Journal of Engineering and Applied Sciences
MEDWELL PUBLICATIONS

The Application of Particle Swarm Optimization Algorithm to Increase the Accuracy of MLP Neural Network for Prediction of Breast Cancer

[1]Mohammad Saeid Zahedi, [2]Zohreh Mortezania and [1]Mahmoud Makkiabadi
[1]Bioinformatics and Computational Biology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[2]Shahid Beheshti University of Medical Sciences, Tehran, Iran

Key words: Breast cancer, particle swarm optimization algorithm, MLP neural network, increasing the prediction accuracy

Corresponding Author:
Mohammad Saeid Zahedi
Bioinformatics and Computational Biology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

Page No.: 3732-3740
Volume: 15, Issue 23, 2020
ISSN: 1816-949x
Journal of Engineering and Applied Sciences

Abstract: Breast cancer usually begins from breast tissue and progresses rapidly. The disease is the most common cancer that women suffer from. With late diagnosis of breast cancer, the likelihood of the relapse of the disease is increased. The earlier breast cancer is diagnosed, the greater the likelihood of successful treatment would be. Also, if cancer is diagnosed in the early stages, the likelihood of the relapse of cancerous tumors is decreased. The presence of various symptoms and features of this disease makes it difficult for doctors to diagnose. The neural network provides the possibility of analyzing patient's clinical data for medical decision making. The purpose of this study is to provide a model for increasing the accuracy of prediction of breast cancer. In this study, patient's information has been collected from the standard database of Mortaz Super Specialty Hospital of Yazd. The medical records of 574 patients with breast cancer having a total of 32 features have been investigated. Each patient has been followed for at least one year. In order to provide a model of prediction of breast cancer, particle swarm optimization algorithm and MLP neural network are used. The proposed model was compared with the methods of the nearest neighbor, Naïve Bayes and decision tree. The results show that the prediction accuracy of the proposed model is equal to 0.966. Also, for the methods of Naïve Bayes, decision tree and the nearest neighbor, prediction accuracy is 0.91, 0.929 and 0.913, respectively. In predicting breast cancer, the proposed model includes minimum error rate and maximum accuracy and validity compared to other models. Naïve Bayes method has maximum error rate and minimum accuracy.

INTRODUCTION

Breast cancer is preventable and early diagnosable and by providing specific strategies, it is possible to reduce the late referral of the patient. Effective treatment of breast cancer is important in order to increase survival, reduce death and improve quality of life. This disease is a major threat to the health of women and is considered as one of the most common causes in reducing the lives of women[1-4]. Breast cancer is due to the excessive growth of abnormal cells in breast[5, 6]. To predict, diagnose and treat the disease, many factors are applied such as the presence of a mass, lymph nodes involvement, nipple niche, breast discharge, etc.[7, 8]. Large similarity in clinical and laboratory symptoms of breast cancer increases the risk of false diagnosis[9]. The mass is the most common symptom of breast cancer that often is detected accidentally by the patient herself and in other cases, it is diagnosed by a physician in physical checkup[1, 2]. This mass may be painful but in most cases, it is painless. In some cases, breast cancer occurs in multiple masses[10-13].

The similarity of clinical and laboratory symptoms of breast cancer increases the likelihood of an error in diagnosis. Diagnosis and prediction of various diseases are possible by the neural network. Discovering useful patterns between disease and clinical and laboratory symptoms of a patient is one of the uses of neural network applications in medicine. A useful model is a model in data that describes the relationship between a subset of patient's data and the disease diagnosis[14]. In recent years with the advances in early diagnosis of the disease, its treatment has been more successful. If breast masses are detected in small sizes, they can be well treated. Neural network is a new method for early diagnosis of breast cancer.

Sheikhpour *et al*.[15] introduced the diagnosis of breast cancer using two-step reduction of the extracted properties of needle aspiration and data analysis algorithms. The use of three data analysis methods for diagnosis of breast cancer was introduced by Delen *et al*.[16] that they were simulated the prediction of cancer by decision tree (C5) with accuracy of 93.6%, artificial neural network with accuracy of 91.2% and logistic regression model with accuracy of 89.2%. Aruna *et al*.[17] predicted breast cancer through the WDBC database by data analysis models such as decision tree and Naïve Bayes and neural network with accuracy of 92.97, 92.61 and 67/93, respectively. Ester, etc., provided a linear discrete analysis in diagnosis of breast cancer using a WDBC database with an accuracy of 96.8%.

The accuracy of neural network's diagnosis and prediction can be increased by setting the parameters of the network using the collective intelligence algorithms. Collective intelligence is actually formed on the basis of the simulation of the collective behavior of a group of animals like birds and fishes[18].

