Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2015/01/jls.htm
2015 Vol.5 (S1), pp. 468-472/Ehsan et al.
Research Article

BREAST CANCER DETECTION BASED ON A HYBRID APPROACH OF FIREFLY ALGORITHM AND INTELLIGENT SYSTEMS

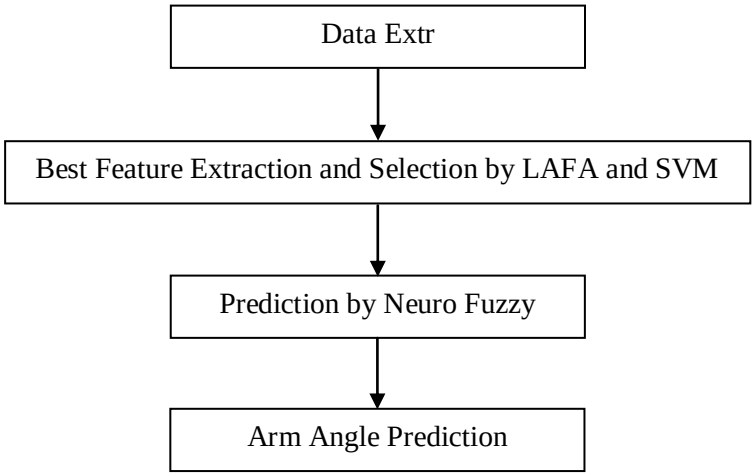
[*]Ehsan Sadeghipour[1], Nasrollah Sahragard[2], Mahmood-Reza Sayebani[2] and Rahman mahdizadeh[3]
[1]Islamic Azad University, Bandar Abbas Branch, Young Researchers and Elite Club, Bandar Abbas, Iran;
[2]Department of Electrical and Computer Engineering, University of Hormozgan, Bandar
[3]Sama Technical and Vocational Training College, Islamic Azad University,
Bandar Abbas Branch, Bandar
**Author for Correspondence*

ABSTRACT
Prediction and diagnosis of different diseases are two influential components of medical sciences. Increasing expansion of science has led to the use of decision support systems to help physician's therapeutic policies. In this article, it was tried to predict and identify breast cancer as the most common cancer among women using artificial intelligent systems. The Data were obtained from the Wisconsin Cancer Diagnosis dataset; these data have been categorized based on 9 different properties. The proposed system had an evolutionary algorithm subsystem and a data SVM mining algorithm in order to select the best properties and to separate them from ineffective properties; 4 effective properties were selected. Finally, it used the nervous fuzzy adaptive system to learn prediction and estimation. Comparing the performance of this method with other methods showed the good performance and high accuracy of the proposed system.

Keywords: *Breast Cancer, Firefly, Intelligent Systems, SVM*

INTRODUCTION
Breast cancer is the most common cancer among women and a considerable growth in the number of cases has been reported in recent years. Cancer begins with uncontrolled division of cells and results in a tumor. These tumors grow rapidly and invade and destroy the surrounding tissue. The symptoms of breast cancer include the presence of lump in breast, change in breast shape and size, difference in breast skin color, chest pain, etc. According to the National Cancer Institute of America, one in eight women develops breast cancer. In Iran, it ranks first among women cancers and the leading cause of cancer deaths in women (Anderson *et al.*, 2006). A lump of fat on the chest is a warning signal for the onset of the disease, although it is not always indicative of malignant and dangerous cancer. Early detection of breast cancer increases the chance of survival by up to 30% (Zhaohui *et al*., 2008). Traditionally breast cancer is diagnosed through fine needle aspiration (FNA) of breast fat without causing injury. It is a suitable test for determining benign or malignant nature of the disease in women. FNA is an invasive, time-consuming and expensive technique and may lead to emotional and psychological distress and concern for patient and his entourage (Mangasarian, 1995). In today's medical world, a detailed and comprehensive analysis by physician seems to be very difficult given the data volume and complexity of the signs and symptoms of the disease. Hence it seems necessary to design an intelligent system as a physician's assistant; a system which is able to discover the existing patterns and to diagnose and predict various types of the disease. Numerous studies have been carried out to predict and diagnose breast cancer. Hayward *et al.*, (2008) studied the performance of different data mining algorithms on a collection of data from various cancers using logistic regression. (Chang *et al*., 2007) presented a new model to predict breast cancer through comparing three techniques of genetic algorithms, applications of artificial intelligence, and data mining. (Einipoor *et al.*, 1388) introduced an approach for classification and detection of cancerous and non-cancerous samples with the help of fuzzy rules and an ant colony-based algorithm. Sung *et al*., (Rakkrit and Terry, 1987) presented a new method to reduce data volume and reduced 56 features of lung cancer to 11. Asmar *et al.*, (Luiza *et al.*, 2001) used the neural network technique and association rules for detection of breast cancer with an accuracy of 70%. Bellaachia *et al*.,

