Inteligencia Artificial 24(67) (2021), 147-156
doi: 10.4114/intartif.vol24iss67pp147-156

INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

Design of Ensemble Classifier Model Based on MLP Neural Network For Breast Cancer Diagnosis

Rahmad Syah[1,*], Siswi Wulandari[2], Arbansyah[3], Amin Rezaeipanah[4]

[1] DS & CI Research Group, Universitas Medan Area, Medan, Indonesia, rahmadsyah@staff.uma.ac.id
[2] Obstetrics Department, Universitas Kediri, Kediri, Indonesia, siswiwulandari@unik-kediri.ac.id
[3] Department of Informatics, Faculty of Science and Technology, Universitas Muhammadiyah Kalimantan Timur, Samarinda 75124, Indonesia, arb381@umkt.ac.id
[4] Department of Computer Engineering, University of Rahjuyan Danesh Borazjan, Bushehr, Iran, amin.rezaeipanah@gmail.com

Abstract Nowadays, breast cancer is one of the leading causes of death women in the worldwide. Substantial support for breast cancer awareness and research funding has helped created advances in the diagnosis and treatment of breast cancer. Data mining techniques have a growing reputation in the medical field because of high diagnostic capability and useful classification and they can help breast cancer diagnosis. In this paper, a Multi-Layer Perceptron Neural Network (MLP-NN) based on Evolutionary Algorithms (EA) is used to automatically classify breast cancer. Here, EA is used to tune MLP parameters such as optimal features, hidden layers, hidden nodes and weights. Ensemble models is a machine learning approach to combine multiple other single models in the prediction process. To improve the performance of the classification model, we use an Intelligent Ensemble Classification method based on MLP, named IEC-MLP. The proposed method was evaluated for the samples of the Wisconsin Breast Cancer Dataset (WBCD) by stacked generalization technique. The proposed method was evaluated for the samples of the Wisconsin database by stacked generalization technique. Experimental results show the advanced performance of the IEC-MLP with ensemble classifiers compared to other algorithms. Accordingly, IEC-MLP was better than GAANN and CAFS algorithms with classification accuracy of 98.74%.

Keywords: Breast Cancer, Ensemble Classifier, Neural Network, Evolutionary Algorithm.

1 Introduction

Breast cancer is one of the most common types of cancer in the worldwide and is the second leading cause of death among women after lung cancer [1]. Each year, the American Cancer Society (ACS) estimates the number of new cancers and deaths in the United States based on the latest data collected. According to ACS, worldwide studies reported 1,762,450 new cancer cases and 606,880 cancer deaths in 2020, are projected to occur in the United States [2]. In recent decades, research on cancer has led to stunning advances in diagnosis and treatment with the appearance of new technologies [2, 3]. An important result of these advances has led to the creation of big data of cancer and made available to the research community, where they can be used to analyse and provide diagnostic models [3]. Due to the time consuming the process of diagnosis by physicians, the use of computer applications can significantly reduce the workload of the human agent [4].

The cancer data analysis can be achieved by data mining and machine learning techniques [4-6]. The main purpose of data mining and machine learning techniques is to create a model with capable of classifying and predicting the outcome of unknown input data according to previous observations [4]. These techniques are effective choices for aggregation, data partition, validation data, and feature selection of the cancer big data, where they can be used to extract the underlying relationships between biological parameters and disease prediction. In general,

ISSN: 1137-3601 (print), 1988-3064 (on-line)

classification and prediction problems have a fundamental role in medical decision making. In this paper, an intelligent ensemble classification [5] method based on multi-layer perceptron neural network (IEC-MLP) is proposed for breast cancer diagnosis. The principal restriction of MLP-NN is related to its parameter settings. Whereas, high dimensions of input features can increase the time and model complexity. The number of hidden layers as well as hidden nodes must be set by the user. Furthermore, due to the manner the weights are trained, it cannot ensure that global minima have been achieved. Therefore, the objective of presenting the IEC-MLP method is to tune the MLP-NN parameters, where evolutionary algorithms are used for this purpose.

MLP-NN is efficient in finding nonlinear associations between input vectors and output vectors, based on training sets. Classification, diagnosis, prediction, and control are typical problems of this nature. Therefore, MLP is suitable for modelling cancer data.

The main contribution of this paper is as follows:

- Improvement of MLP neural network with a homogeneous ensemble classification method
- Simultaneous optimization of all model parameters such as optimal features, hidden layers, hidden nodes and weights
- Comparison of optimization accuracy of model parameters with several evolutionary algorithms

The rest of this paper is organized as follows: Section 2 presents a background of the MLP-NN classifier used in this study, the two evolutionary algorithms and ensemble classification concepts. Section 3 discusses related works to breast cancer diagnosis. Section 4 presents details of the IEC-MLP method for tuning MLP-NN parameters. The experimental results and discussions of the proposed method are reported in Section 5. Section 6 presents conclusions and future work.

