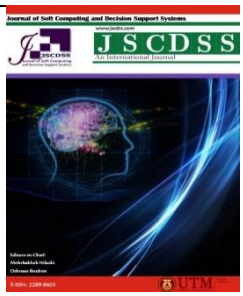
An International Journal
JSCDSS
http://www.jscdss.com
Vol.3 No.5 October 2016: 47-49
Article history:
Accepted 12 August 2016
Published online 16 August 2016

Journal of Soft Computing and Decision Support Systems



E-ISSN: 2289-8603

A New Method for Breast Cancer Diagnosis Using Neural Network and Genetic Algorithms

Atefeh Ahmadi Far [a,*]
[a] Department of Computer Engineering, Khorramabad Branch, Islamic Azad University, Lorestan, Iran

*** Corresponding author email address**: atefehahmadi68@yahoo.com

Abstract

Breast cancer in women is one of the fatal diseases that takes the lives of thousands of women around the world every year while if its common signs and symptoms are identified in time and at the early stages the patients can be easily treated. Unfortunately, the symptoms are usually detected when the cancer has developed and spread in the body and it is too late for treatment and. In addition to genetic factors, other factors such as age and weight play an important role in developing this disease. Breast cancer has hidden patterns which specialists and researchers usually fail to discover without data mining techniques. Breast cancer patients' files can contain valuable information which can be discovered through data mining techniques. One of the characteristics of data mining is the effective searching and computing large sets of data in the medical domain. The aim of this study is therefore to develop a new diagnosis system using the combination of neural networks and genetic techniques. To evaluate the proposed method, we performed several experiments on a breast cancer dataset which is available in UC Irvine machine learning repository. The experimental results show that the method can be used to obtain efficient automatic diagnostic systems for breast cancer with classification accuracy of about %98. The proposed diagnosis system can be used for early detection of breast cancer without needing to undergo clinical trial.

Keywords: Women with breast cancer, Data mining, Neural networks, Genetic algorithms, Pattern discovery, Disease prediction model

1. Introduction

Data mining is practiced to discover hidden patterns from the large amount of data (Piatetsky-Shapiro, 1996). Data mining techniques in medical science, in particular, have many applications. They can help to diagnose the disease most accurately and in the most effective way. One of the interesting, yet challenging practices in data mining is to use valuable information in different patients' records, to discover hidden patterns which can be used in providing predictive models of the disease and accordingly detecting a specific disease.

Breast cancer is the most common cancer among women; excluding non melanoma skin cancers (Karabatak, and Ince, 2009). In the several studies, statistical techniques and artificial intelligence techniques have been used to predict the breast cancer (Chen et al., 2011; Kovalerchuk et al., 1997). According to the World Health Organization, about one-third of the cancer burden could be decreased if cases are detected and treated early (Chen et al., 2011).

Neural network is one of the various and the most widely used method which is applied in data mining. Neural network learns system behavior by using system input-output data (Nilashi et al., 2014a; Nilashi et al., 2014b; Nilashi et al., 2015a; Nilashi et al., 2015b). Neural network has good generalization capabilities. The learning and generalization capabilities of neural network enable it to more effectively address real-world problems. Thus, neural network can solve many problems that are either unsolved or inefficiently solved by existing techniques (Nilashi et al., 2015a). A genetic algorithm has been used as a robust optimization method for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution.

The present study discusses how patients' data can be used by neural network for training and to building the predictive models described above to be used in disease diagnosis. Specifically, we show that how genetic algorithm can increase the predictive accuracy of prediction models of neural networks. In summary, we will show that the combination of neural network and genetic algorithm can be used effectively in diseases diagnosis systems. The present research is motivated by the high rate of breast cancer mortality, and the need for methods that can be efficient without clinical trials and diagnostic mammography, and also estimate the disease to save more lives and improve existing methods of detecting breast cancer. In addition, the most important questions that will be investigated in this study are as follows:

- Is neural network method capable to detect the symptoms and risk of breast cancer in women before metastasis?
- Is the combination of neural networks and genetic algorithms a more effective and accurate way to detect probable breast cancer?

Hence, the main objectives of this study are summarized as following:

- To present a new method for the diagnosis of breast cancer with combination of neural network and genetic algorithm.
- To improve the predictive accuracy of existing breast cancer diagnosis systems.

2. Methodology

There are many statistical models to determine the relationships between variables, which is necessary for predicting a model. In the healthcare field, the main purpose of these methods is to provide a real predictive modeling to be used in diseases' diagnosis systems. Providing predictive models in the field of medicine and especially dangerous diseases has been the subject of numerous investigations (Akay, 2009; Maglogiannis et al., 2009). One of the most important ways to diagnose and discover patterns in data mining, is through artificial neural network. The artificial neural network modeling is employed to give a better understanding and help statistical modeling to be recognized better. The success of artificial neural networks in various areas including disease diagnosis (Ozyilmaz and Yildirim, 2002) and cancer (Karabatak and Ince, 2009) has been proved. An artificial neural network consists of several components (see Fig. 1). Artificial neural network tries to find the best coefficients weights that give more realistic predictive models accurate determination of the coefficients of the neural network is a matter of optimization. To optimize the coefficients, evolutionary algorithms can be used effectively in selecting the most appropriate weights, in order to find the predictive model closest to the real model. The current study, has tried to give an explanation on predicting breast cancer in women using a combination of neural networks and genetic algorithms to determine a more accurate and efficient predictive diagnosis model.

