Breast Cancer Recognition using Association Rules and Optimized Support Vector Machine

Sharafeddin Toumajpour, Abolfazl Ebrahimi
Department of Mechnical Engineering, Payame Noor University (PNU) P.O. Box19395-3697 Tehran, Iran
sharafeddin.toumajpour@gmail.com
Department of Electrical Engineering, Sirjan Branch Islamic Azad University, Sirjan, Iran.
abolfazl.ebrahimi32@yahoo.com

Abstract- Breast cancer is an occurrence of malignant neoplasm within human female breast tissue. Breast cancer is the most prominent known killer of women between the ages of 35 and 54 and males are also sometimes diagnosed with breast cancer. Effective diagnosis of breast cancer remains a major challenge and early diagnosis is extremely important in helping prevent the most serious manifestations of the disease. In this paper a hybrid system is proposed to early and accurate breast cancer detection. The proposed system is based on association rules (AR) and support vector machine (SVM). In the proposed method, AR is used for reducing the dimension of breast cancer database and SVM is used for intelligent classification. This has been confirmed Gaussian radial basis function (GRBF) has better performance among other kernel functions. In SVM with GRBF kernel function, the kernel width has very important roles for its accuracy. For this purpose, we used particle swarm optimization (PSO) algorithm to find the optimal value of kernel width. To evaluate the performance of proposed hybrid system (AR+SVM+PSO), we used Wisconsin breast cancer database (WBCD). The simulation results show that the proposed system has excellent performance in breast cancer detection.

Index Terms— Breast cancer, kernel function, SBM, PSO, kernel function.

1 INTRODUCTION

Breast cancer is the most common cancer and cause of cancer death in women in the world [1]. It also leads in terms of numbers and complexity of available treatment options resulting in decision making difficulties regarding the most appropriate treatment choice [2]. Methods have been developed to assist in predicting outcome and to support clinical decision making in breast cancer management. It is confirmed that the early detection and accurate diagnosis of this disease can ensure a long survival of the patients [3].

Breast cancer is the abnormal growth of cells occurring in breast tissue. The term "breast cancer" refers to a malignant tumor that either begins in the cells of the lobules, which are the milk- producing glands, or the ducts, the passages that drain milk from the lobules to the nipple. Less commonly, breast cancer can begin in the stromal tissues, which include the fatty and fibrous connective tissues of the breast [4].

Mammography is a specific type of breast imaging that uses low-dose x-rays to detect cancer early – before women experience symptoms – when it is most treatable. Mammograms can be used to check for breast cancer in women who have no signs or symptoms of the disease. Screening mammograms usually involve two x-ray pictures, or images, of each breast. The x-ray images make it possible to detect tumors that cannot be felt. Screening mammograms can also find micro calcifications (tiny deposits of calcium) that sometimes indicate the presence of breast cancer [5, 6].

Mammograms can also be used to check for breast cancer after a lump or other sign or symptom of the disease has been found. Besides a lump, signs of breast cancer can include breast pain, thickening of the skin of the breast, nipple discharge, or a change in breast size or shape; however, these signs may also be signs of benign conditions. A diagnostic mammogram can also be used to evaluate changes found during a screening mammogram or to view breast tissue when it is difficult to obtain a screening mammogram because of special circumstances, such as the presence of breast implants [7].

It is still challenging for radiologists to differentiate between benign and malignant tumors. Many investigators believe that automation of mammogram screening analysis increases the rate of early detection. Recently, a number of methods have been applied to automatically identify breast lesions. In [8] a fuzzy rule-based method has been proposed to characterize the shape of nodules in mammograms. In [9], three distinct regions in mammograms are compared: i) regions containing tumors; ii) regions containing calcifications and iii) regions containing no injuries. Then authors have proposed a fuzzy inference system based on the intensity of the pixels in above three regions to determine presence and type of lesion (i.e. tumor or calcification). Although this technique provides better interpretation ability, but it suffers from the computational complexity associated with handling a large number of initially generated inappropriate and sometimes inaccurate rules [10].