In this study, it was introduced a model for increasing the accuracy of MLP neural network's diagnosis and prediction based on the particle swarm optimization algorithm by designing a questionnaire, completing and collecting data sets related to patients' records in Mortaz super specialty hospital, Yazd.

MATERIALS AND METHODS

Data sets collection: To prepare the data collection related to breast cancer, first, a standard questionnaire was designed on the basis of new features and then completed and collected in Mortaz super specialty hospital, Yazd. The data are related to 2014-2015. The data set includes 574 samples of which 17 samples lack complete information. The samples lacking complete information were estimated by the maximum frequency method. For each patient, 32 features are registered. The initial values interval of patient's clinical features is shown in Table 1. The class names and the number of available samples in each class are shown in Table 2. The appropriate data format as the input of data analysis is effective in results and output. If the values of the dataset's features are in a different domain, the likelihood of an error in the findings will be increased. Normalization is said to put the data of a statistical population in the same domain[19]. In the proposed model, the normalization method is Max/Min and in the interval of [1, 0][20].

MLP neural network: The neural network in medicine is used by learning process to predict diseases. The method identifies the relationships and patterns between data through processors called neurons. The neural network approach provides a mapping between the input space (input layer) and the optimal space (output layer). The input layer receives the data and delivers it to the hidden layers. Information is processed in hidden layers and delivered to the output layer. Each network is trained by receiving some examples. Learning is done through the training process. Network learning is performed when the difference between the predicted values and the actual information is acceptable. The trained neural network is used for prediction with a new set of data[21]. The main features of the artificial neural network include the ability to learn patterns, high processing speed, the ability to generalize knowledge after learning, flexibility to deal with unwanted errors and not to cause significant disturbance if there is an error[22].

Particle swarm optimization algorithm: Particle swarm optimization algorithm or PSO uses social behavior of bird or fish groups during food search to guide the

Table 1: The interval of the values of patients' clinical features

The values interval in data set	Feature	The values interval in data set	Feature
[25 67]	age	0 = there is not any	The replacement of HRT hormone
[0 51]	Marriage period	1 = there is from 1-5 years	
[115 189]	Height	2 = there is from 6-10 years	
[49 110]	Weight	3 = there is from 11-15 years	
[16 37]	BMI	4 = there is >15 years	
[0 9]	The number of natural childbirth	0 = no niche	Nipple nicheN/R
[0 5]	The number of cesarean delivery	1= there is niche and it is old	
[0 3]	The number of aborsement	2 = there is niche and it is new	
[8 16]	Age of the first menstruation	0 = no pain	Breast pain
[16 37]	Age of the first natural childbirth	1 = little	
[18 41]	Age of the first cesarean delivery	2 = moderate	
[15 32]	Age of the first aborsement	3 = much	
0 = no history	The infertility history of patient	0 = no area	Pain area in breast
1 = there is history		1 = one area	
[43 56]	Age of menopause	2 = propagated	
0 = no history	Patient's family history of	0 = no discharge	Breast discharge
1 = third grade family	breast cancer	1 = there is discharge	
2 = uncle		(self discharge)	
3 = aunt		2 = there is discharge	
4 = brother		(press discharge)	
5 = sister		0 = no color	The color of breast discharge
6 = child		1 = bloody	
7 = father		2 = white and milky	
8 = mpther		3 = green and black	
0 = no change	The rate of change in the previous	4 = brown or pink	
1 = little change	mammography image	5 = colorless	
2 = moderate change	Rather than the current state of	0 = no mass	Mass
3 = a lot of change	the patient	1 = there is mass	
0 = no use	The type of contraceptive pill used	1 = mass in right breast	Position of the masses
1 = HD pill	by the patient	2 = mass in left breast	
2 = LD pill		3 = two-sided mass	
3 = other pills		[6 0]	Mass size
0 = <1 year	Duration of patient taking a	0 = no stiffness	Palpable stiffness in breast
1 = from 1-5 years	contraceptive pill	1 = there is stiffness	
2 = from 6-10 years		1 = there is stiffness	
3 = from 11-15 years		0 = no lactation	Lactation period
4 = >15 years		1 = one year	
0 = there is history	History of patient's hysterectomy	2 = two years	
1 = there is no history		3 = three years	
		4 = four years	
		5 = five years and more	
		0 = Menstruation is not	The cause of normal
		discontinued	menstruation cessation
		1 = normal menopause	
		2 = Premature menopause	
		3 = Removal of the womb	
		4 = Other causes	
		0 = Healthy	Cancer status
		1 = benign	
		2 = Malignant	

