ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse42952020.pdf
https://doi.org/10.30534/ijatcse/2020/42952020

Neural Network for Breast Cancer Prediction

Mohammad Said El-Bashir[1]
[1]Faculty of Computer Science and Information Technology
Al-al Bayt University - Jordan, mohdelb@aabu.edu.jo

ABSTRACT

The aim of this study was to establish a method for breast cancer prediction, mainly based on blood test results. The study utilized data collected from 166 participants, half of whom were diagnosed with breast cancer. The data consist of information like the body mass index (BMI), age, and results of seven blood tests that encircle homeostatic model assessment (HOMA) and analyses of glucose, insulin, leptin, adiponectin, resistin, and monocyte chemotactic protein 1 (MCP-1). These seven blood tests are considered as indicators of breast cancer. However, this study was intended to identify the particular parameters of these that are most indicative of breast cancer. Two experiments were conducted by using the feed-forward, back-propagation artificial neural network. The first experiment considered all the foregoing nine parameters while the second experiment was run on only seven parameters because a decision tree for feature selection identified seven of the original nine parameters as strong predictors of breast cancer. Values of the prediction accuracy were 93.1% in the first experiment with nine predictors and 94% in the second experiment with seven predictors. Moreover, specificity and sensitivity were calculated to assess the performance of the proposed method with the two sets of predictors. The indications are promising. As such, this study has the contribution of providing rapid, non-invasive means of breast cancer prediction for early treatment.

Key words: Breast Cancer, Blood Test, Decision Tree, Neural Networks.

1. INTRODUCTION

Breast cancer is one of the diseases affecting a high proportion of females worldwide. Though, statistics of the national cancer institute for the years (2009-2015) show that the percentage of survival from this disease is 89.9% [1]. This proportion is good. However, the 10.1% of deaths necessitates finding ways to diagnose this cancer early enough to treat it. In this regard, there is consensus among experts on that regular screening for breast cancer is highly beneficial in early detection of the cancer, which helps much in treating it effectively.

The first examination of breast cancer continues to be the self-test whereby the woman checks her breast for abnormal masses by touch screening. For further screening, the X-rays may be used to detect tumors more accurately [2]. Another way is to take a biopsy (cells or tissues) from the suspected area for pathologic analysis [3]. Within this context, various studies addressed diagnosis of breast cancer (e.g., [4]), but most of the proposed methods are expensive and some of them call for surgery. An appealing approach to breast cancer diagnosis that is easy to perform and widely available is blood testing.

Bearing in mind that some tests can be made to examine the body cells in order to find early indications of cancer, notable research efforts have been, and are still, made to replace the invasive breast cancer diagnostic methods in common practice [5]. In this context, medical databases are constructed to serve as repository of data that can be employed for breast cancer prediction. Therefore, the aim of this study was to examine the potential for nine parameters to serve as reliable early indicators of breast cancer. These were age, BMI, and seven blood parameters (insulin, glucose, HOMA, leptin, resistin, adiponectin, and MCP-1). These parameters have been linked with obesity-associated breast cancer [1]. Pedictive models using data on these parameters, or some of them, will have the substantial contribution of providing rapid, non-invasive means of early breast cancer detection for early treatment [6].

The BMI is counted as one of the breast cancer diagnosis variables. It is mass of the person divided by the square of her/his height. For the adults, the normal range of BMI is 18.5-24.9 kg/m2. The blood glucose test is a simple test of the level of sugar in the blood. Normally, level of glucose in the blood is less than 100 mg/dL after fasting for eight hours. A higher than normal level of glucose means that there is less insulin than normal in the body as glucose has inverse relationship with insulin. Both of them was included in the present study as factors that influence breast cancer. In addition, the homeostatic model assessment (HOMA) depends on insulin and glucose. Its normal range is 0.5-1.4. Leptin is a hormone that regulates appetite while adiponectin is a protein hormone that regulates glucose and burns fatty acid. Resistin is a hormone that indicates the likelihood for adipose persons to have Type 2 diabetes. Its normal range is 7-22 ng/mL. Monocyte chemotactic protein 1 (MCP1) is member of the small inducible gene (SIG) family. It has role in recruitment of monocytes to the sites of injury and infection. The gene for MCP1 is located on chromosome 17q11.2-q12 [7].

