Normalized Neural Networks for Breast Cancer Classification

Emina Alickovic and Abdulhamit Subasi

Abstract

In almost all parts of the world, breast cancer is one of the major causes of death among women. But at the same time, it is one of the most curable cancers if it is diagnosed at early stage. This paper tries to find a model that diagnose and classify breast cancer with high accuracy and help to both patients and doctors in the future. Here we develop a model using Normalized Multi Layer Perceptron Neural Network to classify breast cancer with high accuracy. The results achieved is very good (accuracy is 99.27%). It is very promising result compared to previous researches where Artificial Neural Networks were used. As benchmark test, Breast Cancer Wisconsin (Original) was used.

Keywords

Breast cancer diagnosis • Supervised learning • Normalized neural networks • Multi layer perceptron • Breast cancer Wisconsin

1 Introduction

Cancer is a set of illnesses where body cells grow, alter, and multiply without control. As a rule, name of the cancer comes from the part of the body where it originated. Due to this, breast cancer refers to the unpredictable and often fast enlargement of cells that begin in the breast tissue. A cluster of rapidly separating cells may form a mass of extra tissue, called tumors. Tumors can either be cancerous (malignant) or non-cancerous (benign). Malignant tumors travel through healthy body tissues and destroy them.

The term, breast cancer, refers to a malignant tumor that has developed from cells in the breast. It is the most common cancer among women in almost all parts of the world. But if it is discovered in the earlier stages, chance to cure it are very high. According to statistics, early stage detection and treatment results in a 98% survival rate but if ti is detected in metastases this plummets to 27% [1].

In reality, one in eight women in the USA might expect to develop breast cancer during the life time [2]. Although in Bosnia and Herzegovina we do not have single register at government level, according to the reports of cantonal health care and hospital registers, breast cancer is the most common malign illness in our country [3]. Therefore, there is great need to develop a technique that will diagnose and classify breast cancer with high accuracy.

Till now, several different techniques have been used for breast cancer diagnosis. One of the most widely used techniques is mammography, but in literature, radiologists show significant differences in interpreting it [4]. Another widely used technique is Fine Needle Aspiration Cytology (FNAC) but its bad side is its modest accuracy rate (around 90%). Therefore, there is a need to develop another technique that will provide better performance for classification of breast cancer. The response to this need is usage of statistical techniques and artificial intelligence techniques. Here we define all data into two groups, either benign (that does not have cancer) or malignant group (strong evidence of having breast cancer). Due to this reason, breast cancer diagnosis can be discussed as classification problem [5–8].

Many researchers used different statistical and artificial intelligence techniques to predict and classify breast cancer techniques. Karabatak and Ince [9] used association rules and neural network to classify breast cancer pattern and they achieved the highest accuracy of 97.4%. Quinlan [10] used 10-fold cross-validation with the C4.5 decision tree method to reach accuracy of 94.74%. Goodman [11] used three different methods, optimized learning vector quantization

E. Alickovic
Department of Electrical Engineering, Linkoping University, 58183 Linkoping, Sweden
e-mail: emina.alickovic@liu.se

A. Subasi (✉)
Department of Information Systems, College of Engineering, Effat University, Jeddah 21478, Saudi Arabia
e-mail: absubasi@effatuniversity.edu.sa

A. Badnjevic et al. (eds.), *CMBEBIH 2019*, IFMBE Proceedings 73,
https://doi.org/10.1007/978-3-030-17971-7_77

(LVQ), big LVQ, and artificial immune recognition system (AIRS) to obtained accuracies of 96.7, 96.8, and 97.2%, respectively. Setiono [12] used feed forward neural network rule extraction algorithm to obtain accuracy of 98.1%. Abbas [13] used Memetic Pareto ANN (MPANN) to achieve accuracy of 98%. Aličković and Subasi [14] employed genetic algorithms in order to eliminate insignificant features. Besides they used different data mining techniques for breast cancer detection. They obtained the highest classification accuracy using the Rotation Forest model with GA-based 14 features.

Artificial Neural Networks (ANNs) are computational models consisted of set of interconnected units called neurons. It is inspired by biological neural system, although functions of ANNs might be quite different from biological neural systems. Many researches on medical diagnosis of breast cancer with Wisconcin breast cancer dataset have been done by using Artificial Neural Networks (ANNs) in recent years, and most of these researches reported high classification accuracy.