2 Background

In this section, an overview of MLP-NN classifier used to build the homogeneous ensemble classification model, two evolutionary algorithms used to tune neural network parameters, and ensemble classification concepts are presented.

2.1 MLP neural network

Multi-Layer Perceptron Neural Networks (MLP-NNs) are supervised learning-based feed-forward neural networks used to solve regression and classification problems [6]. MLP-NN have at least three layers input, hidden and output. Assuming the classification problem, the number of nodes in the input and output layers depends on the number of features and the number of classes in the dataset, respectively. In addition, using MLP-NN requires setting parameters the number of hidden layers, the number of nodes in each hidden layer, the number of learning epochs and weights learning technique. The hidden layer nodes receive signals from the input layer nodes through weighted connections, and after calculating the activation function output, transmits the output to the output layer nodes.

2.2 Evolutionary algorithms

The performance of the MLP-NN classifier model depends mainly on its parameters configuration, where the value of these parameters can vary for different datasets [7]. Therefore, optimization of MLP neural parameters is essential to achieve reliable results [7]. In this paper, MLP-NN parameters are optimized using two evolutionary algorithms, where the aim is to compare the performance of these algorithms for modeling work.

Genetic Algorithm (GA): The algorithm was developed by John Holland during the 1960 and 1970 [8]. Essentially, GA is based on Darwin's theory of evolution, where it works on the basis of a slow gradual process with slight and slow changes in the solution. In GA, the potential solutions, called chromosomes. Also, GA uses operators Crossover and Mutation to produce the next generation.

Particle Swarm Optimization (PSO): The algorithm is a population-based stochastic optimization technique developed by Eberhart and Kennedy in 1995, inspired by the social behavior of bird droppings or fish training [9]. The particle refers to a solution and in each iteration, every particle is updated by following two 'best' values, i.e. 'pbest' and 'gbest'.

2.3 Ensemble classification concepts

Heretofore, many classifiers models have been proposed, however none of them has proved to be the best in all circumstances [1]. In general, the performance of a classifier model varies from one context to another, which makes them unstable [1, 11]. To deal with this problem, ensemble-based classifier methods have been developed. First, learning is done based on the single classifier models $c_1, c_2, \dots, c_q$ and then a meta-classifier is learned that combines the output of the single classifier models. Here, q refers to the total number of single classifier models. The main idea of these methods

is to weigh several single classifiers model and combine them to get a classification that outperforms any of them. Conceptually, the single classifier models that make up the ensemble classification are trained individually on all or part of the dataset. The output of single classifier models is a collection of different outputs that are combined by a specific technique. There are generally two types of ensemble learning, including homogeneous and heterogeneous [11]. The homogeneous type contains several single classifier models and creates a meta-model such as bagging and boosting. Whereas, the heterogeneous type combines at least two different single classifier methods. In this study, the homogeneous type is highlighted for the classification of breast cancer.

3 Related works

There are generally several methods for diagnosing breast cancer, such as physical exam, laboratory tests, imaging tests, and biopsy [2]. In addition to these methods, medical data mining has great potential for exploring the hidden patterns in this field [1]. This section surveys a number of recent data mining methods based on breast cancer diagnosis. Ahmad et al. used Artificial Neural Networks (ANN) and GA as a predictor of breast cancer [12]. They used GA to simultaneously optimize ANN parameters including weights, effective features subset and number of hidden nodes. Kabir et al. developed the CAFS algorithm based on ANN [13]. CAFS can determine the number of hidden nodes during the feature selection process. Moreover, this method uses a grouping technique to select features, where the correlation level [14] is the criterion of grouping. Salehi et al. developed novel data mining on breast cancer survivability using MLP-NN ensemble learners [15]. They used two ensemble learning techniques including stacked generalization and mixture of experts based on MLP-NN. Al-Shargabi et al. introduced an enhanced version of MLP-NN for breast cancer prediction [16].

Bashiri and Geranmayeh proposed a method for tuning ANN parameters using central composite design and GA [17]. Here, GA is applied to find the optimal value of the parameters. Alickovic and Subasi proposed a model based on the normalized MLP-NN for the classification of breast cancer with high accuracy [6]. Ibrahim and Shamsuddin used an MLP-NN based on enhanced non-dominated sorting GA (NSGA-II) to diagnose breast cancer [18]. Here, Pareto optimality method based GA is applied simultaneously to optimize the network structure and reduce the MLP-NN error rate. Kumar et al. focused on a learning model based on MLP Back Propagation neural Network (MLPBPN) to diagnose breast cancer [19]. They compared models of single layer perceptron, multi-layer perceptron, and RBF network. Shelomyanov and Poger examined three algorithms, including Generalized Linear Model (GLM), Support Vector Machine (SVM), and ANN, for the breast cancer diagnosis [20].