In [11] author used 10-fold cross-validation with C4.5 decision tree method and achieved acceptable classification accuracy. In [12] authors used RIAC method to breast cancer diagnosis. In [13] authors used artificial immune recognition system (AIRS) and fuzzy resource allocation mechanism. Simulation results show that this method has good accuracy.

Artificial Neural Networks (ANNs) have been widely used to solve the above problems in several researches [14],[15]. In [14] authors proposed a new, genetically optimized neural network algorithm, for solving breast cancer classification. The proposed method was experimented and compared with the classical ANN applied to the WBCD database. The simulation results show that the optimization of neural network can im-

prove the recognition accuracy significantly. In [15], five different classifiers including SVM, probabilistic neural network, recurrent neural network, combined neural network and Multilayer Perceptron neural networks, were applied and simulations reveal that SVM has better performance compared with other classifiers.

Neural networks offer a number of advantages, including requiring less formal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms. One disadvantage with neural network is the difficulty in understanding how a particular classification decision has been reached and also in determining the details of how a given pattern resembles with a particular class. In addition, there is no systematic way to select the topology and architecture of a neural network. In general, this has to be found empirically, which can be time consuming [16].

Based on the published papers, there exist some important issues in the design of automatic breast cancer recognition system which if suitably addressed, lead to the development of more efficient recognizers. One of these issues is the selection of the features. The better selection of features usually results in higher retrieval accuracy. In this paper, AR technique is used as feature selection technique. AR learning is a powerful and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness [17]. Another issue is related to the choice of the classification approach to be adopted. The developed system uses SVM for recognition. In SVM with GRBF kernel function, the kernel width has very important roles for its accuracy. For this purpose, we used PSO algorithm to find the optimal value of kernel width.

The details of proposed method are presented in the next sections. The second section describes the needed concepts including feature selection technique, classifier and optimization algorithm. The details of proposed method are presented in third section. The WBCD is described in forth section. The simulation results are presented in fifth section and finally the six section concludes the paper.

2. Needed concepts

2.1 Feature selection technique

The formal statement of association rule mining problem was firstly stated in [18] by Agrawal. Let be a set of m distinct attributes, T be transaction that contains a set of items such that T⊆ I, D be a database with different transaction records Ts. An association rule is an implication in the form of X⇒Y, where X, Y⊂I are sets of items called item sets, and X∩Y =∅. X is called antecedent while Y is called consequent, the rule means X implies Y.

There are two important basic measures for association rules, support(s) and confidence(c). $I = I_1, I_2, ..., I_m$ Since the database is large and users concern about only those frequently purchased items, usually thresholds of support and confidence are pre-defined by users to drop those rules that are not so interesting or useful. The two thresholds are called minimal support and minimal confidence respectively, additional constraints of interesting rules also can be specified by the users. The two basic parameters of Association Rule Mining (ARM) are: support and confidence.

Support(s) of an association rule is defined as the percentage / fraction of records that contain X ∪Y to the total number of records in the database. The count for each item is increased by one every time the item is encountered in different transaction T in database D during the scanning process. It means the support count does not take the quantity of the item into account. For example in a transaction a customer buys three bottles of beers but we only increase the support count number of *Beer* by one, in another word if a transaction contains an item then the support count of this item is increased by one. Support (s) is calculated by the following formula:

$$Support\ (XY) = \frac{Support\ count\ of\ XY}{Total\ number\ of\ transaction\ in\ D} \quad (1)$$

From the definition we can see, support of an item is a statistical significance of an association rule. Suppose the support of an item is 0.1%, it means only 0.1 percent of the transaction contain purchasing of this item. The retailer will not pay much attention to such kind of items that are not bought so frequently, obviously a high support is desired for more interesting association rules. Before the mining process, users can specify the minimum support as a threshold, which means they are only interested in certain association rules that are generated from those item sets whose supports exceed that threshold. However, sometimes even the item sets are not so frequent as defined by the threshold, the association rules generated from them are still important. For example in the supermarket some items are very expensive, consequently they are not purchased so often as the threshold required, but association rules between those expensive items are as important as other frequently bought items to the retailer.