In this research, we propose normalized Multi Layer Perceptron (MLP) neural network to classify types of breast cancer. This method consists of two stages. In the first stage, we normalized our dataset and then in the second stage we used Multi Layer Perceptron for classification. We use Wisconsin breast cancer (Original) dataset as benchmark test.

This paper is organized as follows. In Sect. 2, we give description of dataset used. In Sect. 3, theoretical background about Artificial Neural Networks, Multilayer Perceptron and Normalization is explained. In Sect. 4 we present out experimental results. In Sect. 5, we give final conclusion and possible future improvements.

2 Wisconsin Breast Cancer Database Overview

Breast cancer is one of the most spread cancers among women. Based on rates from 2005–2007, one in eight women is affected by this cancer during their lifetimes [15]. Breast cancer can also occur in man, although it is not that common. Although some of the risks such as aging, genetic risk factors, family history, menstrual periods, not having children, obesity, etc. that increase chances for development of breast cancer are known, it is not known how these risk factors causes cells to become cancerous. Many researches are being done currently to answer this question and understanding how certain alterations in DNA can cause normal breast cells to develop into cancerous is in a great progress [9].

Performance is evaluated based on the model using the Wisconsin breast cancer dataset (Original) to classify the types of breast cancer as either benign or malignant. This dataset contains nine features summarized in Table 1 and each of these features is represented by some number between 1 and 10. These data is collected by Dr. William H. Wolberg at the University of Wisconsin—Madison and this dataset can be found on UCI Machine Learning Repository. Hospitals and it contains 699 records taken from 699 different persons and 241 (65.5%) records are malignant and 458 (34.5%) records are benign. Out of these 699 records, it contains 16 instances with missing attribute values. We tested out proposed method on set containing 683 data to prove efficiency of our method.

3 Theoretical Background

3.1 Artificial Neural Networks (ANNs)

ANN is an attempt to model information processing capabilities of nervous systems. It is a set of interconnected simple computation units called neurons. Link connecting neurons have weights. It performs useful computations through a process of learning, function that modifies these weights of the network to achieve desired high performances. The output of each neuron is computed by using an activation function such as sigmoid and step. ANNs are trained by experience, when applied an unknown input to the network it can generalize from past experiences and give new result [16–18]. There is an art when we design neural network because we need to pay attention to many parameters. When we design ANN, we first need to decide about ANN model, then we need to decide about number of layers and then about the number of neurons in each layer.

3.2 Multi Layer Perceptron (MLP)

Multi Layer Perceptron is a neural networks consisting of input layer, one or more hidden layers and output layer. It is one of most widely used supervised learning neural network architecture. The input signal travels from beginning to end of the network in forward direction. So, for MLP networks, output of one layer is applied to be input of the following layer. Figure 1 illustrates the architecture of MLP. The output of the MLP network is given in Eq. (1), where M is the number of the layers in the network. MLPs have been applied to solve many diverse problems with high efficiency by training them in a supervised manner using error back propagation algorithm [18].

$$a_m^{m+1} = f_m^{m+1}\left(Wm^{+1} a_m^m + b_m^{+1}\right);\ m = 1, \ldots, M-1 \quad (1)$$

Back propagation algorithm is a learning that consists of two passes (forward and backward pass) through an entire

Table 1 MLP architecture and training parameters

Architecture	
The number of layers	3
The number of neuron on the layers	Input: 9
	Hidden layer: 5
	Output layer: 1
The initial weights and biases	Random
Activation functions	Linear
Training parameters	
Learning rule	Levenberg–Marquardt
	Back propagation
Performance	0.01
Momentum	0.6
Learning Rate	0.2
Maximum number of epochs	100