4 The proposed method

In general, there is no consensus on the classification method that can provide the best performance in any circumstance. However, ensemble classification methods recently attracted more attention. In this paper, homogeneous ensembles based on MLP-NN were developed, where the parameters values of the ensemble members are evaluated based on evolutionary algorithms. The proposed method consists of two stages parameters optimizations and ensemble classification. Here, GA and PSO are investigated for tuning parameters. These parameters include optimal features, hidden layers size, the number of nodes in each layer and weights values. In addition, the outputs of the classification methods are combined using the stacked generalization technique. Stacked generalization is an ensemble technique that uses a new model to learn how to best combine the predictions from several models trained on dataset. The architecture of proposed IEC-MLP method is depicted in Figure 1.

Generally, the process of diagnosing breast cancer is performed in three levels. In the first level, the entire training set is split into q blocks with equal size. Then in the second level, $c_i, \forall i = 1,2,\dots,q$ MLP classifier is trained on $b_i, \forall i = 1,2,\dots,q$ block. Here, each of the MLP-NNs used at this level is trained separately by the evolutionary algorithm. Therefore, samples of each block are entered into MLP-NNs, and output is obtained. Having finished the training phase of this level, a new training data is formed which is sent as an input to the meta-classifier of the third level. The meta-classifier is an MLP-NN that learns the mapping between each single classifier of second level with the actual labels.

In IEC-MLP method, the selected features subset (sf_i) and predicted samples label (p_i) obtained from c_i classifier are used as new data. Therefore, the features in the new dataset are $\{\cup (sf_1, sf_2, \dots, sf_q), p_1, p_2, \dots, p_q\}$, where symbol $\cup$ represent the union of sets. In addition, the real label of each sample is also recorded in this data. These newly fabricated data are prepared temporarily for training through the stacked generalization technique. This technique combines the performed predictions to create a meta-classifier in a new dataset. Having been extracting this dataset, the MLP meta-classifier is trained on it by an evolutionary algorithm. Accordingly, the best combination of MLP classifiers output of second level can be produced to predict the real samples label.