(2005) utilized three techniques of neural networks, decision tree, and Bayesian to predict breast cancer survival rates. Nauck and Kruse used the fuzzy method to detect breast cancer (1999).
In this paper, a general method is described first, and then the topics are studied according to the following flowchart. As is clear, the best features were selected from databank based on the data obtained by LAFA and SVM algorithms, and were used to categorize and estimate the accuracy of the results. Prediction and diagnosis of breast cancer was then determined through fuzzy neural network.



Flowchart 1: The proposed flowchart for prediction and estimation of breast cancer Firefly algorithm using learning automata

Firefly algorithm is an optimization technique and a parallel search algorithm based on population and is introduced by doctor Yang in 2008. This algorithm is inspired from fireflies which produce short flashes as a protective system and to attract mates or prey. The rate and rhythm of the flashes, as well as the time interval between them attract two sexes toward each other. The intensity of light (I) is decreased following an increase in distance (R) from light source. The transmitted light is used as the formulated objective function. Three important properties of FA algorithm are:
1) Firefly is brighter and more attractive when it moves accidentally and all fireflies are unisexual;
2) Attractiveness of firefly is proportional to the brightness and distance from it, and decrease in light intensity is calculated by light absorption coefficient γ. Brightness of firefly is determined by the objective function value.
3) The distance between each firefly is obtained from the following equation, where $X_{i,k}$ is the k^{th} component of spatial coordinates and the i^{th} firefly:

$$r_{i,j} = \sqrt{(\boldsymbol{x_i} - \boldsymbol{x_j})^2 + (\boldsymbol{y_i} - \boldsymbol{y_j})^2} \quad (1)$$

Movement of firefly and its attraction to a brighter firefly is determined as follows:

$$\boldsymbol{x_i} = \boldsymbol{x_i} + B_0 e^{\gamma r_{ij}^2} \left(x_j - x_i\right) + a(rand - 1/2) \quad (2)$$

Where a denotes the randomization parameter and rand is a random number between [0, 1]. B indicates the attractiveness of the light source. The parameter γ is determined according to changes in attractiveness and is very efficient in determining the rate of convergence. In the original version of the algorithm, each member of a group of fireflies moved toward a point where their best experience had occurred. Given the shortcomings in the original version of the algorithm, the searching action through sharing of other bugs in movement of fireflies were somewhat resolved in the improved version (Shariat *et al.*, 2012). B0 values, which reflect the attractiveness of the light source, change depending on the emitted light and in different conditions. In this method, the value of each firefly is equal to the difference between the

emitted light from that firefly in previous repeats and the best amount of emitted light obtained from other bugs. Therefore, the amount of B0 for any firefly in each stage is equal to the value of each firefly divided by the mean value of each one. More efficient values are obtained from functions using this method.

SVM Classification Algorithm

This algorithm is a supervised learning method and is used for classification and regression. This is a relatively new method and showed a good performance in recent years than the older methods of classification. The working basis of SVM classifier is linear classification in which it is trying to select a line that has a more confident margin.

The equation is solved through finding the optimal line for data by means of the QP methods well-known in solving the restricted problems. To classify high complex data, a machine should bring the data to a higher dimension space using the function φ before linear classification.