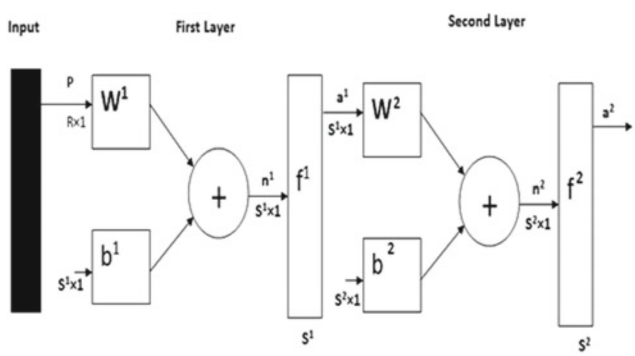

Fig. 1 Two-layer MLP network where P is an input vector, W is the weight vector, S1 and S2 are the number of neurons in the first and second Layer respectively, a is the output of the layer and f is transfer function

network. In the forward pass, input vector is applied to sensory nodes and as output it produces the actual response of the network. In this step, weights are all fixed. In the backward pass, the weights are adjusted in accordance with an error-correction rule. Error signal is produced by subtracting network response from a target (desired) response and this error signal is propagated through the network. Weights are adjusted to make network response close to desired response as much as possible [18].

3.3 Normalization

Training of ANNs could be made more proficient if particular preprocessing steps on the network inputs and targets are done. Network input processing functions converts inputs into a structure more appropriate for the network usage. The normalization process for the raw inputs has large impact on data preparation to be more fitting for the training. If normalization was not used, training of the ANNs would have been very time-consuming. There are several different types of data normalization. Data minimization can be used to scale data in the same range, so each input attribute has minimized bias within ANN. It can also accelerate training time by starting the training process for all attributes within the equal range. Different techniques use different rules such as min max rule, sum rule, product rule, etc. [19]. In this paper, we will discuss two different normalization techniques: Min-Max Normalization and Mean and Standard Deviation Normalization.

(a) *Min Max Normalization*: this method rescales attributes from one range to another. Often, features are rescaled to be in range [0 1] or [−1 1]. It is accomplished by using linear interpretation formulas such as Eq. (1).

$$x' = (x_{\max} - x_{\min}) \times \frac{(x_i - x_{\min})}{(x_{\max} - x_{\min})} + x_{\min} \qquad (2)$$

When min-max normalization is applied, all attributes will be within the same new range. Good side of this normalization is preserving exactly all relationships in the data [19].

(b) *Mean and Standard Deviation Normalization*: Another approach is to normalize mean and standard deviation of the training set. It normalizes inputs and targets in such a way that they will have zero mean and unity standard deviation. After training set is pre-processed by using normalization, then these new inputs are used to train network.

4 Experimental Results

In this section, evaluation of performance of our proposed model was done on Wisconsin breast cancer (Original) dataset. As it was already mentioned, this data set contains 699 records with 16 records having missing data. Our proposed model is given in Fig. 2.

As it can be seen in Fig. 2, we first normalized out training data using Mean and Standard Deviation Technique. Then this normalized data was used to improve Multi Layer Perceptron NN performances. By using Normalization technique together with Neural Networks, we developed a technique that is able to classify breast cancer types with very high accuracy rate. It is worth of mentioning here that we achieved significant improvement for classification by using first normalization and then MLP.

Simulation has been made by using MATLAB 2008. A variety of networks were developed and tested with random initial weights. The network is trained one hundred times. 80–20% was used. Eighty percent of records (546 records) were used for training and twenty percent of records (137 records) were used for testing. The training data set are used for learning the breast cancer pattern and then generate the decision rule(s). The testing data set which have not been used to train the system that are used to test the results.

Experimental results are taken as average of 100 runs and measured in term of classification accuracy Training parameters that gave us the highest accuracy can be seen in Table 1. While making experiments, different values of parameters were used, and the results achieved can be seen in Table 2. As we can see from table two, the most important impact on performances of our system is the choice of transfer function. If we used any other function except linear, or combination of linear and other function. Accuracy rate is very poor. In average, it is around 75%. But, if we use linear transfer function, accuracy rate is extremely high. It is 99.27%., and this is one of the highest accuracies achieved in researches by using Artificial Neural Networks. So, by first normalizing data in preprocessing stage and after that using a linear transfer function we achieved improvement of classification rate for around 32%., what is a very significant improvement. From this we can see that the choice of appropriate function is crucial and the most important factor for correct classification of breast cancer records. When we used transfer functions other then only linear and change other parameters such as learning rate or momentum, we got certain change in average accuracy, but very small and accuracy achieved was very low. This can also be seen in Table 2.

After we selected only linear function for transfer function, we got enormous improvement in accuracy, 99. 27%. So, from here we can conclude that although we select different values for training parameters, the most important factor is transfer function and as transfer function, we need to use only pure linear transfer functions in order to get the highest accuracy.