Confidence of an association rule is defined as the percentage/fraction of the number of transactions that contain X∪Y to the total number of records that contain X, where if the percentage exceeds the threshold of confidence an interesting association rule X⇒Y can be generated.

$$\text{Confidence } (X \mid Y) = \frac{\text{Support (XY)}}{\text{Support (X)}} \quad (2)$$

Confidence is a measure of strength of the association rules, suppose the confidence of the association rule X⇒Y is 80%, it means that 80% of the transactions that contain X also contain Y together, similarly to ensure the interestingness of the rules specified minimum confidence is also pre-defined by users. Association rule mining is to find out association rules that satisfy the pre-defined minimum support and confidence from a given database [19]. The problem is usually decomposed into two subproblems. One is to find those item sets whose occurrences exceed a predefined threshold in the database, those item sets are called frequent or large item sets. The second problem is to generate association rules from those large item sets with the con-

straints of minimal confidence. Suppose one of the large item sets is $L_{k,}$ Lk = $\{I_{1,} I_{2,...,} I_{k-1,} I_k\}$, association rules with this item sets are generated in the following way: the first rule is $\{I_{1,} I_{2,} ..., I_{k-1}\} \Rightarrow \{I_k\}$, by checking the confidence this rule can be determined as interesting or not. Then other rule are generated by deleting the last items in the antecedent and inserting it to the consequent, further the confidences of the new rules are checked to determine the interestingness of them. Those processes iterated until the antecedent becomes empty. Since the second sub problem is quite straight forward, most of the researches focus on the first sub problem. More details regarding the AR can be found in [18].

2.2. Classifier

The basic idea of SVM is to transform the data to a higher dimensional feature space and find the optimal hyper plane in the space that maximizes the margin between the two classes. Consider a training data set $\{(x_1,y_1), (x_2,y_2),...(x_M,y_M)\}$, i=1,2,…,M where M is the total number of training vectors, is the ith d-dimensional input vector, and $y_i \in \{1,-1\}$ is known target. The training of SVM involves the solution of the following quadratic optimization problem:

$$Minimize \ \frac{1}{2}W^T W + C\sum_{i=1}^{M}\xi_i \quad (3)$$

$$\text{Subject to } y_i(W^T\phi(x_i)+b) \geq 1-\xi_i \ , \quad \xi_i \geq 0 \quad (4)$$

Where ξ_i are slack variables, measuring the degree of misclassification of the sample x_i, C is the error penalty factor, penalizing the non-zero ξ_i, the bias b is a scalar, representing the bias of the hyper plane, w is the vector of hyper plane coefficients, defining a direction perpendicular to the hyper plane, the index i labels the M training cases, and the map function ϕ is a transformation to map the input vectors into a high-dimensional feature space (see Fig. 1). The optimization problem becomes a trade-off between the margin maximization and training errors minimization.

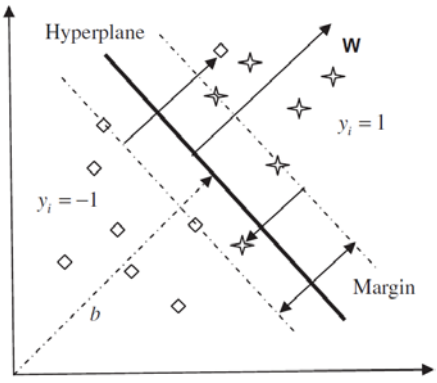

Fig. 1. A geometric interpretation of the classification of SVM for data set with two classes.

