IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 13, Issue 3, Ver. II (May. - June. 2018), PP 01-12
www.iosrjournals.org

A Sub-Optimum Feature Selection Algorithm for Effective Breast Cancer Detection Based On Particle Swarm Optimization

Aya Hossam[1], Hany M. Harb[2], And Hala M. Abd El Kader[3]
[1,3]*(Electrical Engineering Department, Faculty of Engineering (Shoubra)/ Benha University, Egypt)*
[2]*(Computers and Systems Engineering Department, Faculty of Engineering/Azharuniversity,Egypt)*
Corresponding Author:Aya Hossam

Abstract:*Breast cancer (BC) disease is considered as a leading cause of death among women in the whole world. However, the early detection and accurate diagnosis of BC can ensure a long survival of the patients which brought new hope to them. Nowadays, data mining occupies a great place of research in the medical field. The Classification is an effective data mining task which are widely used in medical field to classify the medical dataset for diagnosis. Based on the BC dataset, if the training dataset contains non-effective features, classification analysis may produce less accurate results. To achieve better classification performance and increase the accuracy, feature selection (FS) algorithms are used to select only the effective features from the overall features. This paper proposed a sub-optimum FS algorithm based on the wrapper approach as evaluator and Particle Swarm Optimization (PSO) as a search method for the classification of BC dataset. The proposed PSO-FS algorithm uses a PSO algorithm to estimate and search for the significant and effective features subset from overall features set. Support Vector Machine (SVM), Artificial Neural Network (ANN), and Bayes Network (Bayes net) classifiers were used as evaluators to the optimized feature subset out from PSO search method. The Experimental results showed that the proposed PSO-FS algorithm is more effective by comparing with other two traditional FS search methods which are Beast First, and Greedy Stepwise in terms of classification accuracy and performance.*

Keywords: *Breast cancer, feature selection, Particle Swarm Optimization, Classifiers.*

Date of Submission: 24-05-2018 Date of acceptance: 08-06-2018

I. Introduction

BC is the most common cancer in women worldwide. It is the second leading cause of cancer deaths among females [1]. The early diagnosis of disease can lead to successful treatment and save life of the patients [2]. There are several imaging techniques for detecting BC such as MRI imaging, ultrasound imaging, Mammography, and Thermography. Breast thermography is a new imaging technique which is a relatively new screening method based on temperature a tumor may produce [3,4,5]. One of the important steps to diagnose the BC is classification of the thermal images' results into normal and abnormal cases. Early detection needs a precise and reliable breast diagnosis procedure that allows physicians to distinguish between normal breast thermal images and abnormal ones [6]. For this purpose, there are various computer-based solutions to serve as the breast diagnosis procedure and assist the physicians to specify the result of thermal images of patients. These systems are called Medical Diagnostic Decision Support (MDDS) systems and it can increase the natural capabilities of human diagnosticians for complex cases of medical diagnosis [7].

One of the challenges that faces these systems is the great number of features. Some of these features may be irrelevant to the mining task. Therefore, these features effect on management of dataset and cause of decreasing the accuracy of the classification algorithm [8,9]. FS method is used to cope with this problem. It is used to select a features subset from the original overall features present in a given BC dataset that provides most of the useful information [9,10]. This process of data reduction helps in reducing the number of features, and removes irrelevant, or noisy data. This reduction appears great effects on speeding up data mining algorithm, and improving classification performance such as predictive accuracy and result comprehensibility [11].

FS methods can be broadly divided into two categories: filter and wrapper approaches [12]. In filter approach, the search process is independent of a classifier algorithm, and it generally uses some techniques to record the selected subset. On the other hand, the best feature subset of the wrapper approach is evaluated by using a machine learning algorithm that is the classification engine. The filter approach has a disadvantage. In this approach, the process of selecting the best subset of features is independent to the type of classification algorithm. Due to this drawback, filter approach may cause a bad effect on the result of classifier algorithms because the subset is just selected based on correlation between data records. However, the wrapper approach doesn't have this mentioned drawback because the best feature subset is selected by techniques based on the type of classification. By considering the performance of the selected feature subset on a particular learning algorithm, the wrapper approach can usually have better results than the filter approach [13,14].

The first step of wrapper based FS methods is searching for the best subset of features among the wide variety of possible subsets of features. Recently, evolutionary computation (EC) techniques are well-known for their global search ability. Particle swarm optimization (PSO) is a relatively recent EC technique, which is computationally less expensive than some other EC algorithms. Therefore, PSO has been used as an effective technique in FS. It can be employed to perform the search step in wrapper based FS method . It has some advantages such as its simple mathematical operations, a small number of control parameters, quick convergence and ease of implementation [15,16].

The objective of this paper is to propose a sub-optimum FS algorithm for effective BC detection based on PSO named PSO-FS algorithm. The breast regions of interests (ROIs) were automatically extracted from BC thermal images by using the automatic segmentation method proposed in [17]. There are two types of features (first order statistical and texture) which extracted from the enhanced ROI of breast thermal images. So, the BC dataset which used in this paper contains about twenty nine features and it was prepared in [17]. To evaluate the optimized feature subset out from PSO search method, these features were provided as input to three classifiers which are Support Vector Machine (SVM), Artificial Neural Network (ANN), and Bayes Network (Bayes net). Our work was compared with other approaches that used the traditional FS search methods such as Best First, and Greedy Stepwise. The results showed more enhanced classification performance by providing the features selected by our proposed approach.

