basset et al. [42] proposed a wrapper based feature selection approach by improving the Grey Wolf optimization algorithm with two-phase mutation. The experiments were performed on both WDBC and BCCD datasets, and the performance of selected features was measured on the KNN classifier.

Ontiveros-Robles and Melin [43] worked on the development of a general type-2 fuzzy inference system for computer-aided diagnosis systems. In another work [44], they used a shadowed type-2 fuzzy inference system due to its good approximation and less computational cost as compared to previously used general type-2 fuzzy inference system. In both works, the experiments were performed on all three breast cancer datasets used in this study.

3. Datasets

In this paper, three UCI machine repository datasets have been used. The Wisconsin original Breast Cancer Database, from the University of Wisconsin Hospitals, Madison, has 699 instances which were taken from Fine Needle Aspirates of human breast tissue. There are nine attributes for each record to identify it as benign or malignant. The instances with the missing values (16) are discarded from the dataset. Therefore, only 683 cases have considered in this work containing 444 instances of the benign type and rest 239 of the malignant type. More information is available at [37,38].

The second dataset used in this study is the Wisconsin Diagnostic Breast Cancer Database. The dataset contains 569 samples belonging to two classes, namely, malignant and benign. There are 357 benign cases and 212 malignant cases. The digital images of a fine-needle aspirate of the patient's breast tissues were used to calculate features. Ten real-valued features were extracted from the cell nuclei of the breast tissues. Then, the mean, the standard error, and the mean of the three largest values were calculated from all features. Therefore, the total features obtained from all images were 30. Refer to [37,38] for more information.

The third dataset used in this study is the Breast Cancer Coimbra Dataset which was prepared at University of Coimbra. This dataset has 116 samples out of which 52 are healthy volunteers while rest of the women had cancer as confirmed from the positive mammography followed by histological examination. From each participant, nine clinical, demographic, and anthropometric features are extracted to predict the presence of breast cancer. More information about the dataset is available at [45].

4. Simultaneous feature weighting and parameter optimization

In this section, the concept of feature weighting and Multilayer feedforward NNs are described. Further, the proposed wrapper method based on Ant Lion Optimization is presented.

4.1. *Feature weighting*

Feature weighting works on the idea that some features are more important than others to solve a classification problem, and therefore, the contribution of each feature in pattern classification should be different. A feature weighting algorithm assigns higher weights to relevant features and lower weights to less relevant and redundant features. It is different from feature subset selection since the latter approach works for data that has entirely relevant or irrelevant features only. In fact, feature subset selection is the subset of feature weighting approach where the weights are restricted to 0 and 1 only. Therefore, feature weighting achieves equivalent or higher classification accuracy than feature subset selection. The weights are assigned according to the discrimination ability of features and these can be linear, polynomial or used as an arbitrary function [46]. The real-valued weights are assigned to features within the given search range.

In this work, a linear weight assignment approach is used to improve the accuracy in which the original features are multiplied with weights values. Suppose a data (D) having N instances and d features is represented as $\{x_{i,n}, y_n | n \in N \, and \, i \in d\}$ where x is the training data and y represents the corresponding labels. The linearly feature weighted data (x') will be given as follows:

$$x_i' = W_i \cdot x_i \tag{1}$$

where x_i is the i^{th} feature and W_i is the corresponding weight of i^{th} feature. It changes the feature space of the classification problem by expanding the space of highly weighted features while shrinking the space of nominally weighted features.

It is a challenging continuous search problem as the complexity of the weighting problem increases with an increase in features. Therefore, the wrapper based methods, which employ optimization algorithms, are more popular than the filter methods [21,47,48] for finding the feature weights. Another advantage is that these methods help to tune the weights according to the classifier that may vary from classifier to classifier. This approach has been used by various authors [21,47–50] to obtain better predictive models for many classification problems and combined it with various machine learning algorithms such as k -NN [47,48], Naïve Bayes [51] and Support Vector Machines [21].

4.2. Neural networks

Artificial neural networks are the most popular machine learning algorithms that have been used extensively in many application areas because of their higher classification performance [52]. Artificial neural networks are universal function approximation algorithms that can model linear as well as non-linear data with desired accuracy [53]. The multilayer perceptron is the type of artificial neural network, a three-layered feedforward network that is simple and has well-established generalization properties. The network consists of the input layer, hidden layers and output layer where learning occurs with the notion of weights between each layer. The data is fed into the classifier through the input layer to produce output at the hidden layer as follows:

$$H_j = \emptyset_H \left(\sum_{i=1}^{d} w_{j,i} x_i' \right) \tag{2}$$

where d is the number of inputs neurons, $w_{j,i}$ represents the weights between input and hidden layer, $\emptyset_H$ represents the activation function at the hidden layer. The classification decision came at the output layer which is calculated as follows:

$$y_k = \emptyset_O \left(\sum_{j=1}^{M} w_{k,j} H_j \right) \tag{3}$$

where M represents neurons at the hidden layer, $w_{k,j}$ represents weights between the hidden and output layer and $\emptyset_O$ represents the activation function at the output layer. In this work, *tan-sigmoid* and *softmax* activation functions are used at the hidden layer and output layer respectively. The *tan-sigmoid* function is given as follows:

$$\emptyset_H() = \frac{2}{1 + e^{-2}} - 1 \tag{4}$$

and the *softmax* function is given as follows:

$$\phi_o() = \frac{e}{\sum_{i=1}^{c} e} \tag{((5))}$$

In this work, we have trained the neural network using the backpropagation learning algorithm. This algorithm updates the weights backward, from the output layer, through the hidden layer to the input layer by propagating the difference between the desired output and the predicted output. It is given as follows:

$$E = \frac{1}{2} \sum_{n=1}^{N} (y_n - z_n)^2 \tag{6}$$

where y_n represents the predicted value, z_n represents the actual value, and E is the mean squared error (MSE). The weights vector is adjusted at each iteration until the error is minimized or total iterations are completed. Various training algorithms are suggested in the literature for finding weights at hidden layers. In this paper, we are considering a number of the hidden neurons and different training algorithms that are optimized simultaneously with feature weights to maximize the classification accuracy on breast cancer data. Table 1 shows the detail of the backpropagation training algorithms considered in the study.

4.3. Proposed work

The proposed work is a data preparation approach that makes the data more relevant to achieve high classification performance. The block diagram of the proposed work is shown in Fig. 1. Initially, data is pre-processed with *tanh* method which helps to tackle the problem of outliers as well as feature dominance in terms of numerical range. Then, feature weighting is applied which alters the feature space to achieve the higher classification. Moreover, the parametric values of BPNN are also considered in the work because different training algorithms may converge to different solutions, and the number of neurons to be used with training algorithm has no standard rule. These parametric issues affect the performance of classifier [54] and need to be optimized for the considered data. Therefore, these two parameters are also optimized simultaneously along with the feature weights. Otherwise, the computational cost will be enormous, if exhaustive search of parametric values is performed while searching for weights. The combination of feature weights and parameters increases the search space tremendously to search for the optimal solution which may cause the entrapment of optimization algorithms in local optima is possible. Due to this, Ant Lion optimization is used due to its better convergence as compared to other algorithms. This algorithm has been used widely for searching feature subsets as well as

Table 1 – Training Algorithms for Neural Network used in the work.