In particular, if the data are perfectly linearly separable, then $\xi_i = 0$, and the separating hyper plane that creates the maximum distance between the plane and the nearest data (i.e., the maximum- margin equals $\|w\|^{-2}$) is the optimal separating hyper plane. To solve non-linear classification tasks, a mapping function is usually employed to map the training samples from the input space into a higher-dimensional feature space. This allows the SVM to fit the maximum-margin hyper plane in the transformed feature space. Any function that satisfies Mercer's theorem [20] can be used as a kernel function. In the proposed method we used GRBF as SVM kernel function:

$$GRBG: K(x_i, y_i) = \exp(-g\|x_i - y_i\|^2) \quad (5)$$

2.3. Optimization algorithm

The basic operational principle of the particle swarm is reminiscent of the behavior of a group, for example, a flock of birds or school of fish, or the social behavior of a group of people. Each individual flies in the search space with a velocity which is dynamically adjusted according to its own flying experience and its companions' flying experience, instead of using evolutionary operators to manipulate the individuals like in other evolutionary computational algorithms. Each individual is considered as a volume-less particle (a point) in the N-dimensional search space. At time step t, the ith particle is represented as: $x_i(t)=(x_{i1}(t),x_{i2}(t),....,x_{iN}(t))$. The set of positions of m particles in a multidimensional space is identified as $X=\{x_1,....,x_j,....,x_l,....,x_m\}$. The best previous position (the position giving the best fitness value) of the ith particle is recorded and represented as $P_i(t)=(p_{i1},p_{i2},.....,p_{iN})$. The index of the best particle among all the particles in the population (global model) is represented by the symbol g. The index of the best particle among all the particles in a defined topological neighborhood (local model) is represented by the index subscript l. The rate of movement of the position (velocity) for particle i at the time step t is represented as $V_i(t) = (v_{i1}(t), v_{i2}(t), ..., v_{iN}(t))$. The particle variables are manipulated according to the following equation [21]):

$$v_{in}(t) = w_i * v_{in}(t-1) + c_1 * rand1(.) * (p_{in} - x_{in}(t-1)) + c_2 * rand2(.) * (p_{gn} - x_{in}(t-1)) \quad (6)$$

$$x_{in}(t) = x_{in}(t-1) + v_{in}(t)$$

Where n is the dimension. $(1 \leq n \leq N)$, c_1 and c2 are positive constants, rand1(.) and rand2(.) are two random functions in the range [0,1], and w is the inertia weight. For the neighborhood $(lbest)$ model, the only change is to substitute p_{ln} for p_{gn} in the equation for velocity. This equation in the global model is used to calculate a particle's new velocity according to its previous velocity and the distance of its current position from its own best experience $(pbest)$ and the group's best experience $(gbest)$. The local model calculation is identical, except that the neighbor-

hood's best experience is used instead of the group's best experience. Particle swarm optimization has been used for approaches that can be used across a wide range of applications, as well as for specific applications focused on a specific requirement. Its attractiveness over many other optimization algorithms relies in its relative simplicity because only a few parameters need to be adjusted.

3. Proposed method

In this paper an intelligent system is proposed for breast cancer tumor type recognition. This system consists of three-stages: feature selection stage, classifier stage and optimization stage. The main structure of proposed system is shown in figure 2.

Feature selection techniques are used for three reasons:

1) Simplification of models to make them easier to interpret by researchers/ users [18].

2) Shorter training times.

3) Enhanced generalization by reducing over fitting [18] (formally, reduction of variance [18]).

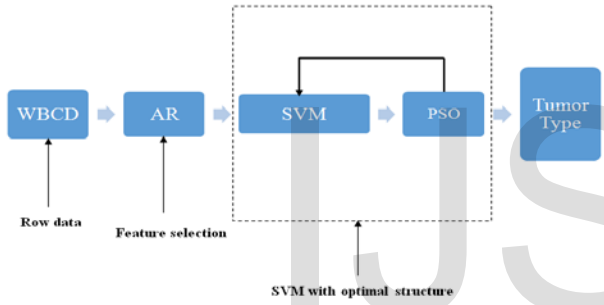

Fig 2. The main structure of proposed system

In the first stage, the input feature vector dimension is reduced and effective features selected by using AR techniques. This provides elimination of unnecessary data. In the second stage, classifier uses these inputs and classifies the breast cancer data. In classifier stage, we used SVM. As mentioned, GRBF has better performance among other kernel functions. In SVM with GRBF kernel function, the kernel width (γ) has very important roles for its accuracy. For this purpose, we used PSO algorithm to find the optimal value of γ.