II. The Proposed BC Detection Model

This section discusses the materials and methods used in this work to build the proposed model for BC detection. It also introduces a background about PSO-based wrapper FS technique and the types of classifiers that used to validate our proposed algorithm. Figure 1 shows the flow chart of the proposed BC detection model. This model is based on the proposed PSO-FS algorithm. The aim of this algorithm is to achieve an optimum feature subset with minimum number of features providing efficient classification accuracy.

2.1 Dataset Description

In this paper, the resultant BC dataset that collected in [18] was used. In [18], the breast thermograms were collected from an open online data base PROENG (http://visual.ic.uff.br/) which called DMR-IR database [19]. About 200 (90 normal and 110 abnormal) cases with their thermograms were considered for this paper. After preprocessing and segmentation processes were applied on the thermograms, the required statistical and texture features were extracted to form our dataset. The authors in [18] proposed a new automatic segmentation method for breast thermograms and this method appeared its ability for best extracting the breast ROI image from breast thermograms. It lead to extract a required dataset with high accuracy. This BC dataset contains about twenty nine attributes. Each instance has one of two possible classes: normal or abnormal. The total number of instances are about 400 due to dividing each thermogram into left and right breast images. The description of these features and the resultant dataset was introduced in [18].

2.2 Feature Selection Using PSO

2.2.1 The Feature Selection Process

Extracting effective and useful data from a large collections of data has now a special concern within the data mining community. Researchers realize that the FS is an integral component in the implementation of successful data mining task. It is considered as an active research area for decades in domains as machine learning and data mining. FS is a process that decreases the number of attributes in dataset by selecting the effective subset of original features based on a certain criteria.Generally, FS is a multi-objective step. It aims to investigate two main objectives, which are the increasing of classification performance and the decreasing of the number of features.

The FS procedure consists of four basic steps which are subset generation, subset evaluation, stopping criterion, and result validation. It starts with subset generation which is a search procedure that uses a certain search strategy to produce candidate feature subsets for evaluation. Then each candidate subset is estimated due to an evaluation criterion and compared with the previous best result. If the new evaluated subset is better, the previous one is replaced with it, else remains the same. The subset generation and evaluation process still repeated until it satisfies the given stopping criteria. Finally, the selected best feature subset must be validated using the prior knowledge or various data tests. Search strategy and evaluation criteria are two necessary key points in the study of FS. Figure 2 shows the flowchart of FS process.

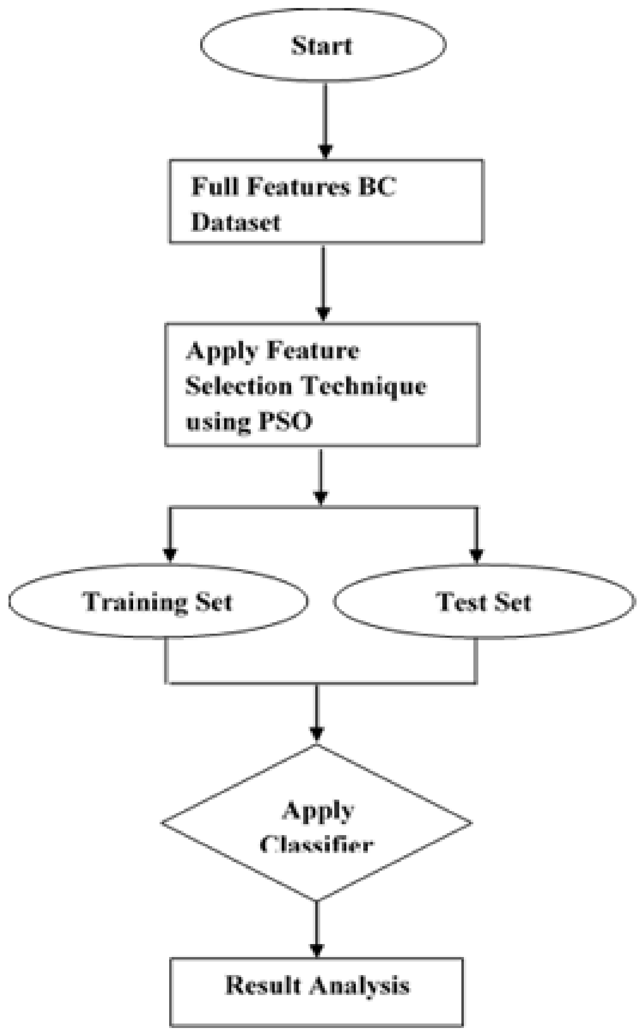

Fig.1: The flowchart of the proposed BC model based on a sub-optimum PSO-FS algorithm.

The evaluation criteria of FS algorithms has two categories which are [12]: the filter, and wrapper.

- The filter approach evaluates and selects feature subsets based on general characteristics of the dataset without depending on any mining algorithm. It is suitable for dataset with high dimensionality. To select the best subset of features, these methods use ranking and space search methods on the basis of strategy that they are following. In ranker based methods, every feature independently ranked by the uses of descriptive score functions and sorted in decreasing order on the basis of significance score. Although ranker based method is much more efficient in computationally, but poor to examine redundant features.