The results obtained in this research have been compared to results obtained by using neural networks from previous

Table 2 Classification accuracy obtained with proposed method using different parameters

Number of neurons in the layers	Transfer functions	Learning rate	Momentum	Training function	Number of epochs	Accuracy (%)
20–20–1	tansig–tansig–purelin	0.2	0.6	tarinlm	10	75.31
20–20–1	tansig–tansig–purelin	0.2	0.6	tarinlm	100	75.01
20–20–1	tansig–tansig–purelin	0.15	0.5	tarinlm	100	75.27
20–20–1	tansig–tansig–purelin	0.1	0.6	tarinrp	100	75.19
20–20–1	logsig–tansig–purelin	0.1	0.6	taringdm	50	73.36
20–20–1	logsig–tansig–purelin	0.2	0.6	tarinlm	50	75.26
20–1	logsig–purelin	0.2	0.6	tarinlm	50	75.79
20–1	tansig–logsig	0.15	0.6	tarinlm	50	75.18
20–1	tansig–purelin	0.15	0.5	tarinlm	50	75.36
20–1	purelin–tansig	0.15	0.6	taringdm	100	75.18
10–5–1	tansig–tansig–purelin	0.1	0.6	tarinlm	100	75.19
10–1	logsig–tansig–purelin	0.1	0.5	tarinrp	100	75.19
20–20–1	purelin–purelin–purelin	0.2	0.6	tarinlm	100	99.27
10–1	purelin–purelin	0.1	0.5	taringm	100	99.27
10–1	purelin–purelin	0.2	0.6	taringdm	50	99.27
10–1	purelin–purelin	0.2	0.6	tarinlm	50	99.27

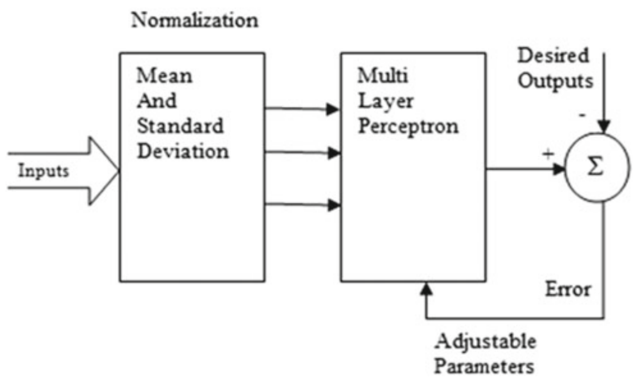

Fig. 2 Block diagram of proposed system for classification of breast cancer types

Table 3 Classification accuracy obtained with proposed method and classification accuracies obtained using neural networks in previous researches

Reference	Technique	Accuracy (%)
Proposed method	*Normalized multi layer perceptron*	*99.27*
Ubeyli [22]	Multilayer perceptron neural networks	91.92
Karabatak and Ince [9]	Association rules and neural networks	97.4
Setiono [12]	Feed forward neural network rule extraction algorithm	98.1
Ubeyli [22]	Recurrent neural network	98.15
Ubeyli [22]	Probabilistic neural network	98.61
Albrecht et al. [20]	Logarithmic simulated annealing with the perceptron algorithm	98.8
Marcano-Cedeño et al. [21]	AMMLP	99.26
Aličković and Subasi [14]	Rotation forest	96.78
Aličković and Subasi [14]	Rotation forest and GA	99.48

researches. We can see this comparison of the results done on the same dataset in Table 3. As it is shown in Table 3, this approach gave much better results than previous research done by using Artificial Neural Networks. One of the possible reasons why linear transfer function gave us the highest accuracy rate for Wisconsin breast cancer data set is the usage of normalization preprocessing.

5 Conclusion

Breast cancer is one the most widely spread cancers among women worldwide today. But, if it is diagnosed in its early stage, it can be cured very high chances. There are different methods used to diagnose and classify it, such as mammography etc. These methods have several drawbacks, and there is a need to find a better method. In recent years, artificial intelligence started to be in use in this area and it showed very good classification results. Our research is done with purpose to contribute to better classification of breast cancer. Our method tested on Wisconsin breast cancer dataset that contains nine attributes showed very good classification accuracy of 99.27%. Method we used in our research is normalization technique combined with Multi Layer Perceptron neural network. A result achieved in this research is very promising compared to the earlier reported classification techniques for mining breast cancer data. This research demonstrated that normalization of data is also very important.