S. No.	Algorithm	Details
1	Conjugate gradient (CG) algorithms	A line search is performed along the direction of the conjugate gradient to determine the step size at each iteration. Many variants of CG algorithms have been proposed based on a search for the next direction from the previous directions.
1.1	Conjugate Gradient with Powell/ Beale Restarts (CGPBR)	The algorithm resets the search direction to the negative of the gradient periodically.
1.2	Fletcher-Reeves Conjugate Gradient (FRCG)	The search direction is given as the ratio of the square of the current gradient to the square of the previous gradient.
1.3	Polak-Ribiére Conjugate Gradient (PRCG)	The search direction is given as the inner product of the change in the previous gradient with a current gradient which is divided by the square of the previous gradient.
1.4	Scaled Conjugate Gradient (SCG)	Performing a line search at each iteration of conjugate gradient algorithms is computationally expensive. This algorithm avoids the line searching by combining the model-trust region with the conjugate gradient approach.
2	Resilient Backpropagation (RB)	Uses the sign of the derivative only to update the weights. The magnitude of the derivative is ignored to improve the convergence speed.
3	Variable Learning Rate Backpropagation (VLRB)	The algorithm combines the gradient descent momentum and adaptive learning for updating the weights.
4	Levenberg-Marquardt (LM)	This algorithm uses a combination of the Gauss-Newton method and gradient descent methods. The approximation of Hessian matrix is achieved using Jacobian matrix.
5	Quasi-Newton Methods	These methods rely on the approximation of Newton's direction without using second-order derivatives. These methods optimize faster but require higher computations due to the calculation of Hessian Matrix.
5.1	Broyden–Fletcher–Goldfarb–Shanno Quasi-Newton (BFGS)	The algorithm uses line search and updates the approximation of the Hessian matrix at each iteration for convergence.
5.2	One Step Secant (OSS)	This algorithm combines the conjugate gradient and BFGS. The computation cost is reduced by considering the previous Hessian matrix as the identity matrix.

parameters of classifiers and training the neural networks due to its good exploration and exploitation capabilities [55–57].

4.3.1. *Data normalization*

In this work, the data is preprocessed with data normalization method to mitigate the effects of dominant features and outliers before its evaluation on BPNN. *tanh* normalization is a widely used data pre-processing method that is selected because when feature weighting is utilized, it improves the classification performance higher as compared to other popular normalization methods such as z-score and min-max [58]. Initially, this method transforms the data by using Hampel estimators [59] which are based on influence function (Ψ). This function is given as follows:

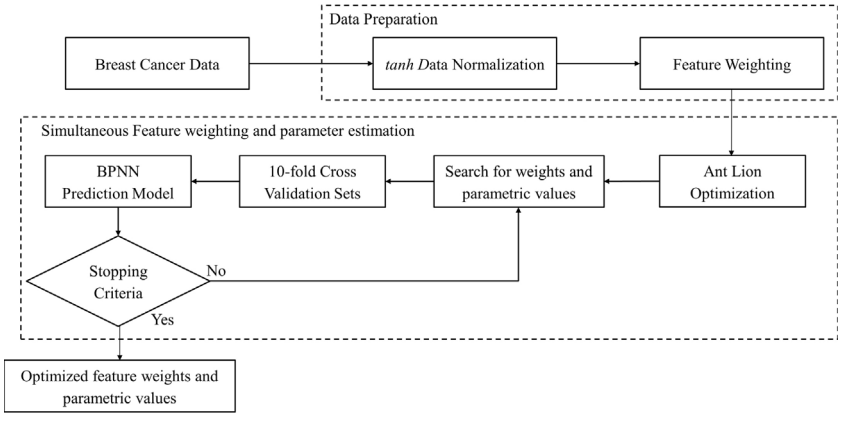

Fig. 1 – Flowchart of the proposed work for optimizing feature weights and parameters of BPNN simultaneously.

$$\Psi(p)=\begin{cases} p & 0\le|p|\le u_1 \\ u_1\times sign(p) & u_1<|p|\le u_2 \\ u_1\times sign(p)\times\left(\dfrac{u_3-|p|}{u_3-u_2}\right) & u_2<|p|\le u_3 \\ 0 & |p|>u_3 \end{cases} \tag{7}$$

where $p_i = x_{i,n} - med(x_i)$, med is a function for calculating the median value, $sign(p)=1$ for $p\ge 0$ and $sign(p)=-1$ for $p<0$. This function suppresses the influence of values at the tails of data distribution with the help of parameters u_1, u_2 and u_3 whose values are set to $quantile_{0.7}(|p|)$, $quantile_{0.85}(|p|)$ and $quantile_{0.95}(|p|)$ respectively [60]. Then, data is normalized by measuring the statistical properties of each feature. It includes computing the mean and standard deviation of each transformed feature to obtain the normalized feature ($\hat{x_i}$). It is given as follows:

$$\hat{x_i}=\frac{1}{2}\left\{tanh\left(0.01\left(\frac{x_i-\mu_i^G}{\sigma_i^G}\right)\right)+1\right\} \tag{8}$$

where μ^G and σ^G represent the mean and standard deviation of Hampel estimated i^{th} feature respectively.

4.3.2. Ant Lion optimization

The Ant Lion Optimization (ALO) [61] algorithm is used to search for optimal feature weights and parametric values of neural networks simultaneously. The algorithm is inspired by the hunting behavior of antlions that catch their prey, ants, by digging a cone-shaped pit in the sand. The size of the trap mainly depends upon the hunger level of the antlion. Higher the hunger, larger will be the pit and vice-versa. This hunting process of the antlion in nature was modeled mathematically to solve the optimization problems. This algorithm has been used for searching the feature subsets [56,62], training neural networks [56], and parameter of classifiers [55] which is related to this work.

The algorithm works by assuming P antlions and ants in d dimensional problem space which is defined as: $ALO=\{AL_i, A_i | i\in P$, where P represents the population of antlions and ants. The initial positions of the antlions (AL) are determined randomly within the given search space. Further, considering the nature of the movement of ants for finding food in the given search space as stochastic, their movement was modeled as a random walk. A random variable is defined as follows:

$$s=\begin{cases} 1 & if(r>0.5) \\ 0 & else \end{cases} \tag{9}$$

where r is generated in the interval of [0,1] randomly. The random walk of the ants is calculated as a normalization of the cumulative sum of total iterations using the following equation:

$$X^t=[0, cumsum(2s(t_1)-1)\ cumsum(2s(t_2)-1),\ldots, cumsum(2s(t_T)-1)] \tag{10}$$

where T is maximum iteration and t represents the current iteration. Min-max normalization approach is used to keep the position of the ants in the predefined search space. The newer calculated positions will easily surpass the boundary conditions because of no relation between the resultant sum of eq. 10 and boundary conditions. Therefore, the newer position of the ants are modified as follows:

$$X^t=\frac{(X^t-a_i)\times(d_i^t-c_i^t)}{b_i-a_i}+c_i^t \tag{11}$$

where a_i represents the lower bound and b_i represents the upper bound of the random walk's in i^{th} dimension, c_i^t represents the lower bound and d_i^t represents the upper bound of the i^{th} dimension at t^{th} iteration. To reflect the trapping of ants in the antlion's pit, the lower and upper bounds of the dimensions are updated. For ant, A_h^t at the current iteration, it is given as follows:

$$c_i^t=AL_m^t+c_i^t \text{ and } d_i^t=AL_m^t+d_i^t \tag{12}$$

where AL_m^t represents the position of the selected m^{th} antlion around which ants are trapping. The process of catching the prey by sliding it downwards inside the pit has reflected through an adaptive decrease in radius of ants' random walk and given as follows:

$$c_i^t=\frac{c_i}{I} \text{ and } d_i^t=\frac{d_i}{I} \tag{13}$$

where c_i is the lower bound and d_i is the upper bound of i^{th} dimension for the given problem and I represents the ratio which is further defined as follows:

$$I=L\frac{t}{T}$$

where L is the constant parameter that controls the level of exploitation. This parameter is modified to meet the requirements of the feature weights which are required up to a few decimal places only. The consumption of ants by antlions has represented in terms of fitness value. If position of the consuming ant is better than the existing position of the antlion, the newer position of the antlion will update to the current position of an ant which is given as follows:

$$AL_m^{t+1}=A_h^t \quad if\ fit(AL_m^t)<fit(A_h^t) \tag{14}$$

The fitness function is classification accuracy, and it is given as follows:

$$fit=\frac{Correctly\ \ Classified\ \textbf{Instances}}{Total\ Instances} \tag{15}$$

In each iteration, the best antlion is determined from all antlions that have the highest fitness. Therefore, the elite antlion affects the movement of all ants in each iteration. The ants tend to move towards the elite antlion to achieve higher fitness. However, it may lead to trapping of antlions in local optima. Therefore, elitism is used to maintain the best solution at any iteration during optimization. The elitism is realized using elite (ET) antlion obtained so far and an antlion selected through the roulette wheel (RW) strategy. The newer position of ant is the average of both antlions and is given as follows:

$$A_h^t=\frac{AL_{ET}^t+AL_{RW}^t}{2} \tag{16}$$

Vector	1	2	3	d	$d+1$	$d+2$
Solution Vector	x_1	x_2	…	x_d	p_1	p_2

Fig. 2 – Solution representation for simultaneous feature weighting and parameters of NN classifier.

where AL^t_{ET} represents the best antlion AL^t_{RW} represents the roulette wheel selected antlion at t^{th} iteration. Fig. 2 illustrates the solution representation of the ALO for the simultaneous feature weighting and parameter determination. The solution vector is continuous in nature, where values are real, and integers type only. The weights for features are real (x_d) values whereas the number of neurons (p_1) and selection of training functions (p_2) are integer type values.

The pseudocode of Ant Lion Optimization for the simultaneous search of feature weights and BPNN parameters is given in Algorithm 1.

Algorithm 1: Feature weighting and parameter optimization using ALO

Input: Training data (D), Parameters of Neural Networks, Population of Antlions and Ants (P), Total Iterations (T)
Output: Optimal Feature weights and parametric values (AL_{ET})

Begin
Randomly initialize antlion population
Calculate fitness of antlions using BPNN
Find Elite Antlion (AL_{ET})
Set current iteration (t) = 1
while $t \le T$
  for **i** = 1 to P
    Select two antlions (*ET* and *RW*)
    Create a normalized random walk around two selected antlions (eq. 11)
    Determine the position of i^{th} ant (eq. 16)
  endfor
  Calculate fitness of ants using BPNN
  If an ant is fitter than any antlion, then replace the position of antlion (eq.14)
  If an antlion is fitter than elite antlion, then update elite antlion:
  $AL^{t+1}_{ET} = AL^t_j \quad if\ fit(AL^t_{ET}) < fit(AL^t_j)$
endwhile
end

5. Results and discussions

The proposed wrapper-based method is applied to Wisconsin medical datasets to improve their classification performance. The proposed methodology is implemented in the MATLAB® environment using a system having i7 1.80 GHz processor and 16 GB RAM. The datasets are evaluated using a 10-fold cross-validation procedure to obtain the classification accuracy. Furthermore, we have performed 10 independent executions to avoid the random success of a classifier.

5.1. *Parameter settings*

The features weights are searched within the interval of 0–1, and details of BPNN parameters for experiments are provided in Table 2. Two parameters; Hidden neurons and selection of training algorithms are considered for optimizing the network in this work. The maximum number of neurons for learning the data is different for each dataset because of the number of features. More hidden neurons are used for a higher number of features and vice-versa.

The details of parameter settings for the ALO algorithm during experimentation is provided in Table 3. These parameters are kept fixed for each independent execution to obtain the results. The maximum iterations and population size are considered from [56]. The values of parameter L is adapted according to the need of feature weighting.

5.2. *Classification accuracy with Full feature set*

In this section, the obtained results are provided as well as discussed. Initially, experiments are performed on a full set of features considering hidden neurons and training algorithms. Then, we present the results of the proposed approach which are compared with the full feature set and state-of-the art work for validation.

Fig. 3 shows the accuracy plots of WBC data for the full set of features considering hidden neurons and training algorithms. Overall, fluctuations of accuracy are observed for training the neural networks with all algorithms as number of neurons increase. The maximum accuracy of 97.16 % is achieved with SCG algorithm using 8 neurons while VLRB algorithm with 29 neurons obtains a minimum accuracy of 92.83 %. It is also noticeable for all training algorithms that with the increase in the number of hidden neurons, accuracy does not improve. In fact, LM, RB, SCG, OSS, and VLRB algorithm shows the decline of accuracy with the increase in hidden neurons. VLRB algorithm fails to converge while utilizing certain number of neurons (such as at 2 and 29) as it is typical of this algorithm for overshooting the good

Table 2 – Parameters of neural networks considered in the study.

Parameter	Value	
Hidden neurons	[1,30]	WBC and BCCD
	[1,50]	WDBC
Training algorithms	1	LM
	2	BFGS
	3	RB
	4	SCG
	5	CGPBR
	6	FRCG
	7	PRCG
	8	OSS
	9	VLRB

Table 3 – Experimental settings for the Ant Lion Optimization.

Parameter	Setting	
Maximum Iterations	200	
Number of Antlions and Ants	20	
I (Controlling exploration and exploitation)	5	$t > 0.50T$
	10	$t > 0.75T$
	20	$t > 0.90T$
	50	$t > 0.95T$

solutions when learning rate is increased to speed up the convergence. Moreover, this algorithm has several parameters whose fine-tuning is required to attain convergence [63]. Therefore, with certain neurons, the algorithm fails to converge resulting in significant decline of accuracy.

The accuracy plots of WDBC data using two parameters of neural networks are shown in Fig. 4. High fluctuations of accuracy are observed with all algorithms as number of neurons increase, and using a single neuron results in lower performance for most of the algorithms. The maximum accuracy of 97.30 % is attained by using CGPBR algorithm with 37 neurons while VLRB algorithm achieves a minimum accuracy of 70.96 % using 29 neurons. LM and BFGS algorithms show gain in accuracy while SCG algorithm shows a decline of accuracy with the increase of hidden neurons. The accuracy remains at same level for the SCG, CGPBR, FRCG, PRCG, and OSS while RB and VLRB show significant variations in outcomes at certain number of neurons. RB algorithm fails to converge when 41 and 46 neurons are used which may be due to the occurrence of larger step size that misses the minimum point [64,65]. In the case of VLRB, previously mentioned reasons cause a decline in accuracy when certain number of neurons are used. Interestingly, lowest accuracy on both WBC and WDBC datasets is attained with 29 neurons using this algorithm.