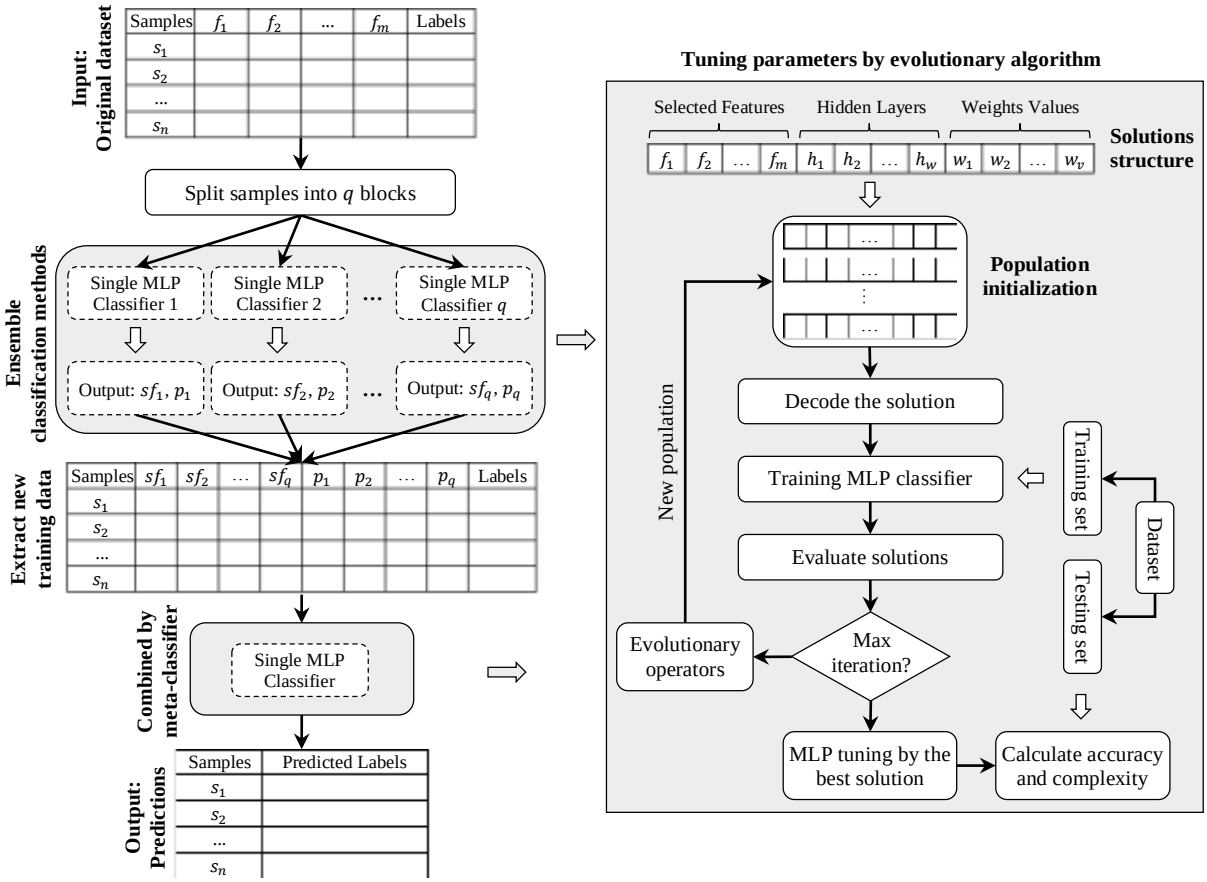

Figure 1. The architecture of the proposed IEC-MLP method.

All evolutionary algorithms basically use the same design for optimization in the sense that almost all of them share some similarity in their principle procedure. Generally, design of these algorithms can be divided into several components including representation, initial population, objective (performance) function, evolutionary operators and termination condition. Here, all components except evolutionary operators are the same in two GA and PSO.

The structure of the solutions representation describes the coding of each individual in the population. The proposed structure consists of three parts selected features, hidden layers and weights values, as shown in Figure 1. The first part is related to the selected features and consists of m elements of binary values, where m is the total number of features. The f_i element is indexed to the i-th feature. The second part of the solution is dedicated to hidden layers and contains the w element, where w represents the maximum number of hidden layers. Here, $h_j, \forall j = 1,2,\dots,w$ refers to the number of nodes in j-th hidden layer. When h_j is zero, this hidden layer is practically not considered in neural network design. In addition, to avoid high complexity in the MLP-NN, h_j is limited to $[0, 10]$ range. Finally, the third part deals with the values of the weights with the v element. $w_k \in [-1, +1], \forall k = 1,2,\dots,v$ is the weight of a communication in the MLP-NN. The number of weights for optimization is determined based on MLP-NN parameters.

Accordingly, the length of each solution in the proposed representation structure is $D = m + w + v$. Based on this structure, the initial population is randomly generated with N_{pop} solutions. In MLP-NN learning, classification accuracy measure on the training set is considered as the objective function. In addition, the maximum number of iterations ($Iter_{max}$) is used as the termination condition. The rest of this section presents the algorithms configuration used to design evolutionary operators.

4.1 GA configuration

In GA there are three types of operators including selection, crossover and mutation for population evolution. The details of these operators are as follows:

Selection: Here, the roulette wheel mechanism [21] is used for the selection operator, where selection probability of i-th chromosome is defined by Eq. (1).

$$P_i = \frac{obj_i}{\sum_{i=1}^{N_{Pop}} obj_i} \quad (1)$$

Where, obj_i is the fitness (objective function) of i-th chromosome and N_{Pop} refers to population size.

Crossover: Here, the Differential Evolution (DE) [21] crossover operator is used. According to the chromosome structure, DE is applied to each section separately with a probability of CR. The main idea of DE is to generate offspring chromosome (X^O) by adding the weighted difference vector between two population chromosomes X^1 and X^2, to a third chromosome, X^0, as given in Eq. (2).

$$x_i^O = \begin{cases} x_i^0 + F.(x_i^1 - x_i^2) & CR > Rnd(0,1) \\ x_i^0 & otherwise \end{cases} \quad (2)$$

Where, X^1 and X^2 are the parents selected for reproduction and X^0 is the best population chromosome. $F \in (0,1)$ is a coefficient to control population evolution. In addition, parameters within chromosomes are indexed with i, which runs for all elements of chromosome parts.

Using a constant coefficient of F can create big differences that produce values outside the problem space. Hence, it is suggested that F change dynamically with Eq. (3).

$$F = \frac{\alpha \,.\, Rnd(0,1)}{max(x_{i,g}^1, x_{i,g}^2)} \quad (3)$$

Where, α is scale factor smaller than 1 and index g indicates the generation to which the chromosome belongs.

Mutation: Here, the Bit Change (BC) mutation operator [21] is used. This operator is applied by the probability of MR for each element of the chromosome, where MR is the mutation rate. Meanwhile, BC is applied to the offspring chromosome from the DE output. The mutation operator is defined according to Eq. (4).

$$x_i^{OM} = \begin{cases} x_i^O + \Delta & MR > Rnd(0,1) \\ x_i^O & otherwise \end{cases} \quad (4)$$

Where, X^{OM} is the offspring after the mutation and Δ determines the scale of random changes in the chromosome. Here, $\Delta \in [-1, +1]$ is for the feature selection part, $\Delta \in [-2, +2]$ is for the hidden layer part and $\Delta \in [-0.01, +0.01]$ is for the weights part.