Table 2: Names and number of class samples

Samples number	Class name
183	Healthy
329	Benign
62	Malignant

population to the promising area in the search space[23, 24]. In PSO, each answer to a problem is the position of a bird in the search space that is called a particle. All particles include a value of merit obtained by the fitness function which is the purpose of optimization. Moreover, each particle also has a component called the velocity that determines its path in the search space. The PSO population contains all the particles called the Swarm. The PSO algorithm includes the two models of velocity and location equations. The coordinates of each particle represent a possible answer associated with two vectors. The vectors of position (Xi) and the velocity (Vi) are the two dependent vectors and related to each i particle in the N-dimensional search space which are respectively expressed as follows:

$$x_i^{k+1} = x_i^{k} + Cv_i^{k+1} \quad (1)$$

$$v_i = [v_{i1}, v_{i2}, ..., v_{iN}] \quad (2)$$

A community of birds consists of a number of particles (possible responses) flying in an appropriate response space to search for optimal solutions. The position of each particle is updated on the basis of the best of its search, the best overall experience of group flight and the vector of the previous particle velocity and based on the following relationships:

$$x_i^{k+1} = x_i^{\,k} + Cv_i^{k+1} \quad (3)$$

$$v_i^{\,k+1} = wv_i^k + c_1 r_1 (Pbest_i^k - x_i^k) + c_2 r_2 (Gbest^k - x_i^k) \quad (4)$$

In which c_1 and c_2 are two positive integer constants, r_1 and r_2 are two random numbers with uniform distribution in the range [0, 1] and w is the inertia weight which is chosen as follows:

$$w = w_{max} - \frac{w_{max} - w_{min}}{iter_{max}} \times iter \quad (5)$$

where, $iter_{max}$ is the maximum number of repeats and iterer is the number of current repeats. $Pbest^k_i$ is the best position of the particle i which is obtained on the basis of the particle's experience and can be expressed as follows:

$$Pbest_i^k = [x_{i1}^{pbest}, x_{i2}^{pbest}, ..., x_{iN}^{pbest}] \quad (6)$$

$Gbest^k$ is the best position of the particle based on the general group experience and consists of:

$$Gbest = [x_1^{gbest}, x_2^{gbest}, \ldots, x_N^{gbest}] \quad (7)$$

And k is the repeat index.

The proposed model: In this study, MLP (Multilayer Perception) network is used to predict breast cancer. This network contains an input layer, one or more hidden layers and an output layer. The available datasets in the files of patients with breast cancer are sent to the input layer for processing. To train the network, the Back Propagation algorithm (BP) is usually used. During the training of the MLP network using the BP learning algorithm, the computation is first performed from the network's input to the network's output and then the calculated error values are released to the previous layers. Initially, the output is calculated as layer by layer and the output of each layer is the input of the next layer. In back propagation mode, the output layers are firstly modulated, since there is a desirable amount for each neuron of the output layer and it is possible to modulate the weights by their help and updating rules.

All features and findings are of particular importance in diagnosis and prediction of breast cancer. In other words, all features do not have the same value. For example, in diagnosis of the disease, two features of the BMI and the presence of a tumor include different significance. It is important to know the value of each feature and how much it plays role in diagnosing the disease. In the study, the value and role of each feature are specified by the neural network and the disease is diagnosed. The training stages of the neural network of the proposed algorithm are:

- Step one: each matrix input value is assigned a random weight
- Step two: the appropriate input and output vector is selected
- Step three: the output of neuron in each layer and so, the output of neurons in the output layer is computed
- Step four: weights are updated to the previous layers by the method of network error propagation
- Step five: the function of the trained network is evaluated

After training, the neural network will be able to accurately predict the noise levels. The overall structure of the used neural network is shown in Fig. 1. Figure 2 shows the trend of the proposed approach.

The weakness in the results of the MLP neural network may be due to the long or uncertain learning time, the inability to properly select the learning coefficients and the inappropriate distribution of the primary weights. One of the factors of disruption in neural network's learning is inappropriate values of parameters. In this study, the PSO algorithm is used to set parameters for learning the network.

The target function for PSO algorithm is defined as: the PSO method consists of the two models of velocity and location equations. The coordinates of each particle represent a possible answer associated with the two vectors. Position (x_i) and velocity (v_i) vectors are two dependent vectors and are related to each particle i in the N-dimensional search space which are expressed as $x_i = [x_{i1}, x_{i2}, ..., x_{iN}]$ and $v_i = [v_{i1}, v_{i2}, ..., v_{iN}]$, respectively:

$$v_h^{\,k+1} = wv_h^k + c_1 r_1 (Pbest_h^k - h^k) + c_2 r_2 (Gbest^k - h^k) \quad (8)$$

$$v_m^{\,k+1} = wv_m^k + c_1 r_1 (Pbest_m^k - m^k) + c_2 r_2 (Gbest^k - m^k) \quad (9)$$