- The wrapper approach requires a predetermined mining algorithm to evaluate the best feature subset. It uses the predictive model for scoring the feature subset. The performance of this mining algorithm is used for the evaluation criterion. It reaches to more identification rates rather than the filters. So, the wrapper approach provides better results than the filter approach.

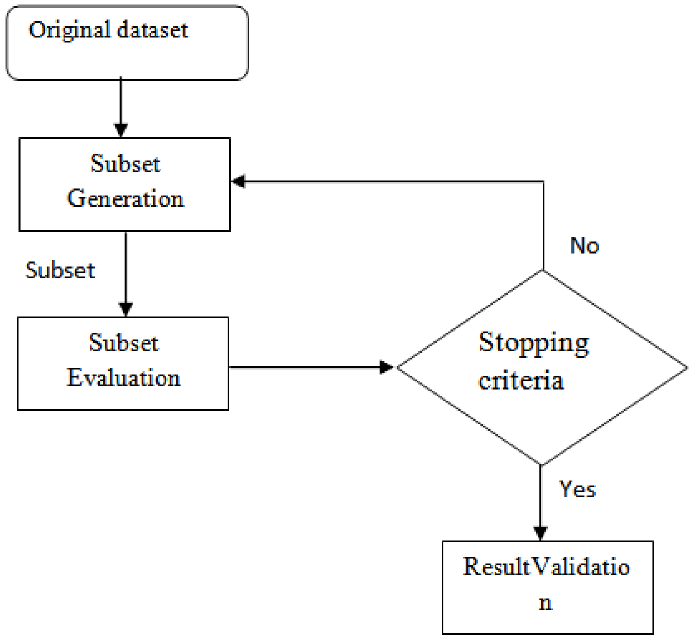

Fig.2: The flowchart of feature selection process.

2.2.2 Particle Swarm Optimization (PSO) based Wrapper Approach

The PSO is an evolutionary computation technique developed by Kennedy and Eberhart [20]. It is based on the behavior of swarm of bees or flock of birds while searching for food. The PSO algorithm maintains a population, named a swarm, of particles, where each particle represents a potential solution to the optimization problem. Each particle has a position in the search space that is represented by a vector $x_i = (x_{i1}.x_{i2}.\cdots\cdots.x_{iD})$, where D is the search space dimensionality. Particles move in the search space searching for the optimal solutions. So that, each particle has a velocity, which is represented by $v_i = (v_{i1}.v_{i2}.\cdots\cdots.v_{iD})$. The velocity value must be in a range defined by parameters Vmin and Vmax.

During movement, each particle updates its position and velocity due to its own experience and that of its neighbors. The best previous position of the particle represents the personal best position of the particle, called local best (lbest), and the best position obtained by the population thus far is called globalbest (gbest). According to lbest and gbest values, PSO searches for the optimal solutions by updating the velocity and the position of each particle using the following equations:

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1)(1)$$

$$v_{id}(t+1) = w * v_{id}(t) + C_1 * r_{1i} * (P_{id} - x_{id}(t)) + C_2 * r_{2i} * (P_{gd} - x_{id}(t))(2)$$

where t represents the t^{th} iteration, d ϵ D refers to the d^{th} dimension in the search space, $\boldsymbol{w}$ is the inertia weight, $\boldsymbol{C_1}$ and $\boldsymbol{C_2}$ are the learning factors which called cognitive parameter, and social parameter, respectively, $\boldsymbol{r_{1i}}$ and $\boldsymbol{r_{2i}}$ are random values uniformly distributed in [0, 1], and finally $\boldsymbol{P_{id}}$ and $\boldsymbol{P_{gd}}$ represent the elements of lbest and gbest in the d^{th} dimension. The personal best position of particle i can be determined by the following equation:

$$y_i(t+1) = \begin{cases} y_i(t) iff(x_i(t+1)) \geq f(y_i(t)) \\ x_i(t+1) iff(x_i(t+1)) < f(y_i(t)) \end{cases} (3)$$

The inertia weight can be calculated be the following equation (4):

$$w = \frac{w_{max} - w_{max}}{tmax} * t_{current} \ (4)$$

where $tmax$: the maximum number of iterations , $t_{current}$: the number of current iteration. The wrong value selection of these parameters will effect on the speed of PSO algorithm convergence. So, the initial selection would be very important [21].The steps of the implemented proposed PSO-FS algorithm is summarized as shown in flowchart in figure 3.

2.2.3 **Training data and Testing data**

To estimate the proposed model, the overall BC dataset should be divided into two parts: training set and testing set. The training dataset is used to build the machine learning model. While the testing dataset is necessary to measure the performance of this model. These two datasets must be different and created by random sampling. In this paper, one dataset for BC in one file have been created. So, it was important to split this dataset into two subsets for training and testing sets. Due to that, about 70% of the original dataset was considered as training dataset and the rest 30% of original dataset was used as testing dataset. After building the model using this training set, the BC model was tested using the testing set. Figure 4 shows the process of splitting the original dataset into training and testing sets.