4.2 PSO configuration

Each particle in the PSO has a position and is represented by $X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,D}\}$, where D is the dimensionality of the search space. In addition, each particle has a velocity, which is represented as $V_i = \{v_{i,1}, v_{i,2}, \dots, v_{i,D}\}$. In each iteration, each particle updates its position and velocity according to 'pbest' and 'gbest', as given in Eq. (5) and (6).

$$v_k^{t+1} = \omega . v_k^{t+1} + c_1 . r_1 . (pbest_k - x_k^t) + c_2 . r_2 . (gbest - x_k^t) \quad (5)$$

$$x_k^{t+1} = x_k^t + v_k^{t+1} \quad (6)$$

Where, ω is inertia weight to control the velocity impact, c_1 and c_2 are acceleration constants and r_1 and r_2 are random values in $[0, 1]$ range. In addition, parameters within particles are indexed with t, which it refers to iteration in the evolutionary process.

PSO was originally developed for continuous problems. In [9] Binary PSO (BPSO) is presented which can be used for discrete optimization. The BPSO encodes the particles position by a binary string, where x_k^{t+1} is restricted to 0 or 1 based on the sigmoid function. BPSO updates the particle position with the Eq. (7).

$$x_k^{t+1} = \begin{cases} 1 & Rnd(0,1) < \dfrac{1}{1 + e^{-v_k^{t+1}}} \\ 0 & otherwise \end{cases} \quad (7)$$

Accordingly, v_k^{t+1} is transformed to $[0, 1]$ by a sigmoid limiting function. However, this equation is only suitable for feature selection part, since a feature mask is Boolean that '1' indicates feature selection and '0' otherwise. Here, Eq. (8) is proposed for the parts of the hidden layers and weights.

$$x_k^{t+1} = \begin{cases} x_k^t + \Delta & Rnd(0,1) < \dfrac{1}{1 + e^{-v_k^{t+1}}} \\ x_k^t & otherwise \end{cases} \quad (8)$$

Where, Δ is the change rate in the previous position to create a new position.

In addition, the velocity is limited by a predefined minimum and maximum velocity, i.e., $v_k^{t+1} \in [v_{min}, v_{max}]$. Here, $v_k^{t+1} \in [0,1]$ is for the feature selection part, $v_k^{t+1} \in [0,10]$ is for the hidden layer part and $v_k^{t+1} \in [-1,+1]$ is for the weights part.

5 Results and discussion

This section provides the basis for experimental evaluation of the proposed IEC-MLP method based on extensive studies. All simulations and experiments are performed on PC housing an Intel Core i7 processor at 3.3GHz, with 16GB DDRAM-III and Windows 10 64-bit operating system. In addition, the IEC-MLP method is implemented with MATLAB R2019a. Additionally, our IEC-MLP method includes several parameters that their values are set by trial and error technique. Setting the IEC-MLP method parameters is as follows: $q = 3$, $w = 5$, $N_{pop} = 35$, $Iter_{max} = 100$, $CR = 0.85$, $MR = 0.15$, $\alpha = 0.1$, $\omega = 0.75$, $c_1 = c_2 = 0.5$, $N_{imp} = 5$, $\theta = 0.5$, $\varphi = 0.3$, $\xi = 0.7$.

5.1 Breast cancer dataset

Wisconsin database is the most comprehensive source of information on cancer incidence, which available from archive.ics.uci.edu. According to previous studies, the Wisconsin Breast Cancer Dataset (WBCD) has been widely used as a benchmark dataset [22]. For this reason, this dataset is used in this study to analyze the proposed IEC-MLP method. The WBCD dataset contains 9 features and 699 instances, where 16 cases have missing values that are replaced by the mean value. In addition to the WBCD dataset, the performance of the proposed method is tested based on datasets of different sizes including WDBC, WPBC, SEER and Cervical.

5.2 Performance analysis

Performance analysis of the proposed method is performed based on statistical measures including Sensitivity, Specificity and Accuracy. Here, performance measures for all algorithms based on the 10-fold cross-validation technique are reported. With this technique, the dataset is randomly divided into 10 parts, of which 9 parts are used for training and the remaining part for testing. The cross-validation process is repeated 10 times alternately and finally the results are estimated on average. In addition to statistical measures, runtime (s), number of selected features, and number of network connections are used to performance analyze of the proposed method. In general, less number of neural network connections can make the classification model with fewer complexes. Here, Eq. (9) is used to calculate the number of connections, where reference [12] is mapped.

$$No.of\ connection = \alpha.\beta + \beta.Y + \beta + Y \tag{9}$$

Where, α, β and Y are the number of selected features, hidden layer size and output nodes size, respectively. Here, β is calculated by multiplying the number of nodes in all hidden layers. For example, in a network with two hidden layers with 2 and 3 nodes, respectively, the value of β is 2.