In which c_1 and c_2 are two positive integers r_1 and r_2 are two random numbers. $Pbest^k_i$ is the best position of the particle i which is expressed on the basis of the particle's experience and in accordance with the following equation:

$$Pbest_i^k = [x_{i1}^{pbest}, x_{i2}^{pbest}, \ldots, x_{iN}^{pbest}] \quad (10)$$

$Gbest^k$ is the best position of the particle on the basis of the general group experience and is defined as follows:

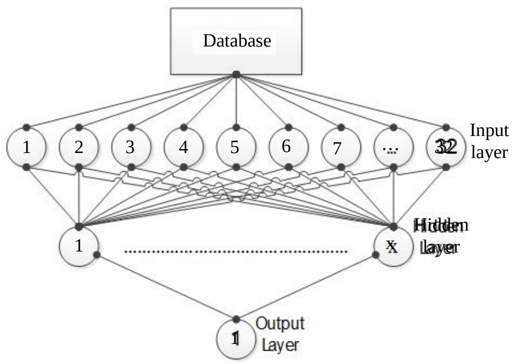

Fig. 1: The structure of artificial neural network

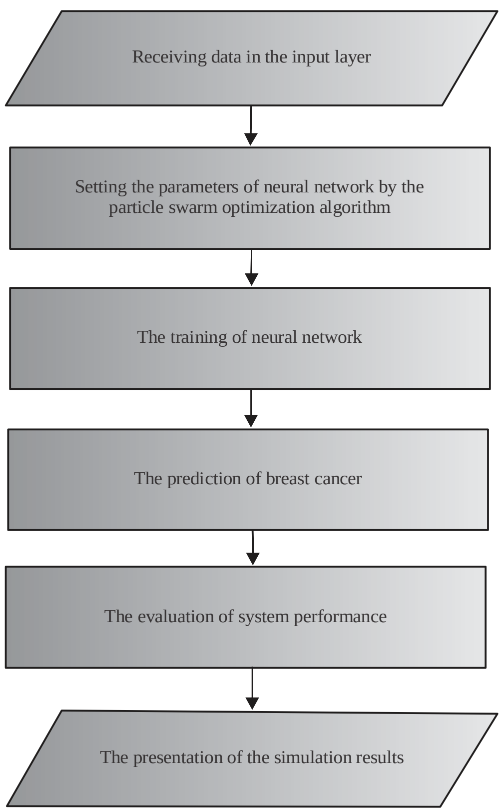

Fig. 2: The trend of the proposed approach

$$\text{Gbest} = [x_1^{gbest}, x_2^{gbest}, \ldots, x_N^{gbest}] \tag{11}$$

In the proposed algorithm, the number of hidden layers and the number of hidden neurons in each layer are considered as particles. Assuming:

m = The number of hidden layers
n = The number of hidden neurons in each layer

The values of n, m must be closer to the Gbest values. The Pbest in the proposed algorithm is obtained

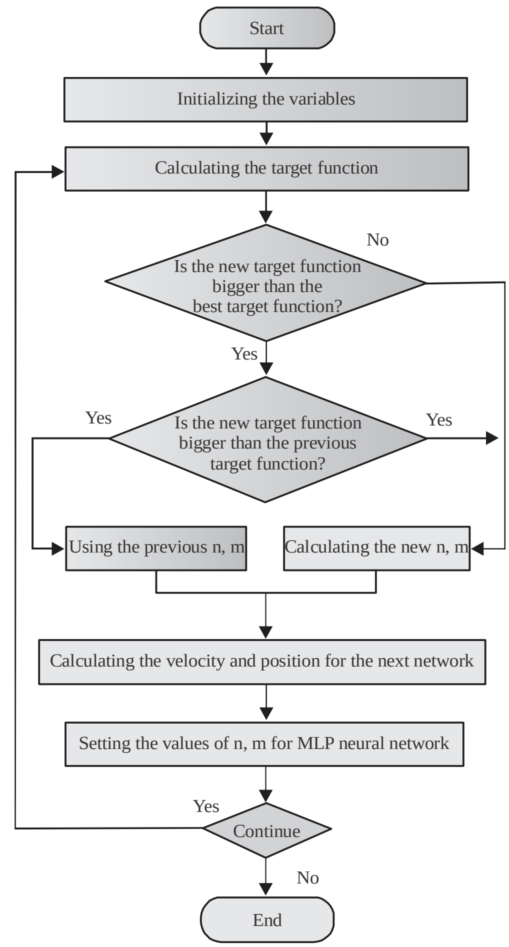

Fig. 3: The chart of the methodology of PSO algorithm for computing the optimal m and n