The accuracy plots of Coimbra data using two parameters of neural networks are shown in Fig. 5. For all algorithms, high fluctuations of accuracy are observed with different neurons. The maximum accuracy of 72.37 % is attained by utilizing LM algorithm with 19 neurons while VLRB algorithm achieves a minimum accuracy of 56.29 % with single neuron. Except LM all other algorithms show gains of accuracy with the increase in hidden neurons. However, at specific number of neurons, sudden decline in accuracy is observed for the CGPBR (9), OSS (13), and VLRB (17 and 21) algorithms. While the reasons for non-convergence of VLRB are already mentioned earlier, OSS algorithm lacks in performance because descent direction is calculated based on rough approximation [66] whereas CGPBR may not converge even in the presence of accurate line search [67].

5.3. *Classification performance with proposed work*

Table 4 outlines the results obtained with the proposed wrapper-based method and also compares these with the full feature set. Three performance metrics, accuracy, sensitivity, and specificity, are used to show the efficiency of proposed method. In the table, minimum, maximum, mean and standard deviation of metrics are provided along with the hidden neurons and the best training algorithm. In case of all features, only minimum and maximum values of best training algorithms are presented over the 10 independent runs.

The significant gain in accuracy on all three datasets indicates the success of the proposed method. Also, the proposed method has improved the sensitivity higher in contrast to specificity for all datasets which shows its superiority for detecting women with breast cancer. For WBC data, the mean values of sensitivity and specificity are almost balanced (approx. 97.80 %), and the sensitivity is improved by more than 1 %. Similarly, for WDBC data, the optimized weighted models increase the chances of detecting cancer patients by improving the detection of malignant cases

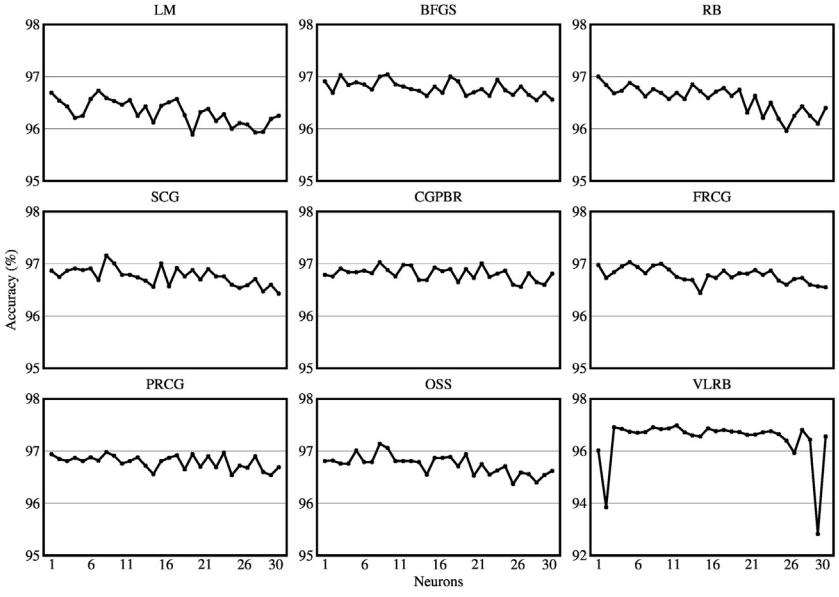

Fig. 3 – Average accuracy on WBC data for different hidden neurons and training algorithms.

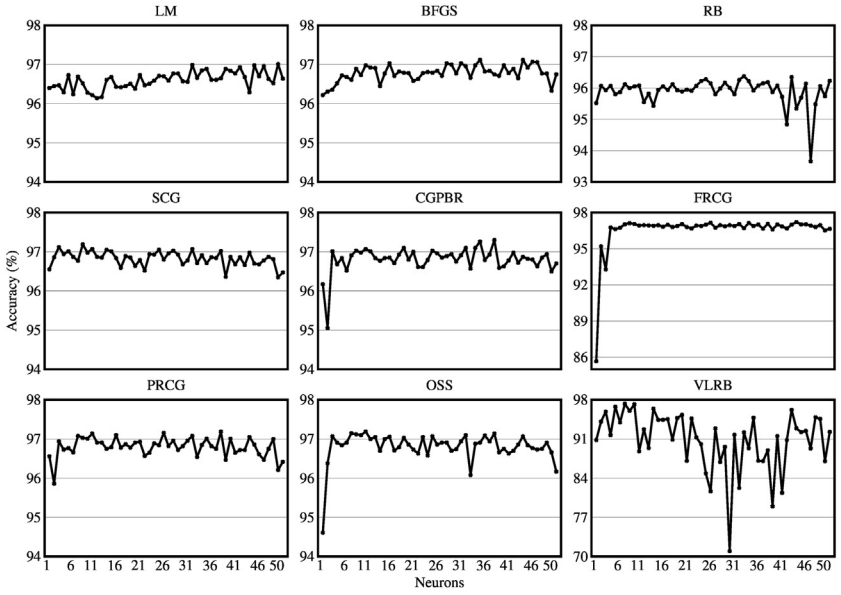

Fig. 4 – Average accuracy for WDBC data for different hidden neurons and training algorithms.

by more than 2 %. For BCCD data, sensitivity is increased by more than 11 %, while the accuracy is improved by 10 %.

BFGS, CGPBR, and LM emerged as the best training algorithms for WBC, WDBC, and BCCD datasets respectively. The average number of hidden neurons required is 12, about 29, and about 13 for original, diagnostic, and Coimbra datasets respectively. It implies that the optimized model requires less hidden neurons and will be more efficient in contrast to full feature set. Moreover, the low standard deviation of all datasets indicates the good convergence capabilities of ALO algorithm.

5.4. *Feature weights*

Fig. 6 presents the feature weights of the best model which are obtained with the proposed approach for all three datasets. Higher weight values signify the higher importance of the features in the learning process and vice-versa. The weights of the WBC data (Fig. 6(a)) indicate that six features contain reliable information when neural networks are used as classifiers. The rest of the features, clump thickness, uniformity of cell shape, and mitoses, have very little contribution in learning process. The weights of BCCD data (Fig. 6(b)) indicate

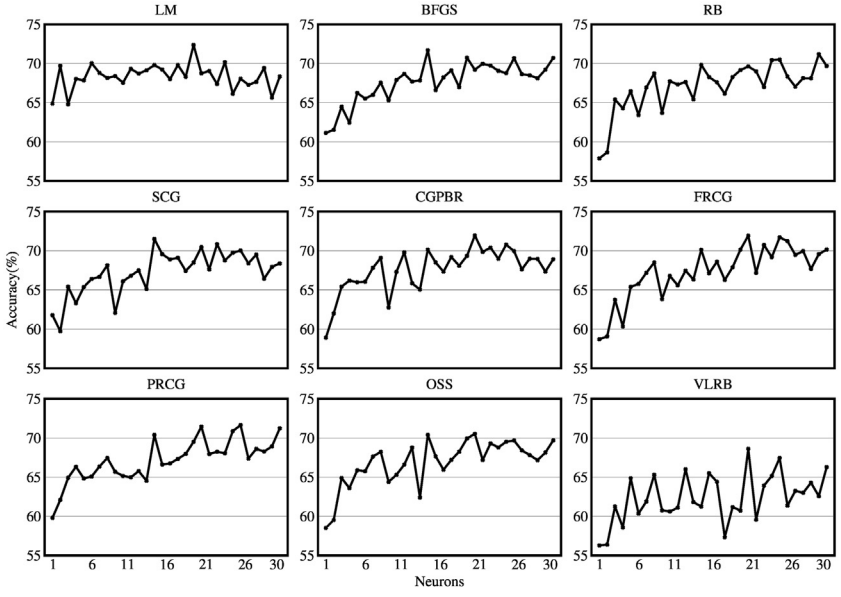

Fig. 5 – Average accuracy for BCCD data for different hidden neurons and training algorithms.