5.3 IEC-MLP method analysis

In this section, the performance analysis of the proposed IEC-MLP algorithm is performed based on various factors. First, a comparison of the performance of GA and PSO algorithms for tuning MLP-NN parameters is shown in Figure 2. This comparison is based on criteria of selected feature number, feature significance, and convergence.

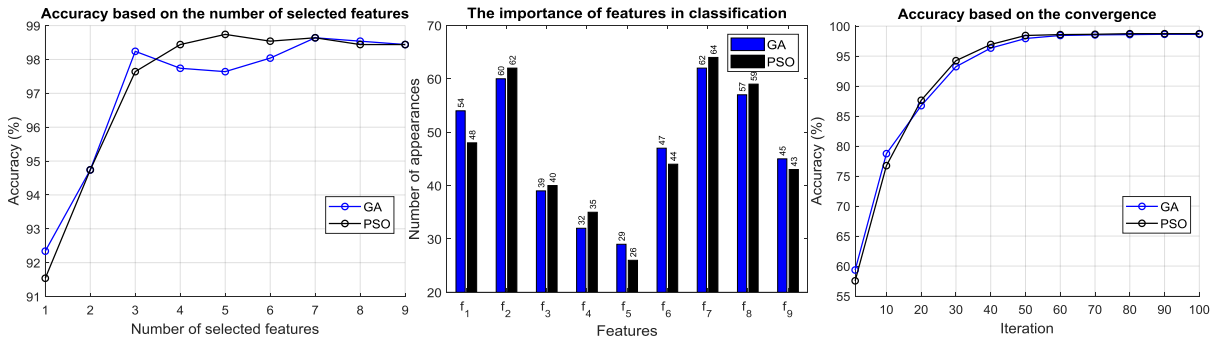

Figure 2. Comparison of the performance of evolutionary algorithms.

In IEC-MLP, the number of optimal features is determined automatically by an evolutionary algorithm. According to the results, GA have the best classify accuracy with 7 features. The best performance for GA is 98.72%. However, PSO

reported the best accuracy of 98.74% with 5 features. Only by feeding in those features that are really important can assist to classification model. Important features are estimated based on the number of appearances in the optimization process. The results for two evolutionary algorithms clearly show that features f_7, f_2 and f_8 are the most important features to diagnose breast cancer, respectively. However, the PSO has given more importance to these features. In addition, the convergence of GA and PSO is almost same and requires only 50 iterations. Overall, the comparison results for the two evolutionary algorithms show that the PSO achieved better results with faster convergence. Hence, in the continuation of this study, the results of this algorithm are used for comparison and experiments.

In the following, modeling performance for cancer diagnosis is compared with/without feature selection. Therefore all features are used in IEC-MLP method. The results of this comparison are shown in Table 1.

Table 1: Performance of IEC-MLP with and without features selection.

Criterion	With features selection	Without features selection
Hidden layer	2 (2 – 3)	2 (3 – 3)
No. of selected features	4.8	9.0
No. of connections	36.7	100.4
Accuracy (%)	98.74	98.25
Selected feature subset	f_1, f_2, f_6, f_7, f_8	f_1 - f_9 (All)

The results show the significant superiority of the IEC-MLP method with feature selection process. Here, IEC-MLP achieves 98.74% accuracy with only features f_1, f_2, f_6, f_7 and f_8. Therefore, irrelevant features can be removed without affecting learning performance. In addition, the network complexity is reduced with feature selection, due to the smaller size of the selected features. The best network produced consists of two hidden layers with 2 and 3 nodes, respectively. Here the number of connections without feature selection process has more than doubled.

The IEC-MLP method uses the stacked generalization technique to generate an ensemble classification model. Here, the performance of the stacked generalization technique is compared with techniques Dempster–Shafer and Mixture of Experts. The mixture of expert's technique is a popular method in ensemble learning, which works based on the divide-and-conquer method [23]. Ensemble classification model based Dempster–Shafer technique incorporates the different classification result by estimating the misclassification cost of each classifier [23]. The values of sensitivity, specificity and accuracy are presented in Table 2 through performing assessments for each technique.

Table 2: Comparison of the performance of ensemble classification techniques.

Techniques	Sensitivity	Specificity	Accuracy
Dempster–Shafer	98.00	99.12	98.61
Mixture of Experts	97.84	99.02	98.47
Stacked Generalization	98.13	99.35	98.74

As it can be seen, the stacked generalization technique obtained an average accuracy of 98.74%, which is considered a high accuracy. Dempster–Shafer technique is in the second rank with an accuracy of 98.61%. However, mixture of expert's technique achieved the lowest accuracy equal to 98.47%. Hence, in the continuation of this study, the results of stacked generalization technique are used for comparison and experiments.