when the values of n, m make our desired criteria achieve their optimum levels. In this case, the updated n, m that makes our intended criteria are optimal is considered as n, m for Pbest. Also, the best global is obtained when the values of n, m are equal to the Gbest values. By placing Pbest and Gbest in Eq. 10, the particle velocity is calculated. Then, by placing the particle's velocity in Eq. 11, the particle position can be calculated. The following relationships are used to calculate the next position of each particle:

$$m^{k+1} = m^k + v_m^{k+1} \tag{12}$$

$$n^{k+1} = n^k + v_n^{k+1} \tag{13}$$

The obtained n_i+1 and m_i+1 from the relations of Eq. 12 and 13 are new particles and they are placed in the position of the available n, m in the preceding stage (Fig. 3). Thus, the values of n, m, defined in MLP neural network as static are converted into dynamic variables in PSO algorithm and changed depending on the network

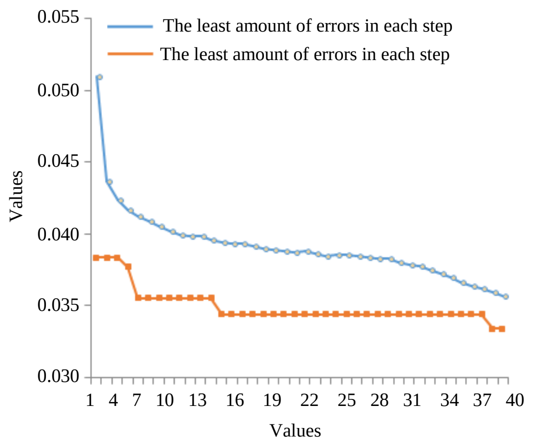

Fig. 4: Error percent for the prediction of breast cancer in the proposed model

conditions. The methodology and the chart of PSO algorithm's work flow for calculating optimal m and n are shown in Fig. 3 and Algorithm 1, respectively.

Algorithm 1; The methodology of PSO algorithm for computing the optimal m and n:

1- start
2- Initializing the variables
3- For each simulation period
3-1. Calculating the target function
3-2. Comparing with the best previous target functions and finding the larger target function (using particle memory)
3-3. Finding an optimal n, m which leads to the maximization of the target function
3-4. Calculating velocity through local and global bests
3-5. Calculating the next position for particles
4- Continuing the service process
5- End

At each stage of the implementation of PSO algorithm, the accuracy of predicting the MLP neural network is increased. In Fig. 4, the error percent has been shown in 40 stages for predicting the disease in the proposed algorithm. As seen, the accuracy of predicting breast cancer is increased in each stage.

The available data in the patient's files are described, simulated and analyzed by MATLAB software (R2013b, The Mathworks Inc., USA).

RESULTS AND DISCUSSION

Findings: The proposed model is compared with the three methods of Naïve Bayes[25], decision tree[26] and nearest neighbor. The relationship between actual classes and predicted classes can be calculated using the Confusion matrix. In the study, the required parameters of the Confusion matrix are indicated.

To compare the proposed model with other methods, the criteria of Accuracy, Sensitivity, Specificity, Precision and F-Measure are used according the following relationships:

$$\text{Accuracy} = (\text{TP+TN})/\text{All} \quad (14)$$

$$\text{Sensitivity} = \text{TP}/(\text{TP+FN}) \quad (15)$$

$$\text{Specificity} = \text{TN}/(\text{FP+TN}) \quad (16)$$

$$\text{Precision} = \text{TP}/(\text{TP+FP}) \quad (17)$$

$$\text{Recall} = \text{TP}/(\text{TP+FN}) \quad (18)$$

$$\text{F Measure} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \quad (19)$$

The required parameters for the relationship between the real classes and the predicted classes:

- TP = The number of records that are correctly recognized as being positive
- TN = The number of records that are correctly recognized as being negative
- FP = The number of records that are incorrectly recognized as being positive
- FN = The number of records that are incorrectly recognized as being negative

Figure 5 illustrates the chart of the results of different method's diagnosis with the criterion of accuracy. As seen, the proposed model has more accuracy than other methods. Also, the comparison of the results of predicting the disease with the criteria of sensitivity and specificity have been shown in Table 3 and 4, respectively. Table 5 and 6 have compared the results of the methods with the Precision and F-Measure criteria, respectively. The comparison results indicate the superiority of the proposed model's performance. In Table 7, the proposed method has been compared with other methods by the criteria of F-Measure, Precision, Sensitivity and Specificity. Table values indicate a better performance of the proposed method.