4. WBCD

Breast cancer is the most common cancer among women; excluding non melanoma skin cancers. This cancer affects one in eight women during their lives. It occurs in both men and women, although male breast cancer is rare. Breast cancer is a malignant tumor that has developed from cells of the breast. Although scientists know some of the risk factors (e.g. ageing, genetic risk factors, family history, menstrual periods, not having children, obesity) that increase a woman's chance of developing breast cancer, they do not yet know what causes most breast cancers or exactly how some of these risk factors cause cells to become cancerous. Research is under way to learn more and scientists are making great progress in understanding how certain changes in DNA can cause normal breast cells to become cancerous [22]. In this study, the WBCD was used and analyzed. They have been collected by Dr. William H. Wolberg (1989–1991) at the University of Wisconsin–Madison Hospitals. There are 699 records in this database. Each record in the database has nine attributes. The nine attributes detailed in Table 1 are graded on an interval scale from a normal state of 1–10, with 10 being the most abnormal state. In this database, 458 (65.5%) records are benign and 241 (34.5%) records are malignant [22].

TABLE 1. WBCD DESCRIPTION OF ATTRIBUTES

	Attribute Description	Value of attributes	Mean		Standard deviation	
			Benign	Malignant	Benign	Malignant
1	Clump thickness	1- 10	2.956	7.195	1.674	2.429
2	Uniformity of cell Size	1- 10	1.325	6.573	0.908	2.720
3	Uniformity of cell Shape	1- 10	1.443	6.560	0.998	2.562
4	Marginal adhesion	1- 10	1.365	5.548	0.997	3.210
5	Single epithelial cell Size	1- 10	2.120	5.299	0.917	2.452
6	Bare nuclei	1- 10	1.413	7.593	1.218	3.126
7	Bland chromatin	1- 10	2.100	5.979	1.080	2.274
8	Normal nucleoli	1- 10	1.290	5.863	1.059	3.351
9	Mitoses	1- 10	1.063	2.589	0.502	2.558

5. Simulation results

In this section we test the performance of proposed system. The computational experiments for this section were done on Intel core i7 with 8 GB RAM using Dell computer. The computer program was performed on MATLAB (version R2012) environment. In order to compare the accuracy of proposed system, the k-fold cross validation technique is used. The k-fold cross validation technique was employed in the experiments, with k=10. The data set was thus divided into 10 portions, with each part of the data sharing the same proportion of each class of data. 9 data portion were used in the learning phase, while the remaining part was applied in test phase. The SVM-training methods were run 10 times to permit each part of the data to take turn as a testing data. The recognition accuracy rate is computed by summing the individual accuracy rate for each run of testing, and then dividing of the total by 10. All the obtained results are the average of 50 independent runs. The objective of the proposed method is to classify a tumor as either benign or malignant based on cell descriptions gathered by microscopic examination. As mentioned, in this study we had performed our experiment on WBCD database taken from UCI Machine Learning repository [22].

5.1. Performance of non-optimized SVM with row data

First we have evaluated the performance of the recognizer without optimization. The obtained results are listed in table 2. This table represents the SVM performance with row data. Also, the value of kernel function width is selected based in

trial and error. For this purpose we tested various values of γ. It can be seen that there is no linear relation between the value of γ and performance of SVM.

TABLE 2. THE PERFORMANCE OF NON-OPTIMIZED SVM WITH ROW DATA

γ	Accuracy (%)	γ	Accuracy (%)
0.5	96.53	5.5	95.43
1	96.25	6	95.32
1.5	94.65	6.5	94.65
2	**96.37**	7	95.66
2.5	95.23	7.5	94.27
3	94.18	8	95.87
3.5	94.36	8.5	94.77
4	93.76	9	95.87
4.5	95.76	9.5	95.54
5	96.03	10	95.28

5.2. Performance of non-optimized SVM with selected features

In next experiment, AR is applied to reduce the feature space. AR is a method to find the associations and/or relationships among items in large databases. So, we can use it to detect relations among inputs of any system and later eliminate some unnecessary inputs. We propose two different techniques to eliminate inputs. These are named as AR1 and AR2, respectively.