The proposed IEC-MLP method is based on an intelligent ensemble classification approach, where it consists of several single classification models. The number of single classification models is important in determining the final accuracy of ensemble classification. Hence, Table 3 shows the final accuracy of the ensemble classification approach based on the different number of single classification models. The results clearly show that the best classification accuracy is achieved when 3 single classification models are used. However, the results with 4 single classification models are also promising.

The results show that the IEC-MLP method performs better than other algorithms except RF+GA on the WBCD dataset. According to the observed results, the IEC-MLP method on other cancer datasets provides promising and acceptable results. Concerning the WDBC dataset, the PCA+LDA+ANNFIS with 98.6% is the best, followed by the IEC-MLP. The IEC-MLP method ranks second place with 98.61% classify accuracy on the WPBC dataset, after the ANFIS+GA algorithm. The IEC-MLP method has better performance on the SEER dataset and offers better results than all algorithms. Due to the high number of samples in this dataset, the proper performance of the IEC-MLP method against

datasets with a high number of samples is proven. Concerning the Cervical cancer dataset, the IEC-MLP method reported the best performance after BDT+Hinslemann algorithm.

Table 3: Comparison of IEC-MLP method with other algorithms on the cancer datasets.

Dataset	Algorithms				
WBCD	NSGA-II [18]	RF+GA [25]	PSO-KDE [26]	SVM+AR [27]	IEC-MLP
	97.01	99.48	98.53	98.00	98.74
WDBC	WAUCE [28]	PSO-KDE [26]	RF+KNN+SVM [29]	PCA+LDA+ANNFIS [30]	IEC-MLP
	97.68	98.45	95.61	98.60	98.52
WPBC	ANFIS+GA [2]	PSO-KDE [26]	GA+CFS+RF [3]	Xgboost [4]	IEC-MLP
	90.00	98.45	86.36	97.71	98.61
SEER	MLP-experts [15]	MLP-SG [15]	Bayesian Belief [5]	Logistic Regression [5]	IEC-MLP
	84.29	83.86	75.90	75.50	88.19
Cervical cancer	Biopsy+Cytology [1]	BTD+Cytology [1]	BTD+Schiller [1]	BDT+Hinslemann [1]	IEC-MLP
	93.70	93.40	90.90	94.10	93.86

The overall results obtained from the proposed IEC-MLP method show that: (i) Using PSO algorithm to tuned MLP-NN parameters provides more accurate classification; (ii) The ensemble classification method provides more accurate classification than their single classification according to the optimization techniques used. (iii) The generalized stack technique is more effective for combining the output of combined classifications than techniques Dempster–Shafer and Mixture of Experts.

6 Conclusion and future works

In this paper, an intelligent ensemble classification method has been performed on the different cancer datasets in order to investigate the ability of MLP-NN for survivability prediction of cancer patients. The objective of the ensemble classification is to obtain reliable predictions based on the stacked generalization technique. Here, IEC-MLP method for parameter optimization of MLP-NN based on evolutionary algorithms is presented. The tuned MLP-NN proven that use of PSO to find the optimal parameters leads to the most accurate predictions. The main advantage of IEC-MLP is the simultaneous adjustment of optimal features, hidden layers, hidden nodes and weights parameters in the MLP-NN. Experiments and comparisons results show that the proposed IEC-MLP method is an effective and useful model for classification of breast cancer. The main purpose of this paper is to use our research in the real detections system and thus help physicians in making useful and better decisions in the future. In addition, the use of other neural network approaches such as LVQ instead of MLP-NN is suggested for future works.

References

[1] Rezaeipanah, A., & Ahmadi, G. (2020). Breast Cancer Diagnosis Using Multi-Stage Weight Adjustment In The MLP Neural Network. *The Computer Journal*, Oxford University Press.

[2] Bilalović, O., & Avdagić, Z. (2018, May). Robust breast cancer classification based on GA optimized ANN and ANFIS-voting structures. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 0279-0284). IEEE.

[3] Singla, S., Ghosh, P., & Kumari, U. (2019). Breast Cancer Detection using Genetic Algorithm with Correlation based Feature Selection: Experiment on Different Datasets. *International Journal of Computer Sciences and Engineering*, 7(4), 406-410.

[4] Ghobaei-Arani, M., Souri, A., Baker, T., & Hussien, A. (2019). ControCity: an autonomous approach for controlling elasticity using buffer Management in Cloud Computing Environment. *IEEE Access*, 7, 106912-106924.

[5] Khorsand, R., Ghobaei-Arani, M., & Ramezanpour, M. (2019). A self-learning fuzzy approach for proactive resource provisioning in cloud environment. *Software: Practice and Experience*, 49(11), 1618-1642.

[6] Khorsand, R., Ghobaei-Arani, M., & Ramezanpour, M. (2018). FAHP approach for autonomic resource provisioning of multitier applications in cloud computing environments. *Software: Practice and Experience*, 48(12), 2147-2173.