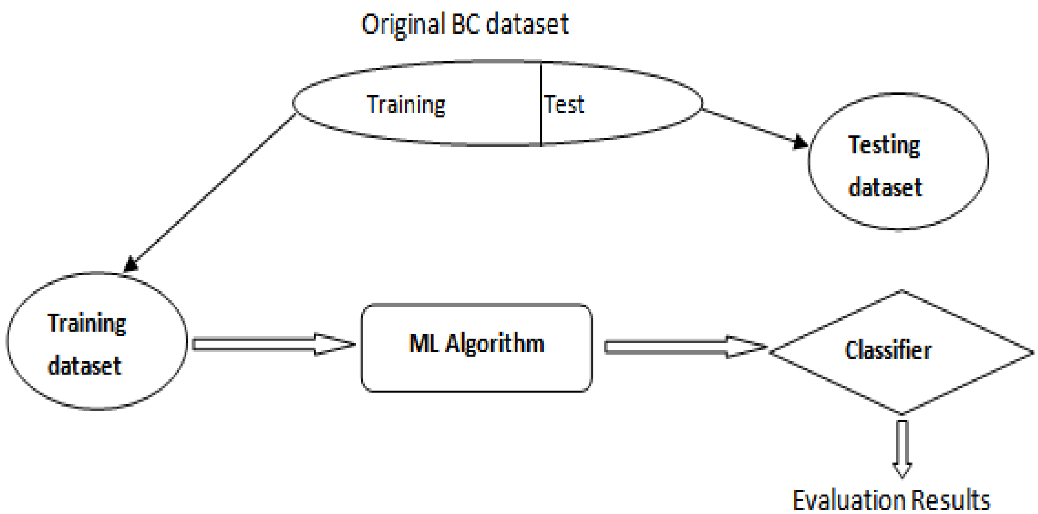

Fig.4: The splitting of original BC dataset into training and testing datasets.

2.2.4 Classifiers

Classifiers are the final step in building the BC detection model. They are used to classify network traffic dataset whenever apply over the dataset. In this paper, the classifiers help to differentiate the dataset into normal and abnormal cases. In WEKA Environment tool, there are about 76 classification algorithms, which are capable to perform the required task. This paper uses three classifiers combined with features selection method to build a classification model of the breast cancer diagnosis to determine whether the case is normal or abnormal. These classifiers are Support Vector Machine (SVM), Artificial Neural Network (ANN), Bayes Network (ByesNet).

2.2.5 Support Vector Machine (SVM)

SVM is a powerful classification algorithm which uses a hyperplane to classify the data set into different classes based on the class membership. The hyperplane can be described as wide line or two parallel lines with maximum distance where there is no data point between them. If there were multiple hyperplanes the best one is the one who is as far as possible from the data point. The goal of it is to point the new data into the right class. This is called the linear classifier [22]. To demonstrate it, suppose there were dataset consists of group of squares and circles.