Table 4 – The results of the proposed wrapper-based method and comparison with the full feature set.

Datasets	Methods	Performance Metrics	Min.	Max.	Mean	Std. Dev.	Hidden Neurons	Training algorithm
WBC	Full Feature Set	Accuracy	96.43 %	97.16 %	–	–	8 (Max.)	SCG
		Sensitivity	95.27 %	96.52 %	–	–		
		Specificity	97.05 %	97.50 %	–	–		
	Feature Weighting	Accuracy	97.66 %	98.25 %	97.85 %	0.17 %	12 (Mean)	BFGS (3 times)
		Sensitivity	97.48 %	97.90 %	97.82 %	0.38 %		
		Specificity	97.75 %	98.42 %	97.86 %	0.26 %		
WDBC	Full Feature Set	Accuracy	95.05 %	97.30 %	–	–	37 (Max.)	CGPBR
		Sensitivity	89.53 %	94.22 %	–	–		
		Specificity	98.34 %	99.13 %	–	–		
	Feature Weighting	Accuracy	98.24 %	98.61 %	98.37 %	0.12 %	29.10 (Mean)	CGPBR (3 times)
		Sensitivity	96.69 %	96.75 %	96.43 %	0.56 %		
		Specificity	99.17 %	99.72 %	99.52 %	0.40 %		
BCCD	Full Feature Set	Accuracy	64.80 %	72.37 %	–	–	19 (Max.)	LM
		Sensitivity	70.76 %	73.02 %	–	–		
		Specificity	57.23 %	71.87 %	–	–		
	Feature Weighting	Accuracy	80.82 %	85.28 %	82.79 %	1.38 %	12.80 (Mean)	LM (7 times)
		Sensitivity	84.29 %	86.90 %	86.05 %	3.70 %		
		Specificity	76.33 %	83.00 %	78.73 %	2.65 %		

that two features, glucose, and insulin have significant contribution in learning the best model. Age, body mass index, homeostasis model assessment index, and resistin features have minor influence in classification, whereas Leptin and monocyte chemoattractant protein-1 features do not contribute to the learning process at all due to their near-zero weight values. The feature weights for WDBC data (Fig. 6(c)) are higher than 0.4 which signifies the importance of all features in the classification. Moreover, weights of 9 features are near to unity which shows that their normalized values played an important role in classification.

5.5. *Comparison with previous works*

The results attained with the proposed work are compared with the work of the other researchers to validate the superiority of our work. Table 5 shows the classification accuracy based comparison of the proposed approach for original and diagnostic datasets in a decade long research. For the sake of fair comparison, the previous works that have used 10-fold cross-validation to determine their classification performance are reported. In the table, the information about the normalization method and dimensionality reduction, if used, is also provided along with the classifier.

It can be seen that the proposed approach has outperformed the existing approaches where our method attains 97.85 % mean accuracy in the WBC dataset as compared to previous work that has the highest accuracy of 97.80 %. Similarly, in the case of WDBC dataset, the proposed method attains 98.37 % mean accuracy, showing improvement as compared to the previous work that has 98.18 % highest

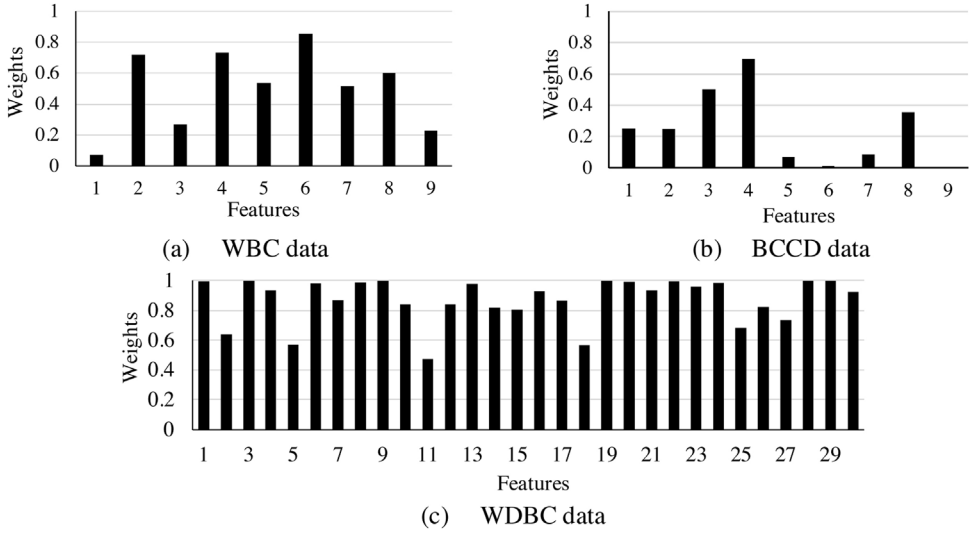

Fig. 6 – Feature weights of the best model for three datasets.

Table 5 – Comparison of the proposed work with previous research based on 10-fold cross-validation.

Author	Year	Data Normalization	Approach (Features + Classifier)	Mean Accuracy	
				WBC	WDBC
Maglogiannis [24]	2009	No	FULLSET + ANN	–	97.90 %
Li [25]	2011	No	FULLSET + KSLPDA	–	97.33 %
Salama [34]	2012	No	FULLSET+(ENSEMBLE and SMO)	97.28 %	97.71 %
Zheng [27]	2014	No	FS + SVM	–	97.38 %
Ghosh [36]	2016	Yes	FULLSET + NFS	97.80 %	95.90 %
Haque [68]	2016	No	FULLSET + ENSEMBLE	–	97.14 %
Ahmadi [69]	2016	Yes	FR + SVM	–	93.53 %
Nilashi [28]	2017	No	FR + CART-FC	–	93.20 %
Sayed [29]	2017	No	FS + KNN	–	90.28 %
Sudha [22]	2017	No	FS + BPNN	95.31 %	–
Yang [20]	2017	No	FULLSET + ID3	95.28 %	–
Phan [21]	2017	Yes (Min-Max)	FW + SVM	97.28 %	–
Shahnaz [31]	2017	No	FULLSET + MLP	–	97.89 %
Shahnaz [31]	2017	No	FULLSET + DNN	–	98.06 %
Agarap [32]	2018	Yes (z-score)	FULLSET + DNN	–	87.96 %
Zhang [30]	2018	No	FS + KNN	–	98.18 %
Wang [37]	2018	No	FULLSET + ENSEMBLE	97.10 %	97.68 %
Dorado [70]	2018	No	FS + MLP	–	97.00 %
Naik [71]	2019	No	FS + OGCNN	–	93.54 %
Liu [38]	2019	No	FS + BPNN	96.30 %	97.50 %
Rao [72]	2019	No	FS + GBDT	–	92.80 %
Li [39]	2019	No	FULLSET + FFNN	93.75 %	90.94 %
Ontiveros-Robles [43]	2019	No	FULLSET + GT2-FIS	97.01 %	95.74 %
Ontiveros-Robles [44]	2019	No	FULLSET + ST2-FIS	97.25 %	96.26 %
Abdel-Basset [42]	2019	No	FS + KNN	–	94.82 %
Proposed Approach	2019	Yes (*tanh*)	FW + BPNN	97.85 %	98.37 %