The AR1 technique uses all input parameters and their all records to find relations among the input parameters. If we find rules that have enough support value and high confidence value, than we can eliminate some inputs thanks to these rules. In the AR form $(X \Rightarrow Y)$ item set also depend on X item set. Thus, we can eliminate all items in Y item set. So, these are not necessary to use in SVM inputs.

Especially, we can use AR2 with classification problems. AR2 uses all input parameters but not all their records. We find only large item sets for every class. All items in these large item sets are most important items to classification. Thus, we can only use these items to classify all data. If an item of large item set of any class is large in other classes and it has different value, this item must be used as SVM inputs.

In this study, we used AR1 and AR2 to reduce the number of SVM inputs for breast cancer detection problem. We eliminated only one input parameter of SVM By using AR1 technique. Because, one of the rules is:

Input : 1-3-8-9 $\Rightarrow$ 2

Value : 1-1-1-1 $\Rightarrow$ 1 Confidence is 100%

According to this rule; if the value of 1st, 3rd, 8th and 9th input parameters are 1, the value of 2nd input parameter is 1. Then it says that 2nd input already depend on others. So we did not use

2nd input parameter in SVM input.

WBCD has two classes. These are benign and malignant classes. Using AR2, we found large item sets of benign and malignant classes given as follows:

Input: 2-8-9

Value: 1-1-1 (large item sets for benign class)

Input: 6

Value: 10 (large item for malignant class)

According to this large item set, we can say that 2nd, 8th and 9th input parameters already can define benign class and 6th input parameter can define malignant class. These parameters are the most important parameters for breast cancer detection problems. So, we only used these inputs in SVM.

The performance of non-optimized with selected features is listed in tables 3 and 4. It can be seen that the feature selection has positive effect on SVM performance.

TABLE 3. THE PERFORMANCE OF NON-OPTIMIZED SVM WITH SELECTED FEATURES (AR1)

γ	Accuracy (%)	γ	Accuracy (%)
0.5	98.42	5.5	97.96
1	98.25	6	98.27
1.5	98.54	6.5	97.88
2	97.68	7	98.37
2.5	98.35	7.5	98.17
3	**98.58**	8	98.43
3.5	98.37	8.5	97.94
4	98.36	9	98.25
4.5	98.25	9.5	98.17
5	97.28	10	98.27

TABLE 4. THE PERFORMANCE OF NON-OPTIMIZED SVM WITH SELECTED FEATURES (AR2)

γ	Accuracy (%)	γ	Accuracy (%)
0.5	97.87	5.5	97.56
1	98.03	6	97.85
1.5	97.85	6.5	97.81
2	**98.15**	7	98.04
2.5	97.76	7.5	97.94
3	98.03	8	97.85
3.5	98.04	8.5	97.79
4	97.94	9	97.58
4.5	98.11	9.5	98.02
5	97.85	10	97.96

5.3. Performance of proposed method

In this experiment, the performance of proposed method is evaluated. For this purpose, we used selected features by AR1 and AR2 as input of optimized SVM. In the optimized SVM, the value of kernel function width is selected by PSO. In the PSO, there are several coefficients whose values can be adjusted to produce a better rate of convergence. The constants c_1 and c_2 represent the weighting of the stochastic acceleration terms that pull each particle toward the *pbest* and *gbest* positions. Thus, adjustment of these constants changes the amount of `tension' in the system. Low values allow particles to roam far from target regions before being tugged back, while high values result in abrupt movement toward, or past, target regions. The inertia weight w controls the impact of the previous histories of velocities on the current velocity, thus influencing the trade-off between global (wide-ranging) and local

(nearby) exploration abilities of the `flying points' [21]. Table 5 shows the coefficient values in the PSO algorithm.