DOI: 10.9790/2834-1303020112

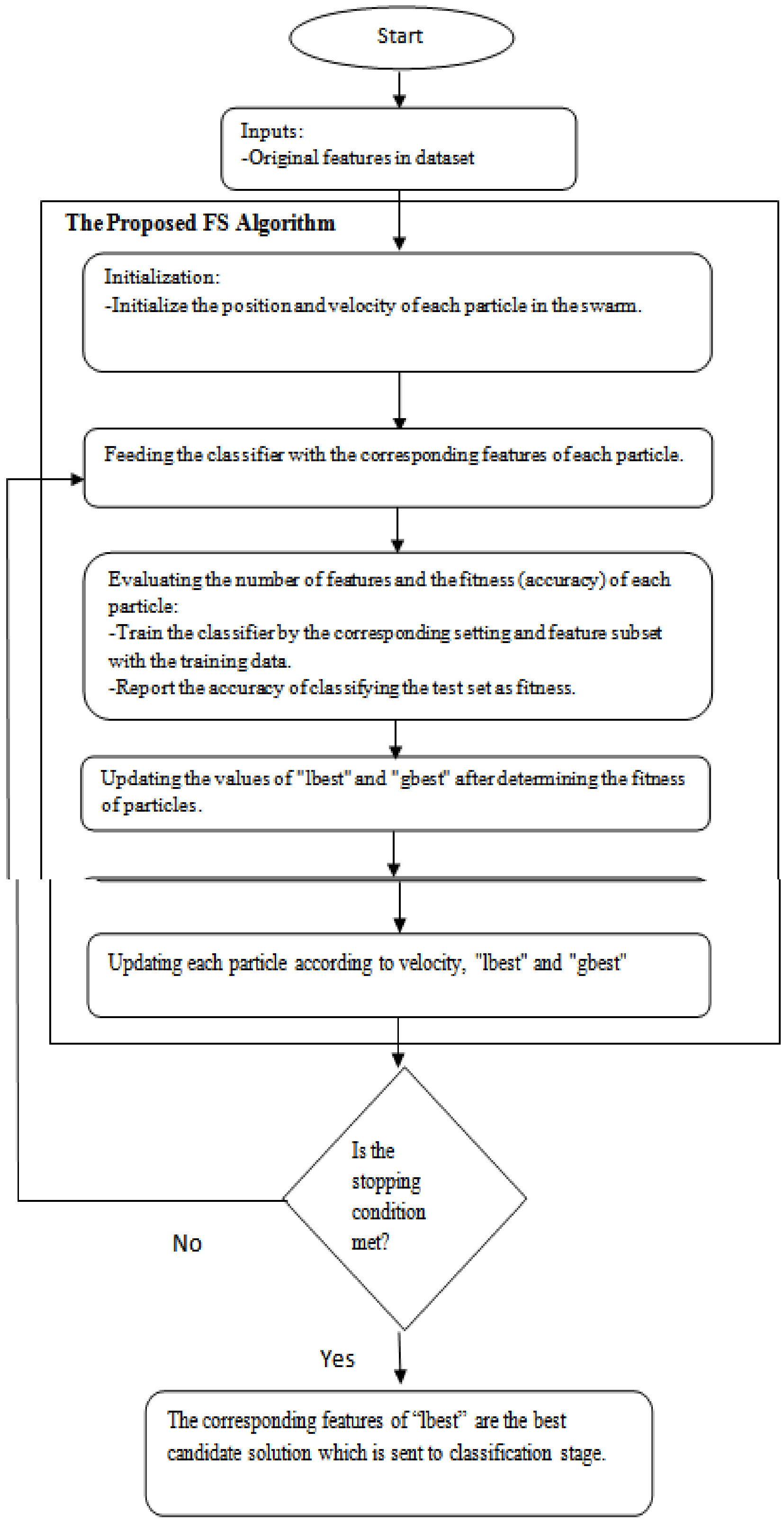

Fig. 3: Flowchart of the proposed PSO-FS algorithm

The hyperplane is going to separate the sets into two classes as shown in figure 5. In some cases, there are some data sets that don't permit linear classifier. Here, they come up with the non-linear classifier, which maps the data into a higher dimensional space and then uses the simple linear classifier [23].

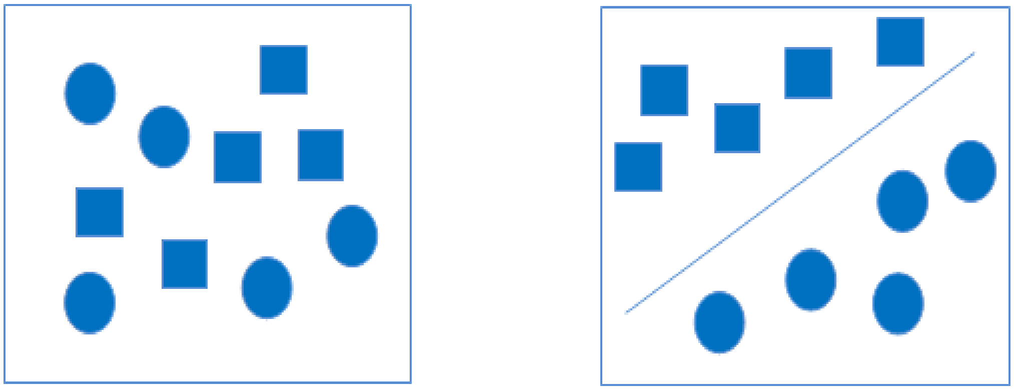

Fig.5: The process of classifying data sets into two classes using the hyper-plane.

2.4.2 Artificial Neural Network (ANN):

Artificial Neural Network (ANN) is an intelligent system inspired by the human brain. These networks are densely interconnected networks of Processing Elements together with rule to adjust the strength of the connections between the units in response to externally supplied data. The medical applications of ANNs mostly depend on their ability to handle classification problems including classifications of diseases or to estimate prognosis [24]. ANNs are highly adaptive structures with numerous adaptable coefficients which have to be set on specific values.

NNs consists generally of at least two physical elements which are: Neurons and a weighted link. Neurons are the process element and the weighted link is used to connect these elements together. There are three types of neurons identified by: input neurons, hidden neurons and output neurons. For input neurons, they receive data from element outside the network, while the output neurons use their produced output externally. However, the hidden neurons are located between input and output neurons in order to receive the input from input neurons in the network and its produced output is used as an input for output neurons.

Neurons can form multiple layers to create what is called the Multilayer Perceptron (MLP) which is the most popular type of ANNs. In MLP, the neurons are collected into layers. The first layer is the input layer which contains input neurons and the last layer is the output layer which contains output neurons. They represent the overall input and output of the network. There is one or more hidden layers which contains hidden neurons between these two layers. Each node in the input layer has directed connections to nodes in the hidden layer. Also, Each node in the hidden layer has directed connections to nodes in the output layer. Figure 6 shows the construction of MLP Neural Network

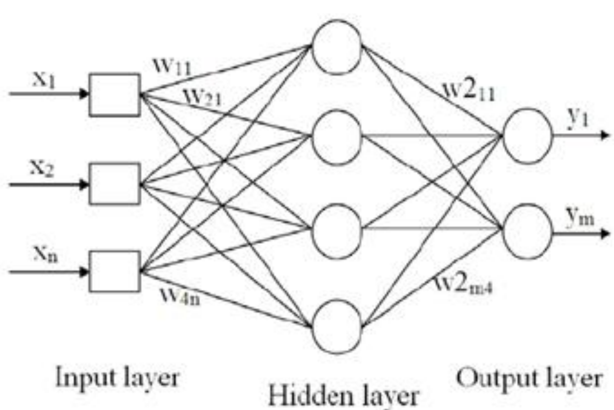

Fig.6: the construction of MLP Neural Network.