NFS = Neuro-Fuzzy System, AIS = Artificial Immune System, FS = Feature Selection, FC = Fuzzy Classifier, MDC = Minimum Distance Classifier, FR = Feature Reduction, KSLPDA = Kernel Self-optimized Locality Preserving Discriminant Analysis, FW = Feature Weighting, KNN=*k*-Nearest Neighbor, DNN = Deep Neural Networks, OGCNN = One-pass Generalized Classifier Neural Network, GBDT = Gradient Boosting Decision Tree, GT2-FC = General Type-2 Fuzzy Inference System and ST2-FIS = Shadowed Type-2 Fuzzy Inference System.

accuracy. Table 6 shows the comparison of proposed work as compared to existing works on the BCCD dataset. The proposed approach attained the mean accuracy of 82.79 % which is higher than the previous work having 82.39 % mean accuracy. Three of the works have reported very less accuracy (less than 76 %) as compared to this work.

From Tables 5 and 6, it has been observed that although data normalization has been useful for constructing accurate predictive models, this step was missing from most of the previous works on all three datasets. Many researchers have also worked without dimensionality reduction methods which also result in lower performance. Some have used feature selection to select the relevant features, but this approach is unable to compete with the feature weighting. Therefore, the proposed work outperformed the existing works, some of which includes the combination of feature weighting and SVM, ensemble-based approaches, and two deep learning approaches. These are the latest trends in the field of machine learning research to attain better classification models. Furthermore, the feature weighting also surpasses the feature selection based approaches which is the expected outcome. Hence, the proposed approach provides an effective learning model that can be used in breast cancer diagnosis system.

Table 6 – Comparison of proposed work with previous works on BCCD data based on 10-fold cross-validation.

Author	Year	Data normalization	Approach (Features + Classifier)	BCCD
Silva Araújo [40]	2019	No	FULLSET + FNN	81.04 %
Singh [41]	2019	No	FS + MG-SVM	82.39 %
Ontiveros-Robles [43]	2019	No	FULLSET + GT2-FIS	75.95 %
Ontiveros-Robles [44]	2019	No	FULLSET + ST2-FIS	75.45 %
Abdel-Basset [42]	2019	No	FS + KNN	73.63 %
Proposed Approach	2019	Yes (*tanh*)	FW + BPNN	82.79 %

FS = Feature Selection, FNN = Fuzzy Neural Networks, MG-SVM = Medium Gaussian SVM, GT2-FC = General Type-2 Fuzzy Inference System, ST2-FIS = Shadowed Type-2 Fuzzy Inference System and, KNN=*k*-Nearest Neighbor.

5.6. *Analysis of training algorithms for breast cancer data*

In this work, nine backpropagation training algorithms have been utilized to measure the performance on breast cancer datasets. The experiments involving all features show that these algorithms are susceptible to lack of convergence depending upon the number of neurons which represents unstable learning, which may be due to lack of parameter tuning or internal working of the algorithms. Therefore, number of neurons and training algorithms are optimized simultaneously along with the weights to attain the best solution without making efforts to improve the backpropagation training algorithms or fine-tuning of their internal parameters.

Furthermore, it is difficult to generalize the choice of backpropagation training algorithm for breast cancer prediction as the nature of features (in statistical terms) varies with the source (images, signals and clinical) and type (texture, color, wavelet, spatial and shape) of extraction as it does in case of datasets used in this study. The experimental results support this fact since the optimal choice of training algorithms is different for each dataset. The scaled conjugate gradient is good for WBC data, while Conjugate gradient with Powell/Beale restarts emerged as the best choice for WDBC dataset. Levenberg-Marquardt is best for BCCD dataset while Powell/Beale restarts algorithm fails to converge.

5.7. *Implication of the proposed work*

The practical breast cancer diagnosis is a complex, long, dynamic, specialized process where doctors require different information at each stage for better decision making. The stages for diagnosis includes early detection, early diagnosis, clinical diagnosis, imaging diagnosis and pathological diagnosis [73,74]. With the help of machine learning algorithms, many prediction models for this problem have been presented by using data from these stages, such as clinical, demographic and anthropometric data [41,45], mammographic images [75,76], ultrasound images [77–79] pathological images [80–82] molecular and genomic data [83–85] or their combination [86]; however, there is no improvement in interpretive accuracy. The problem of overdiagnosis, false positives and false negatives is common while relying on screening of breast cancer. It has been found that both overdiagnosis and false positives account for 10 % of screened women whereas false negatives occur in approximately 15–20 % of screened women [87]. Although 90 % of false positives are resolved, it has an adverse impact on participants due to unnecessary anxiety, additional imaging work-up, and breast biopsy. Overdiagnosis and false negatives remain the important limitation of existing research works due to their lower performance. In fact, lower sensitivity has been a major bottleneck in breast cancer screening [80,88].

The proposed method will have a significant contribution in practical breast cancer diagnosis as it increases the sensitivity of malignant cases while improving the classification accuracy. The sensitivity on WBC, WDBC, and BCCD datasets has been improved by more than 1 %, 2 %, and 11 % respectively. This work aims to recommend feature weighting acts as a better substitute for improving classification performance in contrast to conventional approach that involves feature extraction followed by classification process directly as well as feature subset selection.

6. Conclusion

Several studies have been conducted on modeling procedures for breast cancer classification, but the selection of the appropriate technique has been a challenge for the researchers for developing an effective diagnosis system. In this paper, a wrapper method is proposed for the classification of breast cancer data that uses Ant Lion optimization for searching feature weights and optimal parametric values of the neural networks simultaneously. The data is normalized with *tanh* method before its evaluation. The obtained results show gains in accuracy and help to identify the number of neurons required at the hidden layer and the best training algorithm for each dataset. While comparing our outcomes with previously reported results, the improvements are observed which validates the efficacy of proposed methodology. In conclusion, the combination of data normalization with feature weighting and parameter determination yields fruitful results in achieving the high-performance neural networks classification models. In future work, the proposed work will be extended to other medical problems as well as to deep neural networks.

Studies in humans and animals

No human and animals are involved in this work.

CRediT authorship contribution statement

Conceptualization, Methodology, Software, Writing - original draft, Validation.Supervision, Visualization, Investigation, Resources.Supervision, Writing - review & editing.

REFERENCES

[1] Ericeira DR, Silva AC, De Paiva AC, Gattass M. Detection of masses based on asymmetric regions of digital bilateral mammograms using spatial description with variogram and cross-variogram functions. Comput Biol Med 2013;43:987–99.

[2] Jerez-Aragonés JM, Gómez-Ruiz JA, Ramos-Jiménez G, Muñoz-Pérez J, Alba-Conejo E. A combined neural network and decision trees model for prognosis of breast cancer relapse. Artif Intell Med 2003;27:45–63.