In this study, a collection of less famous and effective patterns in the incidence of breast cancer has been processed on the basis of a local database. Designing and completing questionnaires, recording clinical symptoms and laboratory results in the used datasets has been compiled by the authors of the paper in super specialty Mortaz hospital, Yazd. Research on breast cancer prediction has some limitations including the small number of patients for creating the model, missing data and incomplete variables. In this study, the acceptable

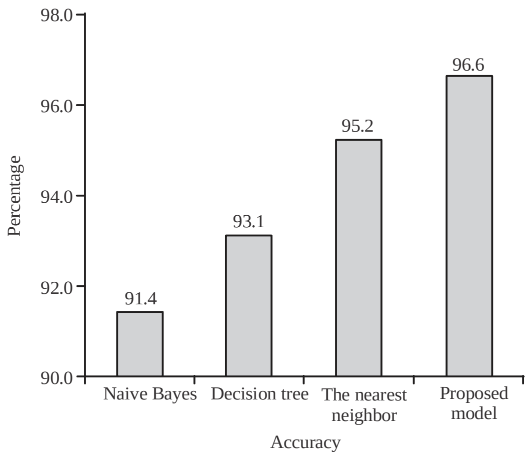

Fig. 5: The chart of the results with the criterion of accuracy

Table 3: The comparison of the results with the criterion of sensitivity

The proposed model	The nearest neighbor	Decision tree	Naïve Bayes	Class name
0.943	0.850	0.833	0.833	Healthy
0.975	0.981	0.923	0.925	Benign
0.977	0.981	0.931	0.831	Malignant

Table 4: The comparison of the results with the criterion of specificity

The proposed model	The nearest neighbor	Decision tree	Naïve Bayes	Class name
0.973	0.975	0.960	0.968	Healthy
0.975	0.980	0.978	0.980	Benign
0.976	0.981	0.978	0.975	Malignant

Table 5: The comparison of the results with the criterion of precision

The proposed model	The nearest neighbor	Decision tree	Naïve Bayes	Class name
0.959	0.945	0.898	0.910	Healthy
0.976	0.980	0.961	0.981	Benign
0.975	0.982	0.931	0.876	Malignant

Table 6: The comparison of the results with the criterion of F-measure

The proposed model	The nearest neighbor	Decision tree	Naïve Bayes	Class name
0.950	0.895	0.890	0.870	Healthy
0.976	0.981	0.942	0.951	Benign
0.978	0.981	0.931	0.853	Malignant

Table 7: The overall comparison of data analysis methods

Precision	Sensitivity	Specificity	F-measure	Class name
0.919	0.899	0.967	0.908	Naïve Bayes
0.925	0.925	0.970	0.925	Decision tree
0.949	0.952	0.975	0.949	The nearest neighbor
0.967	0.968	0.975	0.967	The proposed model

number of patients with appropriate variables and the least amount of missing data have been used. The present study predicts breast cancer using patient's clinical and laboratory features and through genetic algorithm and data analysis. The proposed model of the study includes the normalization of features, the selection of features, weighting to features and the introduction of a method for predicting breast cancer through weighted features.

Mass features, tumor size, family history, the amount of changes in the image of the previous mammography compared with the current image of the patient had the highest efficacy in the categorization of the proposed model. Also, the features of the type of delivery, breast pain and age of the first aborsement had the least effect on categorization. The variables of the type of the used edible oil, fabric material of the bra and use of armpit's spray had no effect on the diagnosis and prediction of malignancy of breast cancer. In this study, in addition to identifying the most important features, we achieved a better performance in terms of accuracy, sensitivity and specificity indicators through the selection of features based on the genetic algorithm.

Aruna *et al*.[17] predicted breast cancer using observer classifiers such as Naïve Bayes and Decision Tree on the datasets of WBCD database. Based on the criteria of Accuracy, Sensitivity, Specificity, Precision, Recall and F-Measure in the simulation results, the proposed model has better results than this study[17]. Land and Warhegen used the support vector machine on the breast cancer datasets for prediction[27]. The results of the experiments showed that the support vector machine has the accuracy of 96.7%. Kiyan and Yildirim[28] by using the RBF and MLP methods for predicting breast cancer, reached the accuracy of 96.18 and 95.74%, respectively. Charasia *et al*.[29] achieved the accuracy of 96.4% using the SVM method. Sarvestani *et al*.[30] predicted malignancy of breast cancer by comparing the mean of squared error in multi-layer, competitive and radius-based neural networks that the best accuracy was for radius-based neural network. Lavania *et al*.[31] obtained the accuracy of 94.84% through the data of WBCD database and decision tree with a two-stage categorization.