2.4.3 Bayesian Network:

A Bayesian network [25], also called belief networks or Bayes nets, is considered as a probabilistic graphical model which can be used to build models from data and/or expert opinion. This classifier is a type of a directed acyclic graph (DAG) which is popular in the machine learning, the statistics, and the artificial intelligence fields. The DAG is realized by using two sets which are: the set of nodes and the set of directed edges. For nodes, they are drawn as circles identified by the variable names and represent random variables. These nodes may be latent variables, observable quantities, unknown parameters or hypotheses. For edges, they represent direct dependence among the variables and drawn by arrows between nodes. Each node is associated with a probability function. The task of this probability function is to take a particular set of values for the node's parent variables as input and gives the probability of the variable represented by the node. For example, the variable BC has two sets which are: "Normal" and "Abnormal. Each one of these states has a probability value, for each node, these probability values sum to 1.

Bayesian networks introduces efficient algorithms that perform inference and learning. The dynamic Bayesian network is a type of Bayesian networks which model sequences of variables (e.g. protein sequences or speech signals). Generalizations of Bayesian networks that can represent and solve decision problems under uncertainty are called influence diagrams.

III. Estimation of Model Performance

To analyze and evaluate the results, data mining tool kit WEKA [25] is used. This software helps to create the BC detection models. The classification models can be evaluated using different performance measures such as Classification Accuracy, Root mean squared error (RMSE), Kappa statistic, True Positive Rate (TP-Rate), False Positive Rate (FP-Rate), Precision, Recall, and F-Measure Index. These several standard terms have been defined for the two class confusion matrix (Normal and Abnormal):

A. Accuracy:

The Accuracy metric is necessary to estimate the overall correctness of the BC model and it can be determined by dividing the sum of right classifications over the total number of classifications, as shown in the following equation:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

B. Root mean squared error (RMSE):

The RMSE measure is a frequently-used measure of the differences between values predicted by a model or an estimator and the values actually observed from the thing being modeled or estimated.

C. Kappa statistic:

The Kappa statistic measure is a very useful measure that can treat very well with both multi-class and imbalanced class problems. It can be defined as in the following equation:

$$K = \frac{(P(A)-P(E))}{(1-P(E))} \quad (6)$$

where, P (A) is the observed agreement, and P(E) is the expected agreement. It basically helps to know how much better the classifier is performing.

D. False Positive Rate (FP-Rate):

The FP-Rate defines the rate of negative cases that were classified as positive incorrectly. It is calculated by equation(7) :

$$FP-Rate = \frac{FP}{FP+TN} \quad (7)$$

E. Precision:

The Precision is another metric that represents the rate of the positive cases that were predicted correctly, and is determines using the following equation:

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

F. Recall or True Positive Rate (TP-Rate):

The Recall or TP-Rate introduces the ratio of the correctly identified positive cases. It is calculated as follows:

$$Recall = TP - Rate = \frac{TP}{TP+FN} \quad (9)$$

G. F-Measure:

In some cases, it is very important to have higher precision, but in other cases higher recall may be very important. However, in most cases, we try to improve both values. F-Measure is said to be the combination of these values, and in the most common form, it is the harmonic mean of the both:

$$F - measure = \frac{2*Precision \ *Recall}{Precision \ + Recall} \quad (10)$$

IV. Performance Evaluation and Results

In this section, the performance of each classifier based on the proposed sub-optimum PSO-FS algorithm has been analyzed. This was done using the features in the BC data set that prepared in [18]. Then, the results of were compared with Best First search, and greedy step wise search. To benchmark the proposed PSO-FS algorithm, different implementations and comparisons were performed using other FS search algorithms such as Best First, and greedy step wise. The purpose of this paper is to experiment that the use of the proposed PSO-FS algorithm increases the classifier performance and improve its accuracy. For this purpose, firstly each classifier was trained with whole "original features" and then the results were compared with the same classifier which has been trained by the subset of features obtained by the proposed "PSO-FS algorithm". Then, the results due to the proposed algorithm were also compared with the other FS algorithms which are BEST first and greedy stepwise.

For PSO, the individual weight is 0.34, inertia weight is 0.33, and population size is 20. In this paper, The PSO search method is applied on the three classifiers which are: SVM, MLP and Bayes net. Table no 1 shows the performance of the three classifiers without using any FS method as where the whole 29 attributes were used. Table no 2 shows the important effect of using the proposed sub-optimum PSO-FS algorithm on the performance of each classifier. It also shows the total number of attributes that were used after filtration process due to PSO search method. As shown in Table no 2 the performance and accuracy of the tree classifiers is better than using the whole attributes in Table1. Table no 3 and table no 4 show the effect of using other widely used FS search methods, which are Best First and Greedy Step wise, respectively, on the performance of classifiers and on the reduction of the total number of features. They achieved slightly better performance than using whole attributes in the dataset.

Table no 1 Performance of all Classification models using whole 29 attributes.

Classifier	Accuracy	RMSE	Kappa statistics	TP-Rate	FP-Rate	Precision	Recall	F-Measure	Class
SVM	96.21%	0.1946	0.9243	0.928	0.000	1.000	0.928	0.962	Normal
				1.000	0.072	0.926	1.000	0.962	Abnormal
MLP	95.83 %	0.2068	0.9168	0.921	0.000	1.000	0.921	0.959	Normal
				1.000	0.079	0.919	1.000	0.958	Abnormal
BayesNet	93.94%	0.2327	0.8782	0.971	0.095	0.918	0.971	0.944	Normal
				0.905	0.029	0.966	0.905	0.934	Abnormal

Table no 2 Performance of all Classification models using proposed PSO-FS algorithm.