Equation (3) shows the relationship of data classifier hyper plane. Equations 4 and 5 are the relations of parallel hyper planes based on maximum-margin. When plotting these simple functions, the distance between two margin planes is $\frac{2}{||w||}$.

$$w \times x - b = 0 \quad (3)$$
$$w \times x - b = -1 \quad (4)$$
$$w \times x - b = 1 \quad (5)$$

Where x denotes the input variable, w the normal vector of separator line, and b is the intercept of separator line. Linear function is used in this study to reduce computation.

Adaptive Fuzzy Neural Network (ANFIS)

Fuzzy system corresponds to fuzzy IF-THEN rules and is not analyzable with classical probability theories. The objective of fuzzy logic is to extract accurate results using a set of rules that are defined by the expert. On the other hand, neural network is capable to learn and can determine the network parameters using the observed data, so that the desired output is achieved for the desired input. However, neural networks cannot use human knowledge and cannot conclude using linguistic expressions in contrast to fuzzy systems.

Therefore, the adaptive fuzzy-neural network called ANFIS was proposed in 1999 to achieve the learning ability of neural networks and the deduction properties of fuzzy in TSK fuzzy model. In addition to learning ability of neural network and deduction ability of fuzzy systems, these networks can find any nonlinear model or mapping which are able to precisely relate inputs to outputs. Therefore, ANFIS is a multi-layer neural network based on fuzzy system whose structure is shown in Figure 3.

Circles and squares represent the fixed and the adaptive nodes, respectively. All nodes in the first layer are adaptive and its output is the membership degree of input fuzzy.

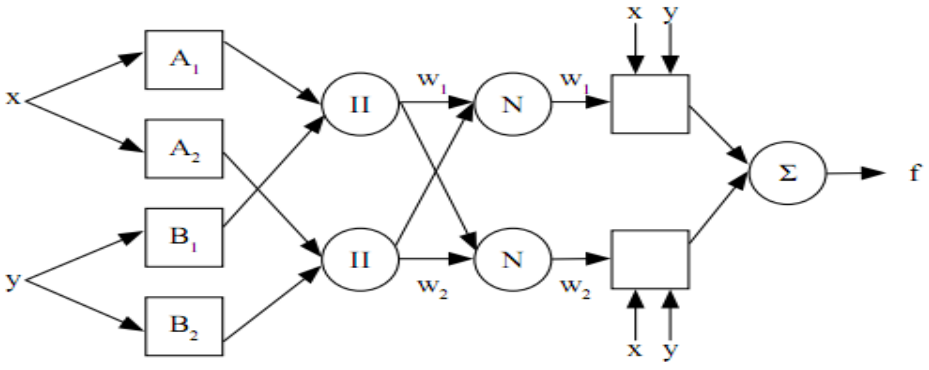


Figure 1: Neuro-fuzzy adaptive system

Procedure

The Data Used

Wisconsin Breast Cancer Dataset (WBCD) is the result of efforts made in the hospital of this university for diagnosis of breast fat obtained by FNA. This dataset contains 683 samples including 444 benign and

239 malignant tumors. Nine features obtained from FNA results are defined as parameters related to the diagnosis of benign from malignant tumors, each having an integer value between 1 and 10 (Table 1).

Table 1: Parameter related to the diagnosis of benign from malignant disease.

Row	Measured parameter	Variable name
1	Tumor thickness	X_1
2	Uniformity of cell size	X_2
3	Uniformity of cell shape	X_3
4	Marginal adhesion	X_4
5	Single epithelial cell size	X_5
6	Bare nuclei	X_6
7	Bland chromatin	X_7
8	Normal nuclei	X_8
9	Mitosis	X_9

Feature Selection

Feature selection is an issue which is considered in the field of machine learning and in statistical standpoint of pattern recognition. Feature selection is in fact choosing the features that have maximum power in prediction of the output (Jensen, 2005).

Optimal subset depends on the problems that should be solved (Newman *et al.*, 1998). A common pattern recognition system consists of 4 parts, feature extraction, feature selection, classification design and training, and testing. In order to select the best features for classification and categorization of data in this paper, the fuzzy neural network was used for implementation, simulation and testing of firefly algorithm optimization technique using learning automata.