TABLE 5. COEFFICIENT VALUES IN THE PSO ALGORITHM

Number of particles	20
Acceleration constant	3
Maximum velocity	6
Maximum number of iterations	50
Size of the local neighborhood	2
Constants $c_1 = c_2$	1.8

The obtained results using proposed system are listed in table 6. It can be seen that the optimization of SVM parameter can increase the classification accuracy rate.

TABLE 4. THE PERFORMANCE OF NON-OPTIMIZED SVM WITH SELECTED FEATURES (AR2)

Input	γ	Accuracy (%)
WBCD	1.43	97.60
AR1	0.29	99.14
AR2	0.67	98.96

5.4. Comparison and discussion

For comparison purposes, table 7 gives the classification accuracies of our method and previous methods applied to the same database. As can be seen from the results, proposed method obtains excellent classification accuracy.

TABLE 7. CLASSIFICATION ACCURACIES OBTAINED WITH PROPOSED METHOD AND OTHER CLASSIFIERS FROM LITERATURE.

Ref. no	year	Method	RA (%)
[11]	1996	C4.5	94.74
[12]	1996	RAIC	95.00
[23]	1996	LDA	96.80
[24]	1999	NEF CLASS	95.06
[25]	1999	Fuzzy-GA1	97.36
[26]	2000	Neuro- rule 2a	98.10
[27]	2003	SFC	95.57
[28]	2007	LLS	96.00
[17]	2009	AR+NN	97.40
This study	2015	AR+SVM+PSO	99.14

5. CONCLUSION

One in every eight women is susceptible to breast cancer, at some point of time in her life. Early detection and effective treatment is the only rescue to reduce breast cancer mortality. Accurate classification of a breast cancer tumor is an important task in medical diagnosis. In this study we proposed an accurate system to early breast cancer recognition. In first step, non-optimized SVM is applied to WBCD. The SVM with $\gamma = 2$ led to 96.37% recognition accuracy. In next step, we applied AR technique to select the best features. The obtained results show that the application of feature selection algorithm can enhance the recognition accuracy. Finally we used PSO to find the best value of kernel function parameter. The optimized SVM with selected features led to 99.14% accuracy. The proposed system has high recognition accuracy and therefore we recommend the proposed system to breast cancer tumor type recognition.

REFERENCES

[1] Endogenous Hormones and Breast Cancer Collaborative Group. Steroid hormone measurements from different types of assays in relation to body mass index and breast cancer risk in postmenopausal women: Reanalysis of eighteen prospective studies. Steroids, Volume 99, Part A, July 2015, Pages 49-55.

[2] Y. Okumura, R. Nishimura, K. Nakatsukasa, A. Yoshida, N. Masuda, M. Tanabe, T. Shien, S. Tanaka, N. Arima, Y. Komoike, T. Taguchi, T. Iwase, H. Inaji, M. Ishitobi, Collaborative Study Group of Scientific Research of the Japanese Breast Cancer Society. Change in estrogen receptor, HER2, and Ki-67 status between primary breast cancer and ipsilateral breast cancer tumor recurrence. European Journal of Surgical Oncology (EJSO), Volume 41, Issue 4, April 2015, Pages 548-552

[3] Lifu Xiao, Qian Chen, Yangzhe Wu, Xiaojun Qi, Anhong Zhou. Simultaneous topographic and recognition imaging of epidermal growth factor receptor (EGFR) on single human breast cancer cells. Biochimica et Biophysica Acta (BBA) - Biomembranes, Volume 1848, Issue 10, Part A, October 2015, Pages 1988-1995.

[4] Jeffrey N. Weitzel. The Genetics of Breast Cancer: What the Surgical Oncologist Needs to Know. Surgical Oncology Clinics of North America, Volume 24, Issue 4, October 2015, Pages 705-732.