Classifier	No. of features	Accuracy	RMSE	Kappa statistics	TP-Rate	FP-Rate	Precision	Recall	F-Measure	Class
SVM	11	98.48%	0.1231	0.9697	0.971	0.000	1.000	0.971	0.985	Normal
					1.000	0.029	0.969	1.000	0.984	Abnormal
MLP	10	97.76%	0.158	0.9535	0.967	0.006	0.996	0.967	0.981	Normal
					0.994	0.033	0.952	0.994	0.972	Abnormal
BayesNet	8	96.97%	0.1961	0.9391	1.000	0.063	0.945	1.000	0.972	Normal
					0.937	0.000	1.000	0.937	0.967	Abnormal

Table no 3 Performance of all Classification models using Best First search method.

Classifier	No. of features	Accuracy	RMSE	Kappa statistics	TP-Rate	FP-Rate	Precision	Recall	F-Measure	Class
SVM	13	96.21%	0.1946	0.9243	0.928	0.000	1.000	0.928	0.962	Normal
					1.000	0.072	0.926	1.000	0.962	Abnormal
MLP	7	95.00%	0.2356	0.9002	0.905	0.000	1.000	0.905	0.950	Normal
					1.000	0.095	0.905	1.000	0.950	Abnormal
BayesNet	5	95.45%	0.2381	0.9088	0.971	0.063	0.944	0.971	0.957	Normal
					0.937	0.029	0.967	0.937	0.952	Abnormal

Table no 4 Performance of all Classification models using Greedy Stepwise search method.

Classifier	No. of features	Accuracy	RMSE	Kappa statistics	TP-Rate	FP-Rate	Precision	Recall	F-Measure	Class
SVM	9	97.73%	0.1508	0.9545	0.957	0.000	1.000	0.957	0.978	Normal
					1.000	0.043	0.955	1.000	0.977	Abnormal
MLP	8	96.21%	0.1977	0.9243	0.928	0.000	1.000	0.928	0.962	Normal
					1.000	0.072	0.926	1.000	0.962	Abnormal
BayesNet	5	95.45%	0.2381	0.9088	0.971	0.063	0.944	0.971	0.957	Normal
					0.937	0.029	0.967	0.937	0.952	Abnormal

Based on the results, average accuracy of SVM, MLP, and Bayes network without FS are 96.21%, 95.83 %, and 93.94%, respectively. While the average accuracy of SVM, MLP, and Bayes network using the proposed PSO-FS algorithm are 98.48%, 97.76%, and 96.97%%, respectively. It is very obvious that using the proposed PSO-FS approach can improve the results in compare with using the original features in dataset. Also, the results in table 3 and table 4 demonstrate that the proposed algorithm enhanced the classification accuracy of almost all the data mining algorithms than the other FS search methods such as Best first and greedy stepwise. The graphical representation of the performance of the these three classification algorithms are represented in Figure 7.

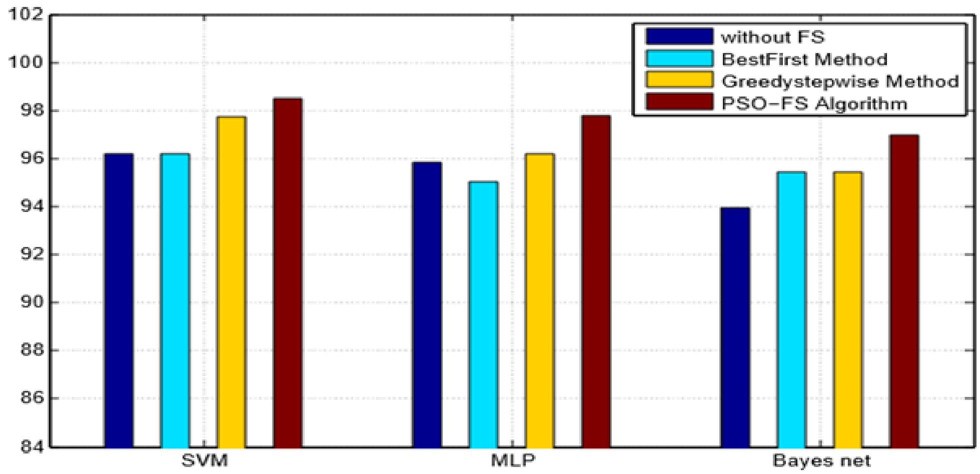

Fig. 7: Classifiers Performance before and after FS algorithms.

V. Conclusion

This paper proposes a sub-optimum FS algorithm for BC detection model which based on a PSO technique. The PSO in the proposed PSO-FS algorithm was used to search for the optimal set of attributes that can help to achieve better classification performance than using the overall attributes. After finding the significant features in the training set, three classifiers, which are SVM, MLP, and Bayes net, were used to classify the test dataset using the significant features only. The proposed PSO-FS algorithm was compared with other two traditional search algorithms in FS process named Best First and Greedy stepwise algorithms. The experimental results showed that the proposed PSO-FS algorithm achieved better classification accuracy and performance than without applying any FS algorithm. It also achieved better results than applying the other two widely algorithms used in FS which are Best First and Greedy stepwise algorithms. The proposed classification approaches due to SVM, MLP, and Bayes net classifiers using the proposed PSO-FS algorithm achieved accuracies reached to 98.48%, 97.76%, and 96.97%, respectively on the test dataset.