Ashkeli *et al*.[32] provided the prediction of relapse of breast cancer using three data analysis techniques. The performed studies show that the accuracy of the three data analysis algorithms, i.e. decision tree, ANN and SVM was 0.936, 0.494 and 0.947, respectively. Kiani and Atashi[33] in the developed model had a specificity and sensitivity of 65 and 96%, respectively. In the study reviewing the past, they were used 809 patients with breast cancer and 89 features of each patient. Atashi and Kiani[34] discovered 100 associative relationships with reliability coefficients above 90%. The 10 relationships of the whole were reported as being significant. Kavrasia, etc., used decision tree to predict breast cancer with the accuracy of 74%. The accuracy of 87% in the diagnosis of breast cancer using the C4.5 tree was the result of the work of Chaurasia and Pal[35].

Salaama, etc., used a fuzzy-neural network approach to predict breast cancer. The prediction accuracy of this method was 95.06%. While the proposed model with the accuracy of 96.6%, the sensitivity of 96.8% and specificity of 97.5% has a better performance in the diagnosis of breast cancer than the methods of Naïve Bayes, decision tree and the nearest neighbor.

In the proposed method, the purpose was to design and evaluate a medical assistant model for the determination of breast cancer malignancy using the PSO algorithm to increase the accuracy of the neural network. The designed medical assistant model in this research has been successful in the diagnosis of malignancy and has performed the categorization with acceptable accuracy. Experiments and simulations showed that the medical assistant system, introduced in this study, reached the accuracy of 96.6% for the data set of local patients with breast cancer in the Mortaz Hospital of Yazd which was higher than similar research on different datasets.

CONCLUSION

Search in medical databases for knowledge and information to predict, diagnose and decide is of data analysis applications in medicine. Hereditary algorithms such as genetic algorithm can be used to optimize data analysis techniques. The proper prediction and diagnosis of breast cancer increases the chances of successful treatment with the use of artificial intelligence and machine learning.

In the paper for prediction and diagnosis of breast cancer, the PSO algorithm was used to optimize neural network results and a new model was proposed. The simulation results show that the proposed model with the accuracy of 0/966 is more accurate than the methods of Naïve Bayes, decision tree and the nearest neighbor.

REFERENCES

01. Wang, J., F. Kato, H. Yamashita, M. Baba and Y. Cui *et al*., 2017. Automatic estimation of volumetric breast density using artificial neural network-based calibration of full-field digital mammography: Feasibility on Japanese women with and without breast cancer. J. Digital Imaging, 30: 215-227.
02. Wahab, A.A., M.I.M. Salim and M.N.C. Aziz, 2017. *In Vivo* Thermography-Based Image for Early Detection of Breast Cancer using Two-Tier Segmentation Algorithm and Artificial Neural Network. In: Application of Infrared to Biomedical Sciences, Ng, E. and M. Etehadtavakol (Eds.). Springer, Singapore, pp: 109-131.
03. Chen, D., G. Qian, C. Shi and Q. Pan, 2017. Breast cancer malignancy prediction using incremental combination of multiple recurrent neural networks. Proceedings of the International Conference on Neural Information Processing, November 14-18, 2017, Springer, Guangzhou, China, pp: 43-52.
04. Sun, W., T.L.B. Tseng, J. Zhang and W. Qian, 2017. Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. Computerized Med. Imaging Graphics, 57: 4-9.
05. Parkin, D.M., P. Pisani and J. Ferlay, 1999. Estimates of the worldwide incidence of 25 major cancers in 1990. Int. J. Cancer, 80: 827-841.
06. Ashley, D.J.B. and R.W. Evans, 1990. Evan's Histological Appearance of Tumours. 4th Edn., Churcill Livingstone, Edinburgh, Scotland, pp: 440-455.
07. Forrest, A.P., H.J. Stewart, D. Everington, R.J. Prescott and C.S. McArdle *et al*., 1996. Randomised controlled trial of conservation therapy for breast cancer: 6-year analysis of the Scottish trial. Lancet, 348: 708-713.
08. Sari, F., S.R. Mahdavi, R. Anbiaee and A. Shirazi, 2014. The effect of breast reconstruction prosthesis on photon dose distribution in breast cancer radiotherapy. Iranian J. Med. Phys., 14: 251-256.
09. Milovic, B. and M. Milan, 2012. Prediction and decision making in health care using data mining. Intl. J. Public Health Sci., 1: 69-78.
10. Adnan, M.H.M., N.A. Rashid and W. Husain, 2012. Data mining for medical systems: A review. Proceedings of the International Conference Advances in Computer and Information Technology-ACIT 2012, August 11, 2012, Hotel Shangri-La, Kuala Lumpur, Malaysia, pp: 17-22.
11. Richie, R.C. and J.O. Swanson, 2003. Breast cancer: A review of the literature. J. Insurance Med. New York Then Denver, 35: 85-101.
12. Mohammadpoor, M., A. Shoeibi and H. Shojaee, 2016. A hierarchical classification method for breast tumor detection. Iranian J. Med. Phys., 13: 261-268.
13. Sheikhpour, R., N. Ghassemi, P. Yaghmaei, J.M. Ardekani and M. Shiryazd, 2014. Immunohistochemical assessment of P53 protein and its correlation with clinicopathological characteristics in breast cancer patients. Indian J. Sci. Technol., 7: 472-479.
14. Cios, K.J. and G.W. Moore, 2002. Uniqueness of medical data mining. Artif. Intell. Med., 26: 1-24.
15. Sheikhpour, R., A.M. Sarram, Z.M.R. Mirakabad and R. Sheikhpour, 2015. Breast cancer detection using two-step reduction of features extracted from fine needle aspirate and data mining algorithms. Iranian Q. J. Breast Dis., 7: 43-51.
16. Delen, D., G. Walker and A. Kadam, 2005. Predicting breast cancer survivability: A comparison of three data mining methods. Artif. Intell. Med., 34: 113-127.