[5] Mark Knox, Angela O'Brien, Endre Szabó, Clare S. Smith, Helen M. Fenlon, Michelle M. McNicholas, Fidelma L. Flanagan. Impact of full field digital mammography on the classification and mammographic characteristics of interval breast cancers. European Journal of Radiology, Volume 84, Issue 6, June 2015, Pages 1056-1061

[6] Mahboubeh Madadi, Shengfan Zhang, Louise M. Henderson. Evaluation of breast cancer mammography screening policies considering adherence behavior. European Journal of Operational Research, Volume 247, Issue 2, 1 December 2015, Pages 630-640

[7] Sabee Molloi, Huanjun Ding, Stephen Feig. Breast Density Evaluation Using Spectral Mammography, Radiologist Reader Assessment, and Segmentation Techniques: A Retrospective Study Based on Left and Right Breast Comparison. Academic Radiology, Volume 22, Issue 8, August 2015, Pages 1052-1059.

[8] A. Vadivel, B. Surendiran, A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories, Comput. Biol. Med. 43 (4) (2013) 259–267.

[9] B. Bayram, U. Acar, An approach to the detection of lesions in mammograms using fuzzy image processing, J. Int. Med. Res. 35 (6) (2007) 790–795.

[10] Hemanth DJ, Vijila CK, Anitha J. Application of Neuro-Fuzzy Model for MR Brain Tumor Image Classification. Int J Biomed Imaging. 2009;16:95–102.

[11] Quinlan J. Improved use of continuous attributes in C4.5. Journal of Artificial Intelligence Research 1996;4: 779-09.

[12] Hamiton H, Shan N, Cercone N. RIAC: A rule induction algorithm

based on approximate classification. In International conference on engineering applications of neural networks, University of Regina 1996.

[13] K. Polat, S. Günes. Breast cancer diagnosis using least square support vector machine. Digital Signal Processing, 17, (2007) 694–701.

[14] Arpit Bhardwaj, Aruna Tiwari. Breast cancer diagnosis using Genetically Optimized Neural Network model. Expert Systems with Applications 42 (2015) 4611–4620.

[15] Ubeyli E. Implementing automated diagnostic systems for breast cancer detection. Expert Systems with Applications 2007;33:1054–62.

[16] Jing-Ru Zhang, Jun Zhang, Tat-Ming Lok, Michael R. Lyu. A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training. Applied Mathematics and Computation 185 (2007) 1026–1037.

[17] Karabatak M,Ince M. An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications 2009;36:3465–9.

[18] Agrawal, R., Imielinski, T., and Swami, A. N. 1993. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, P. Buneman and S. Jajodia, Eds. Washington, D.C., 207- 216.

[19] Agrawal, R. and Srikant, R. 1994. Fast algorithms for mining association rules. In Proc. 20th Int. Conf. Very Large Data Bases, VLDB, J. B. Bocca, M. Jarke, and C. Zaniolo, Eds. Morgan Kaufmann, 487- 499.

[20] Cristianini, N., & Shawe-Taylor, J. (2000). A introduction to support vector machines and other kernel-based learning methods. Cambridge, UK: Cambridge University Press.

[21] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of IEEE International Conference on Neural Networks 1995; 4:1942–1948.

[22] Nauck D, Kruse R. Obtaining interpretable fuzzy classification rules from medical data. Artificial Intelligence in Medicine 1999;16:149–69.

[23] Ster B, Dobnikar A. Neural networks in medical diagnosis: Comparison with other methods. In Proceedings of the international conference on engineering applications of neural networks 1996;427–30.

[24] Nauck D, Kruse R. Obtaining interpretable fuzzy classification rules from medical data. Artificial Intelligence in Medicine 1999;16:149–69.

[25] Pena-Reyes C, Sipper M. A fuzzy-genetic approach to breast cancer diagnosis. Artificial Intelligence in Medicine 1999;17:131–55.

[26] Setiono R. Generating concise and accurate classification rules for breast cancer diagnosis. Artificial Intelligence in Medicine 2000;18:205–17.

[27] Abonyi J, Szeifert F. Supervised fuzzy clustering for the identification of fuzzy classifiers. Pattern Recognition Letters 2003;14:2195–07.

[28] Guijarro B, Fontenla O, Perez B, Fraguela P. A linear learning method for multilayer perceptrons using least squares. Lecture Notes in Computer Science 2007;365–74.