References

[1]. L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, " Global cancer statistics, 2012," CA: a cancer journal for clinicians, Vol. 65, No. 2, 2015, pp. 87-108.
[2]. I. Harirchi, et al., “Breast cancer in Iran: a review of 903 case records,” Public Health, Vol. 114, No. 2, 2002, pp. 143-145.
[3]. E.Y.K. Ng, "A review of thermography as promising non-invasive detection modality for breast tumor", International Journal of Thermal Sciences, Vol. 48, no. 5, 2009, pp. 849-859.
[4]. S.V. Francis, M. Sasikala, and S. Saranya, "Detection of breast abnormality from thermograms using curvelet transform based feature extraction", Journal of Medical Systems, Vol. 38, no. 4, 2014, pp. 1-9.
[5]. J. F. Head, F. Wang, C. A. Lipari, and R. L. Elliott, "The important role of infrared imaging in breast cancer", IEEE Engineering in Medicine and Biology Magazine, Vol. 19, no. 3, 2000, pp. 52–57.
[6]. T. Subashini, V. Ramalingam, and S. Palanivel, “Breast mass classification based on cytological patterns using RBFNN and SVM,” Expert Systems with Applications, Vol. 36, No. 3,2009, pp. 5284-5290.
[7]. R.A. Miller, “Medical diagnostic decision support systems - past, present, and future,” Journal of the American Medical Informatics Association, Vol. 1, No. 1, 1994, pp. 8-27.
[8]. Y. Yuling, “A Feature Selection Method for Online Hybrid Data Based on Fuzzy-rough Techniques,” IEEE computer society, 2009, pp. 320- 324.
[9]. W. Adam, N. Phong, and K. Alexandros, "Model mining for robust feature selection", KDD '12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM New York, NY, USA, 2012, pp. 913-921.
[10]. M. Dash, and H. Liu, “Feature selection for classification,” Intelligent data analysis, Vol. 1, No. 3, 1997, pp. 131-156.
[11]. MIT Lincoln Laboratory: http://www.ll.mit.edu/IST/ idaval/.
[12]. C. Girish, and S. Ferat, " A survey on feature selection methods," Computers and Electrical Engineering, Vol. 40, 2014, pp. 16–28.
[13]. B. Xue, M. Zhang, and W. N. Browne, “New fitness functions in binary particle swarm optimization for feature selection,” IEEE World Congress on Computational Intelligence (WCCI 2012), Brisbane, Australia, 2012.
[14]. R. Kohavi, and G. H. John, “Wrappers for feature subset selection,” Artificial Intelligence, Vol. 97, 1997, pp. 273-324.
[15]. B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms," Applied Soft Computing, Vol. 18, 2014, pp. 261–276.
[16]. L. Li, H. Peng, J. Kurths, Y. Yang, and H. J. Schell nhuber, " Chaos–order transition inforagingbehavior of ants," Proceedings of the National Academy of Sciences, Vol.111, No. 23, 2014, pp. 392–8397.
[17]. Aya Hossam, Hany M. Harb, and Hala M. Abd El Kader, " Performance Analysis of Breast Cancer Imaging Techniques," International Journal of Computer Science and Information Security (IJCSIS), Vol. 15, No. 5, 2017, pp. 48-56.

[18]. Aya Hossam, Hany M. Harb, and Hala M. Abd El Kader, " Automatic Image Segmentation Method for Breast Cancer Analysis Using Thermography,"
[19]. PROENG (2012). Image processing and image analyses applied to mastology. http://visual.ic.uff.br/en/proeng/.
[20]. J. Kennedy, and R. Eberhart, "Particle swarm optimization," Proceedings of the IEEE International Conference on Neural Networks, 1995, pp. 1942-1948.
[21]. G.Pranava, P.V.Prasad," Constriction Coefficient Particle Swarm Optimization for Economic Load Dispatch with Valve Point Loading Effects", 2013 International Conference on Power, Energy and Control (ICPEC).
[22]. C. Cortes, and V. Vapnik, "Support-vector networks," Machine learning, Vol. 20, No. 3, 1995, pp. 273–297.
[23]. L. Auria and R. A. Moro, "Support Vector Machines (SVM) as a Technique for Solvency Analysis," in DIW Wochenbericht, No. 811, 2008, pp. 1-18.
[24]. H. Truong, "Neural networks as an aid in the diagnosis of lymphocyte-rich effusions," Anal Quant CytolHistol, Vol. 17, No. 1, 1995, pp.48-54.
[25]. Ben-Gal, and Irad, "Bayesian Networks". In Ruggeri, Fabrizio; Kennett, Ron S.; Faltin, Frederick W. Encyclopedia of Statistics in Quality and Reliability. John Wiley & Sons. doi:10.1002/9780470061572.eqr089.ISBN 978-0-47001861-3.
[26]. WEKA: Data Mining Software in Java (2008), http://www.cs.waikata.ac.nz/ml/weka.

IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) is UGC approved Journal with Sl. No. 5016, Journal no. 49082.

Aya Hossam "A Sub-Optimum Feature Selection Algorithm for Effective Breast Cancer Detection Based On Particle Swarm Optimization." IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) 13.3 (2018): 01-12.