The proposed method consisted of three steps. In the first step, LAFA algorithm with multiple random subsets started to work and tried to choose subsets with the most useful information contained. In the second step, the introduced feature vectors were used for training SVM classifier.

In order to determine the best and most effective features, different steps of the proposed algorithm (application of optimization and support vector machine algorithms) were run 400 times. Finally, four features were extracted as key features including cell nucleus division, adhesion edges, uniformity of cells, and tumor thickness.

Then, a database with 683 rows and 4 columns was sorted and applied to the fuzzy neural network to predict the correct amount of arm joint angle.

To this end, the data of 400 benign and 200 malignant cancers were used to train the network and the data of 44 benign and 39 malignant cancers were used to test it.

The generated fuzzy inference system uses the Gaussian membership function and in order to optimize the membership function, the hybrid method which is a combination of back-propagation and least squares methods was used.

The efficiency of the proposed method can be evaluated through comparing the obtained results with the results in other papers (Table 2).

Table 2: Performance of the proposed system in the diagnosis of breast cancer

RIAC	C4.5	SVM	LDA	Fuzzy-GA	Fuzzy-AIRS	Proposed system
96%	94.74%	97.2%	96.8%	97.36%	98.5%	99.14%

CONCLUSION

Given the necessity to accurately detect people with this disease, a novel method was proposed in this study based on combination of intelligent systems. The results showed the accuracy of prediction and diagnosis of breast cancer. In this study, adaptive fuzzy neural network was used for adaptation and learning, but since teaching of this type of networks is a challenging issue, a combination of evolutionary

algorithms and data mining systems can be a new idea to increase the efficiency of prediction and estimation of adaptive fuzzy neural network with higher accuracy.

REFERENCES

Anderson WF, Pfeiffer RM, Dores GM and Sherman ME (2006). Comparison of age distribution patterns for different histopathologic types of breast carcinoma. *Cancer Epidemiology, Biomarkers & Prevention* **15**(10)1899-1905.

Bellaachia Abdelghani and Erhan Guven (2005). Predicting Breast cancer survivability using Data MiningTechniques.

Einipour Amin (2009). Cancer detection using fuzzy logic and ant colony algorithm, Master in Science and Research, Khuzestan, 78.

Hayward J, Alvarez S, Ruiz C, Sullivan M, Tseng J and Whalen G (2008). Knowledge discovery in clinical performance of cancer patients. *IEEE International Conference on Bioinformatics and Biomedicine* 51–58.

Jensen R (2005). Combining rought and fuzzy sets for feature selection PhD. Thesis, School of informatics, Univ Edinburgh, 2005.

Luiza Antonie Maria, Osmar R, Za¨ıane and Alexandru Coma (2001). Application of Data Mining Techniques for Medical Image Classification.

Mangasarian O, Nick Street W and Wolberg W (1995). Breast cancer diagnosis and prognosis via linear programming. *Journal of the Operational Research Society* **43** 570-7.

Nauck D and Kruse R (1999). Obtaining interpretable fuzzy classification rules from medical data, *Artificial Intelligence in Medicine* **16** 149-169.

Newman DJ, Hettich S, Blak CLS and Merz C (1998). UCI repository of machine learning database Irvine, CA: University of California, Dep of Information and Computer Science archive ics uci edu/ml/database/pima +indian+database

Rakkrit Duangsoithong and Terry Windeatt (2009). Relevance and Redundancy Analysis for Ensemble Classifiers. Springer-Verlag Berlin Heidelberg.

Shariat Panahi Massood and Moshtaghi Yazdani Navid (2012). Improvement of firefly algorithm using learning automata, 9th international conference of industrial engineering, Tehran.

Wei Pin Chang, Der-Ming and Liou (2007). Comparison of Three Data Mining Techniques with Genetic Algorithm in the Analysis of Breast Cancer Data.

Zachman John A (1987). A Framework for Information Systems Architecture. *IBM Systems Journal* **26(**3).

Zhaohui L, Xiaoming W, Shengwen G and Binggang Y (2008). Diagnosis of breast cancer tumor based on manifold learning and support vector machine. *Proceedings of IEEE International Conference on Information and Automation* **70** 3-7.