3. Results

To evaluate the proposed method, we have performed several experiments and obtained the accuracy of prediction models. To do so, we used a real-world dataset, breast cancer dataset, which has been provided by the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg (Mangasarian et al., 1995). In Table 1, the information of dataset used in this study is presented. From this table, it can be found that totally there are 699 instances and 9 attributes along with a class label for building the prediction models.

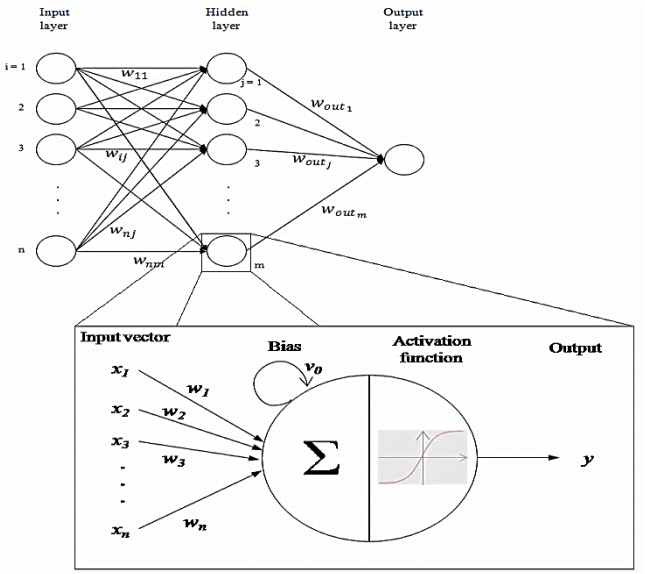


Fig. 1. A neural network with weights edges and components

Classification performance of proposed method was evaluated using the following measure.

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

In the above formula, TP, TN, FP, FN are "True Positive", "True Negative", "False Positive" and "False Negative", respectively. By using the above measure, our method achieved in overall the accuracy about 98% in identifying benign and malignant tumors.

Table 1
Dataset information

# Attribute	Domain
1. Sample code number	Id number
2. Clump Thickness	1 - 10
3. Uniformity of Cell Size	1 - 10
4. Uniformity of Cell Shape	1 - 10
5. Marginal Adhesion	1 - 10
6. Single Epithelial Cell Size	1 - 10
7. Bare Nuclei	1 - 10
8. Bland Chromatin	1 - 10
9. Normal Nucleoli	1 - 10
10. Mitoses	1 - 10
11. Class:	(2 for Benign, 4 for Malignant)
*Number of Instances: 699	
*Number of Attributes: 10 plus the class attribute	

4. Conclusion

In this study, an automatic diagnosis system for detecting breast cancer based on genetic algorithm and neural network (NN) was proposed. To evaluate the proposed method, we performed several experiments on a breast cancer dataset which is available in UC Irvine machine learning repository. The experimental results

An International Journal JSCDSS E-ISSN: 2289-8603

showed that the method can be used to obtain efficient automatic diagnostic systems for breast cancer with classification accuracy of about %98. In the future, we have plan to test the effectiveness of the proposed method on large datasets. In addition, in the future study, we will apply the dimensionality reduction techniques before prediction models to further improve the predictive accuracy of the method.

References

Akay, M. F. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications,36(2), 3240-3247.

Chen, H. L., Yang, B., Liu, J., & Liu, D. Y. (2011). A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis.Expert Systems with Applications, 38(7), 9014-9022.

Karabatak, M., & Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. Expert systems with Applications, 36(2), 3465-3469.

Kovalerchuk, B., Triantaphyllou, E., Ruiz, J. F., & Clayton, J. (1997). Fuzzy logic in computer-aided breast cancer diagnosis: analysis of lobulation.Artificial Intelligence in Medicine, 11(1), 75-85.

Maglogiannis, I., Zafiropoulos, E., & Anagnostopoulos, I. (2009). An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. Applied intelligence, 30(1), 24-36.

Mangasarian, O. L., Street, W. N., & Wolberg, W. H. (1995). Breast cancer diagnosis and prognosis via linear programming. Operations Research, 43(4), 570-577.

Nilashi, M., bin Ibrahim, O., & Ithnin, N. (2014a). Hybrid recommendation approaches for multi-criteria collaborative filtering. Expert Systems with Applications, 41(8), 3879-3900.

Nilashi, M., bin Ibrahim, O., & Ithnin, N. (2014b). Multi-criteria collaborative filtering with high accuracy using higher order singular value decomposition and Neuro-Fuzzy system. Knowledge-Based Systems, 60, 82-101.

Nilashi, M., Ibrahim, O. B., Ithnin, N., & Zakaria, R. (2015a). A multi-criteria recommendation system using dimensionality reduction and Neuro-Fuzzy techniques. Soft Computing, 19(11), 3173-3207.

Nilashi, M., Zakaria, R., Ibrahim, O., Majid, M. Z. A., Zin, R. M., Chugtai, M. W., ... & Yakubu, D. A. (2015b). A knowledge-based expert system for assessing the performance level of green buildings. Knowledge-Based Systems, 86, 194-209.

Ozyilmaz, L., & Yildirim, T. (2002, November). Diagnosis of thyroid disease using artificial neural network methods. In Neural Information Processing, 2002. ICONIP'02. Proceedings of the 9th International Conference on (Vol. 4, pp. 2033-2036). IEEE.

Piatetsky-Shapiro, G. (1996). Advances in knowledge discovery and data mining (Vol. 21). U. M. Fayyad, P. Smyth, & R. Uthurusamy (Eds.). Menlo Park: AAAI press.

An International Journal JSCDSS E-ISSN: 2289-8603