17. Aruna, S., S.P. Rajagopalan and L.V. Nandakishore, 2011. Knowledge based analysis of various statistical tools in detecting breast cancer. Comput. Sci. Inf. Technol., 2: 37-45.
18. Asuncion, F., 2010. UCI machine learning repository. University of California, California, USA.
19. Dodge, Y., D. Cox and Z. Geng, 2003. The Oxford Dictionary of Statistical Terms. Oxford University Press, Oxford, England, Pages: 498.
20. Han, J., M. Kamber and J. Pei, 2011. Data Mining: Concepts and Techniques. 3rd Edn., Morgan Kaufmann Publishers, USA., ISBN-13: 9780123814791, Pages: 744.
21. Hertz, J.A., A. Krogh and R.G. Palmer, 1991. Introduction to the Theory of Neural Computation. Addison-Wesley Publishing Company, Redwood City, California, USA., pp: 224.
22. Haykin, S., 1994. Neural Networks a Comprehensive Foundation. McMaster University, Hamilton, Ontario, Canada.
23. Zahiria, S.H. and S.A. Seyedin, 2007. Swarm intelligence based classifiers. Franklin Institute, Philadelphia, Pennsylvania.
24. Kennedy, J. and R.C. Eberhart, 1995. Particle Swarm Optimization Piscataway. IEEE Press, New Jersey, pp: 1942-1948.
25. Rish, I., 2001. An empirical study of the naive Bayes classifier. Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence Vol. 3, August 4-10, 2001, IBM, New York, USA., pp: 41-46.
26. Rokach, L. and O.Z. Maimon, 2008. Data Mining with Decision Trees: Theory and Applications. World Scientific Publishing, Singapore,.
27. Land Jr., W.H. and E.A. Verheggen, 2009. Multiclass primal support vector machines for breast density classification. Int. J. Comput. Biol. Drug Des., 2: 21-57.
28. Kiyan, T. and T. Yildirim, 2004. Breast cancer diagnosis using statistical neural networks. Istanbul Univ. J. Electr. Electron. Eng., 4: 1149-1153.
29. Chaurasia, S. and P. Chakrabarti, 2013. An approach with support vector machine using variable features selection on breast cancer prognosis. Int. J. Adv. Res. Artif. Intell., 2: 38-42.
30. Sarvestani, A.S., A.A. Safavi, N.M. Parandeh and M. Salehi, 2010. Predicting breast cancer survivability using data mining techniques. Proceedings of the 2010 2nd International Conference on Software Technology and Engineering Vol. 2, October 3-5, 2010, IEEE, USA., pp: V2-227-V2-231.
31. Lavanya, D. and K.U. Rani, 2012. Ensemble decision tree classifier for breast cancer data. Int. J. Inf. Technol. Convergence Serv., 2: 17-24.
32. Ashkeli, A.T., P.A. Ebrahamid, M. Ebrahimi and G.A. Leila, 2012. Prediction of relapse of breast cancer using three data analysis techniques. J. Breast Dis. Iran, 5: 23-34.
33. Kiani, B. and A. Atashi, 2014. Creating a data analysis-based prognostic model to predict the relapse of breast cancer. J. Inf. Health Biomed., 1: 26-31.
34. Atashi, A. and B. Kiani, 2015. Discovering the hidden patterns in the actual data sets of patients with breast cancer using data analysis techniques. J. Inf. Health Biomed., 8: 60-65.
35. Chaurasia, V. and S. Pal, 2014. Data mining techniques: to predict and resolve breast cancer survivability. Int. J. Comput. Sci. Mob. Comput. IJCSMC., 3: 10-22.