We believe that this technique can be useful and can help to physicians to make very accurate diagnostic decisions. In our future investigations, we will pay a lot of attention to evaluate our proposed technique on larger data sets and to apply this algorithm on other applications such as ECG, EEG, and Intrusion Detection and test the efficiency of this algorithm. Further studies of the data can yield more remarkable outcomes. This will also be the center of attention of our prospect work.

References

1. Cancer Facts and Figures| American Cancer Society. Available https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancerfacts-figures-2009.html (2009). Accessed 31 Jan 2019 [Online]
2. Breast Cancer Risk and Risk Factors. Breastcancer.org. Available https://www.breastcancer.org/symptoms/understand_bc/risk. Accessed 31 Jan 2019 [Online]
3. Mušanovic, M., Đapo, M.: RANA detekcija raka dojke. Zavod zdravstvenog Osiguranja Kantona Sarajevo, Ministarstvo Zdravstva Kantona Sarajevo (2009)
4. Elmore, J., Wells, M., Carol, M., Lee, H., Howard, D., Feinstein, A.: Variability in radiologists' interpretation of mammograms. N. Engl. J. Med. **22**, 1493–1499 (1994)
5. Anderson, T.W.: An introduction to multivariate statistical analysis. Wiley, New York (1984)
6. Dillon, W.R., Goldstein, M.: Multivariate Analysis Methods and Applications. Wiley, New York (1984)
7. Hand, D.J.: Discrimination and Classification. Wiley, New York (1981)
8. Johnson, R.A., Wichern, D.W.: Applied Multivariate Statistical Analysis, 5th edn. Upper Saddle River, Prentice-Hall, NJ (2002)
9. Karabatak, M., Ince, M.C.: An expert system for detection of breast cancer based on association rules and neural network. Expert Syst. Appl. **36**, 3465–3469 (2009)
10. Quinlan, J.R.: Improved use of continuous attributes in C4. J. Artif. Intell. Res. **4**, 77909 (1996)
11. Goodman, D., Boggess, L.: Artificial immune system classification of multiple-class problems. In: Intelligent Engineering Systems Through Artificial Neural Networks, Fuzzy Logic, Evolutionary Programming Complex Systems and Artificial Life, vol. 12, pp. 179–184 (2002)
12. Setiono, R.: Generating concise and accurate classification rules for breast cancer diagnosis. Artif. Intell. Med. **18**, 205–217 (2000)
13. Abbas, H.A.: An evolutionary artificial neural network approach for breast cancer diagnosis. Artif. Intell. Med. **25**, 265–281 (2001)
14. Aličković, E., Subasi, A.: Breast cancer diagnosis using GA feature selection and rotation forest. Neural Comput. Appl. **28**(4), 753–763 (2017)
15. National Cancer Institute, Cancer Statistics. Available http://www.seer.cancer.gov/statfacts/html/breast.html. Last visited on 19 May 2011 [Online]
16. Bishop, C.M.: Neural Networks for Pattern Recognition. Clarendon Press, Oxford (1996)
17. Hanbay, D., Turkoglu, I., Demir, Y.: An expert system based on wavelet decomposition and neural network for modeling Chua's circuit. Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2007.03.002 (2007)
18. Haykin, S.: Neural networks, a comprehensive foundation. Macmillan College Publishing Company Inc, New York (1994)
19. Jayalakshmi, T., Santhakumaran, A.: Statistical normalization and back propagation for classification. Int. J. Comput. Theor. Eng. **3** (1), 1793–8201 (2011)
20. Albrecht, A.A., Lappas, G., Vinterbo, S.A., Wong, C.K., Ohno-Machado, L.: Two applications of the LSA machine. In: Proceedings of the 9th International Conference on Neural Information Processing, pp. 184–189. https://doi.org/10.1109/ICONIP.2002.1202156 (2002)
21. Marcano-Cedeño, A., Quintanilla-Domínguez, J., Andina, D.: WBCD breast cancer database classification applying artificial metaplasticity neural network. Expert Syst. Appl. **38**, 9573–9579 (2011)
22. Ubeyli, E.D.: Implementing automated diagnostic systems for breast cancer detection. Expert Syst. Appl. **33**, 1054–1062 (2007)