[3] Razek NMA, Yousef WA, Mustafa WA. Microcalcification detection with and without CAD system (LIBCAD): a comparative study. Egypt J Radiol Nucl Med 2013;44:397–404.
[4] Beheshti SMA, AhmadiNoubari H, Fatemizadeh E, Khalili M. An efficient fractal method for detection and diagnosis of breast masses in mammograms. J Digit Imaging 2014;27:661–9.
[5] Society AC. Cancer facts & figures 2018; 2018 (accessed November 29, 2018) https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf.
[6] Report. Over 17 lakh new cancer cases in India by 2020: ICMR; 2016 (accessed April 16, 2018) http://icmr.nic.in/icmrsql/archive/2016/7.pdf.
[7] Ganesan K, Acharya RU, Chua CK, Min LC, Mathew B, Thomas AK. Decision support system for breast cancer detection using mammograms. Arch Proc Inst Mech Eng Part H J Eng Med 1989-1996 2013;227:721–32.
[8] Papadopoulos A, Fotiadis DI, Costaridou L. Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques. Comput Biol Med 2008;38:1045–55.
[9] Frank A, Asuncion A. UCI machine learning repository, 2008. Irvine, CA: University of California, School of Information and Computer Science. Irvine, CA; 2010. p. 0 [http://archive.ics.uci.edu/ml].
[10] Quinlan JR. Improved use of continuous attributes in C4. 5. J Artif Intell Res 1996;4:77–90.
[11] Nauck D, Kruse R. Obtaining interpretable fuzzy classification rules from medical data. Artif Intell Med 1999;16:149–69.
[12] Pena-Reyes CA, Sipper M. A fuzzy-genetic approach to breast cancer diagnosis. Artif Intell Med 1999;17:131–55.
[13] Chen T-C, Hsu T-C. A GAs based approach for mining breast cancer pattern. Expert Syst Appl 2006;30:674–81.
[14] Übeyli ED. Implementing automated diagnostic systems for breast cancer detection. Expert Syst Appl 2007;33:1054–62.
[15] Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. Expert Syst Appl 2009;36:3240–7.
[16] Karabatak M, Ince MC. An expert system for detection of breast cancer based on association rules and neural network. Expert Syst Appl 2009;36:3465–9.
[17] Azar AT, El-Said SA. Performance analysis of support vector machines classifiers in breast cancer mammography recognition. Neural Comput Appl 2014;24:1163–77.
[18] Senapati MR, Panda G, Dash PK. Hybrid approach using KPSO and RLS for RBFNN design for breast cancer detection. Neural Comput Appl 2014;24:745–53.
[19] El-Baz AH. Hybrid intelligent system-based rough set and ensemble classifier for breast cancer diagnosis. Neural Comput Appl 2015;26:437–46.
[20] Yang S, Guo J-Z, Jin J-W. An improved Id3 algorithm for medical data classification. Comput Electr Eng 2017.
[21] Phan AV, Nguyen MLe, Bui LT. Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems. Appl Intell 2017;46:455–69. http://dx.doi.org/10.1007/s10489-016-0843-6
[22] Sudha M. Evolutionary and neural computing based decision support system for disease diagnosis from clinical data sets in medical practice. J Med Syst 2017;41:178.
[23] Guo H, Nandi AK. Breast cancer diagnosis using genetic programming generated feature. Pattern Recognit 2006;39:980–7.
[24] Maglogiannis I, Zafiropoulos E, Anagnostopoulos I. An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. Appl Intell 2009;30:24–36.
[25] Li J-B, Pan J-S, Chen S-M. Kernel self-optimized locality preserving discriminant analysis for feature extraction and recognition. Neurocomputing 2011;74:3019–27.
[26] Daoudi R, Djemal K, Benyettou A. An immune-inspired approach for breast cancer classification. Int Conf Eng Appl Neural Networks 2013;273–81.
[27] Zheng B, Yoon SW, Lam SS. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Syst Appl 2014;41:1476–82.
[28] Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L. A knowledge-based system for breast cancer classification using fuzzy logic method. Telemat Informatics 2017;34:133–44.
[29] Sayed GI, Hassanien AE, Azar AT. Feature selection via a novel chaotic crow search algorithm. Neural Comput Appl 2017;1–18.
[30] Zhang X, Zhang Y, Gao H, He C. A wrapper feature selection algorithm based on brain storm optimization. Int Conf Bioinspired Comput Theor Appl 2018;308–15.
[31] Shahnaz C, Hossain J, Fattah SA, Ghosh S, Khan AI. Efficient approaches for accuracy improvement of breast cancer classification using wisconsin database. Humanit Technol Conf (R10-HTC) 2017;792–7. IEEE Reg. 10, 2017.
[32] Agarap AF. Deep learning using rectified linear units (relu). ArXiv Prepr ArXiv180308375; 2018.
[33] Krishnan MMR, Banerjee S, Chakraborty C, Chakraborty C, Ray AK. Statistical analysis of mammographic features and its classification using support vector machine. Expert Syst Appl 2010;37:470–8.
[34] Salama GI, Abdelhalim M, Zeid MA. Breast cancer diagnosis on three different datasets using multi-classifiers. Breast Cancer (WDBC) 2012;32:2.
[35] Lavanya D, Rani KU. Ensemble decision tree classifier for breast cancer data. Int J Inf Technol Converg Serv 2012;2:17.
[36] Ghosh S, Biswas S, Sarkar DC, Sarkar PP. Breast cancer detection using a neuro-fuzzy based classification method. Indian J Sci Technol 2016;9:1–15.
[37] Wang H, Zheng B, Yoon SW, Ko HS. A support vector machine-based ensemble algorithm for breast cancer diagnosis. Eur J Oper Res 2018;267:687–99.
[38] Liu N, Qi E-S, Xu M, Gao B, Liu G-Q. A novel intelligent classification model for breast cancer diagnosis. Inf Process Manag 2019;56:609–23.
[39] Li F, Zurada JM, Wu W. Smooth group L1/2 regularization for input layer of feedforward neural networks. Neurocomputing 2018;314:109–19.
[40] Silva Araújo VJ, Guimarães AJ, de Campos Souza PV, Silva Rezende T, Souza Araújo V. Using resistin, glucose, age and bmi and pruning fuzzy neural network for the construction of expert systems in the prediction of breast cancer. Mach Learn Knowl Extr 2019;1:466–82.
[41] Singh BK. Determining relevant biomarkers for prediction of breast cancer using anthropometric and clinical features: a comparative investigation in machine learning paradigm. Biocybern Biomed Eng 2019;39:393–409.
[42] Abdel-Basset M, El-Shahat D, El-henawy I, de Albuquerque VHC, Mirjalili S. A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection. Expert Syst Appl 2020;139112824.
[43] Ontiveros-Robles E, Melin P. Toward a development of general type-2 fuzzy classifiers applied in diagnosis problems through embedded type-1 fuzzy classifiers. Soft Comput n.d.:1–17.