[7] Hosni, M., Idri, A., Abran, A., & Nassif, A. B. (2018). On the value of parameter tuning in heterogeneous ensembles effort estimation. *Soft Computing*, *22*(18), 5977-6010.

[8] Holland, J. H. (1992). Genetic algorithms. *Scientific american*, *267*(1), 66-73.

[9] Kennedy, J., & Eberhart, R. (1995, November). Particle swarm optimization. In *Proceedings of ICNN'95-International Conference on Neural Networks* (Vol. 4, pp. 1942-1948). IEEE.

[10] Atashpaz-Gargari, E., & Lucas, C. (2007, September). Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In *2007 IEEE congress on evolutionary computation* (pp. 4661-4667). IEEE.

[11] Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and systems magazine*, 6(3), 21-45.

[12] Ahmad, F., Isa, N. A. M., Hussain, Z., Osman, M. K., & Sulaiman, S. N. (2015). A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer. *Pattern Analysis and Applications*, *18*(4), 861-870.

[13] Kabir, M. M., Islam, M. M., & Murase, K. (2010). A new wrapper feature selection approach using neural network. *Neurocomputing*, *73*(16-18), 3273-3283.

[14] Kim, Z., & Nevatia, R. (1999). Uncertain reasoning and learning for feature grouping. *Computer Vision and Image Understanding*, *76*(3), 278-288.

[15] Salehi, M., Razmara, J., & Lotfi, S. (2020). A Novel Data Mining on Breast Cancer Survivability Using MLP Ensemble Learners. *The Computer Journal*, *63*(3), 435-447.

[16] Al-Shargabi, B., Alshami, F., & Alkhawaldeh, R. (2019). Enhancing Multi-Layer Perception for Breast Cancer Prediction. *International Journal of Advanced Science and Technology*, 130, 11-20.

[17] Bashiri, M., & Geranmayeh, A. F. (2011). Tuning the parameters of an artificial neural network using central composite design and genetic algorithm. *Scientia Iranica*, *18*(6), 1600-1608.

[18] Ibrahim, A. O., & Shamsuddin, S. M. (2018). Intelligent breast cancer diagnosis based on enhanced Pareto optimal and multilayer perceptron neural network. *International Journal of Computer Aided Engineering and Technology*, *10*(5), 543-556.

[19] Kumar, K. S., Sasank, V. V. S., Praveen, K. R., & Rao, Y. K. (2020). Multilayer Perceptron Back propagation Algorithm for Predicting Breast Cancer. In *Intelligent System Design* (pp. 41-53). Springer.

[20] Shelomyanov, L., & Poger, S. (2020). Applying three machine learning algorithms to three breast cancer diagnosis datasets. *Journal of Computing Sciences in Colleges*, *35*(8), 272-274.

[21] Ghalehgolabi, M., & Rezaeipanah, A. (2017). Intrusion Detection System Using Genetic Algorithm and Data Mining Techniques Based on the Reduction. *International Journal of Computer Applications Technology and Research*, *6*(11), 461-466.

[22] Wang, Z., Wang, R., Gao, J., Gao, Z., & Liang, Y. (2020). Fault recognition using an ensemble classifier based on Dempster–Shafer theory. *Pattern Recognition*, *99*, 107079.

[23] Yuksel, S. E., Wilson, J. N., & Gader, P. D. (2012). Twenty years of mixture of experts. *IEEE transactions on neural networks and learning systems*, *23*(8), 1177-1193.

[24] Onan, A. (2015). A fuzzy-rough nearest neighbor classifier combined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer. *Expert Systems with Applications*, *42*(20), 6844-6852.

[25] Aličković, E., & Subasi, A. (2017). Breast cancer diagnosis using GA feature selection and Rotation Forest. *Neural Computing and Applications*, *28*(4), 753-763.

[26] Sheikhpour, R., Sarram, M. A., & Sheikhpour, R. (2016). Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer. *Applied Soft Computing*, *40*, 113-131.

[27] Ed-daoudy, A., & Maalmi, K. (2020). Breast cancer classification with reduced feature set using association rules and support vector machine. *NetMAHIB*, *9*(1), 34.

[28] Wang, H., Zheng, B., Yoon, S. W., & Ko, H. S. (2018). A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, *267*(2), 687-699.

[29] Kumar, A., & Poonkodi, M. (2019). Comparative study of different machine learning models for breast cancer diagnosis. In *Innovations in Soft Computing and Information Technology* (pp. 17-25). Springer, Singapore.

[30] Preetha, R., & Jinny, S. V. (2020). Early diagnose breast cancer with PCA-LDA based FER and neuro-fuzzy classification system. *Journal of Ambient Intelligence and Humanized Computing*, 1-10.