[44] Ontiveros-Robles E, Melin P. A hybrid design of shadowed type-2 fuzzy inference systems applied in diagnosis problems. Eng Appl Artif Intell 2019;86:43–55.
[45] Patricio M, Pereira J, Crisóstomo J, Matafome P, Gomes M, Seiça R, et al. Using Resistin, glucose, age and BMI to predict the presence of breast cancer. BMC Cancer 2018;18:29.
[46] Wettschereck D, Aha DW, Mohri T. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. Artif Intell Rev 1997;11:273–314.
[47] AlSukker A, Khushaba R, Al-Ani A. Optimizing the k-nn metric weights using differential evolution. Proc Int Conf Multimed Comput Inf Technol 2010;89–92.
[48] Shirazi F, Rashedi E. Feature weighting for cancer tumor detection in mammography images using gravitational search algorithm. 6th Int Conf Comput Knowl Eng 2016;310–3.
[49] Yongxiong W, Li K. Feature and weight selection using tabu search for improving the recognition rate of duct anomaly. IEEE ROBIO 2014;2163–8.
[50] Barros AC, Cavalcanti GD. Combining global optimization algorithms with a simple adaptive distance for feature selection and weighting. Proc IEEE Int Jt Conf Neural Networks 2008;3518–23.
[51] Jiang L, Li C, Wang S, Zhang L. Deep feature weighting for naive Bayes and its application to text classification. Eng Appl Artif Intell 2016;52:26–39.
[52] Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015;61:85–117.
[53] Galeshchuk S. Neural networks performance in exchange rate prediction. Neurocomputing 2016;172:446–52.
[54] Kriesel D. A brief introduction on neural networks; 2007 (accessed October 30, 2019) http://www.dkriesel.com.
[55] Tharwat A, Hassanien AE. Chaotic antlion algorithm for parameter optimization of support vector machine. Appl Intell 2017;1–17.
[56] Singh D, Singh B. Hybridization of feature selection and feature weighting for high dimensional data. Appl Intell 2018;1–17.
[57] Heidari AA, Faris H, Mirjalili S, Aljarah I, Mafarja M. Ant Lion optimizer: theory, literature review, and application in multi-layer perceptron neural networks. Nature-Inspired Optim.. Springer; 2020. p. 23–46.
[58] Singh D, Singh B. Investigating the impact of data normalization on classification performance. Appl Soft Comput 2019105524.
[59] Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. Robust statistics: the approach based on influence functions, 114. John Wiley & Sons; 2011.
[60] Jain A, Nandakumar K, Ross A. Score normalization in multimodal biometric systems. Pattern Recognit 2005;38:2270–85. http://dx.doi.org/10.1016/j.patcog.2005.01.012
[61] Mirjalili S. The ant lion optimizer. Adv Eng Softw 2015;83:80–98. http://dx.doi.org/10.1016/j.advengsoft.2015.01.010
[62] Emary E, Zawbaa HM, Hassanien AE. Binary ant lion approaches for feature selection. Neurocomputing 2016;213:54–65. http://dx.doi.org/10.1016/j.neucom.2016.03.101
[63] Demuth HB, Beale MH, De Jess O, Hagan MT. Neural network design. Martin Hagan; 2014.
[64] Bailey TM. Convergence of Rprop and variants. Neurocomputing 2015;159:90–5.
[65] Hinton G, Srivastava N, Swersky K. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. Cited On 2012;14:8.
[66] Saito K, Nakano R. Partial BFGS update and efficient step-length calculation for three-layer neural networks. Neural Comput 1997;9:123–41.
[67] Dai Y-H, Liao L-Z, Li D. On restart procedures for the conjugate gradient method. Numer Algorithms 2004;35:249–60.
[68] Haque MN, Noman MN, Berretta R, Moscato P. Optimising weights for heterogeneous ensemble of classifiers with differential evolution. IEEE Congr Evol Comput 2016;233–40.
[69] Ahmadi A, Afshar P. Intelligent breast cancer recognition using particle swarm optimization and support vector machines. J Exp Theor Artif Intell 2016;28:1021–34.
[70] Dorado H, Cobos C, Torres-Jimenez J, Jimenez D, Mendoza M. A proposal to estimate the variable importance measures in predictive models using results from a wrapper. Int Conf Min Intell Knowl Explor 2018;369–83.
[71] Naik AK, Kuppili V, Edla DR. Efficient feature selection using one-pass generalized classifier neural network and binary bat algorithm with a novel fitness function. Soft Comput 2019;1–13.
[72] Rao H, Shi X, Rodrigue AK, Feng J, Xia Y, Elhoseny M, et al. Feature selection based on artificial bee colony and gradient boosting decision tree. Appl Soft Comput 2019;74:634–42.
[73] Gu D, Liang C, Zhao H. A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis. Artif Intell Med 2017;77:31–47.
[74] Gradishar WJ, Anderson BO, Balassanian R, Blair SL, Burstein HJ, Cyr A, et al. NCCN guidelines insights: breast cancer, version 1.2017. J Natl Compr Cancer Netw 2017;15:433–51.
[75] Gardezi SJS, Elazab A, Lei B, Wang T. Breast Cancer detection and diagnosis using mammographic data: systematic review. J Med Internet Res 2019;21e14464.
[76] Ting FF, Tan YJ, Sim KS. Convolutional neural network improvement for breast cancer classification. Expert Syst Appl 2019;120:103–15.
[77] Liu L, Li K, Qin W, Wen T, Li L, Wu J, et al. Automated breast tumor detection and segmentation with a novel computational framework of whole ultrasound images. Med Biol Eng Comput 2018;56:183–99.
[78] Virmani J, Agarwal R, et al. Effect of despeckle filtering on classification of breast tumors using ultrasound images. Biocybern Biomed Eng 2019;39:536–60.
[79] Chiao J-Y, Chen K-Y, KY-K Liao, Hsieh P-H, Zhang G, Huang T-C. Detection and classification the breast tumors using mask R-CNN on sonograms. Medicine (Baltimore) 2019;98.
[80] Nahid A-A, Kong Y. Histopathological breast-image classification using local and frequency domains by convolutional neural network. Information 2018;9:19.
[81] Bardou D, Zhang K, Ahmad SM. Classification of breast cancer based on histology images using convolutional neural networks. IEEE Access 2018;6:24680–93.
[82] Alirezazadeh P, Hejrati B, Monsef-Esfahani A, Fathi A. Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images. Biocybern Biomed Eng 2018;38:671–83.
[83] Ortiz AG, Muñoz AS, Parrado MRC, Pérez MÁ, Entrena NR, Dominguez AR, et al. Deciphering HER2 breast cancer disease: biological and clinical implications. Front Oncol 2019;9:1124.
[84] Søkilde R, Persson H, Ehinger A, Pirona AC, Fernö M, Hegardt C, et al. Refinement of breast cancer molecular classification by miRNA expression profiles. BMC Genomics 2019;20:503.

[85] Mahmoodian H, Ebrahimian L. Using support vector regression in gene selection and fuzzy rule generation for relapse time prediction of breast cancer. Biocybern Biomed Eng 2016;36:466–72.
[86] WY-Y Wu, Tabar L, Tot T, Fann C-Y, AM-F Yen, SL-S Chen, et al. Imaging biomarkers as predictors for breast Cancer death. J Oncol 2019;2019.
[87] Seely JM, Alhassan T. Screening for breast cancer in 2018—what should we be doing today? Curr Oncol 2018;25: S115.
[88] Trister AD, Buist DSM, Lee CI. Will machine learning tip the balance in breast cancer screening? JAMA Oncol 2017;3:1463–4.