Researchers around the world are working on improving prediction of breast cancer, and, therefore, its treatment. Dalamaga et al. (2013) studied the relation of breast cancer with the level of resistin in the blood serum. They found that level of resistinmay be a biomarker of the presence of an inflammatory and advanced-stage breast cancer. However, further studies are needed to decide on whether or not the level of resistin in the blood can be reliably used in breast cancer prediction [8]. A similar study was conducted in 2015 by Assiri and Kamel on leptin, resistin, and visfatin as obesity is a factor of cancer [9]. Additionally, the level of serum irisinwere found in 2015 to be good discriminator of benign and malignant breast tumors, with a sensitivity of 62.7% and a specificity of 91.1% [10]. However, it should be highlighted that, in [10], the data were not divided into two sub-sets (training and testing sub-sets). Accordingly, model performance was not assessed on different data, which raises suspicion about model performance and validity.

Characteristics of a group of women suspected to be having breast cancer have been outlined in [11]. The present study uses the same attributes. These researchers [11] subjected data on those characteristics to statistical feature selection technique for better breast cancer diagnosis. The results of their model are promising. Zadeh [12] proposed a fuzzy logic system to extract knowledge about the problem from data. Another proposed approach is a hybrid approach that combines fuzzy logic with the artificial neural network [13]. This combined (or hybrid) approach is used in various research areas, including health information systems [14].

On the other hand, other scientists approached breast cancer from other perspectives and explored potential biomarkers for its diagnosis. For instance, in 2012, Winden [15] studied 10 potential serum cancer biomarkers in an attempt to detect breast cancer at its early stage in samples collected before clinical diagnosis. These biomarkers are haptoglobin, osteopontin, carcinoembryonic antigen 15-3 (CA15–3), prolactin, cancer antigen 125, α-fetoprotein, cancer antigen 19–9, the migration inhibitory factor, and leptin. However, these biomarkers could not successfully differentiate the sample patients from the controls. Before neoadjuvant chemotherapy was developed and evaluated, a model predictive of breast cancer presented in [16] showed pathologic response of patients. Prediction of breast cancer is made by making an ultrasound near-infrared optical tomography to check tumor hemoglobin, whether it has the characteristics of the typical pathologic tumor or not [17].

Araújoet. Al. [18] constructed a hybrid artificial intelligence model that was based on concepts of fuzzy systems and neural networks to forecast breast cancer. Their hybrid model has acceptable prediction accuracy. These intelligent techniques allow for development of ‘IF-THEN’ rules as an expert system. Furthermore, the hybrid model presented in [18] can be employed to support the actions of physicians during treatment of breast cancer at its initial stages [18].

The remainder of this paper is ordered as follows: Section 2 presents the research method, including the artificial neural network (ANN), the decision tree, and a description of the dataset used in this study. Section 3 describes the two experiments made in this study, discusses their results, and compares them with results of relating previous studies. Then, Section 4 concludes this paper by presenting the conclusions drawn from the results.

2. METHODOLOGY

2.1 Data Collection

Databases of information on several parameters taken from patients at different stages of breast cancer have been created. These databases have been made available for the public and are found in the UCI Machine Learning Repository. Artificial intelligence research can be conducted using those databases to predict recurrence of breast cancer [11]. The data employed in this study were originally collected by the Department of Gynecology at the University Hospital Centre of Coimbra (CHUC) in Portugal during the period 2009-2013. The objective of the study for which those data were collected was to assess the effects of metabolic dysregulation and hyperresistinemia on breast cancer. The data were collected from women recently diagnosed with breast cancer and volunteering healthy women. A mammogram was taken for each woman and a biopsy was sent to the pathology laboratory to test for presence of the disease. None of the sample patients had underwent any treatment before the study. Moreover, none of those patients suffered from cancer, or was treated for it, or had an acute illnesses before, or during, the study. All participants signed agreement forms that were passed to the Ethical Committee at CHUC before joining the study. A total of 116 women were included in the study, comprising 64 women diagnosed with breast cancer and 52 healthy volunteers. Thirty eight of those 116 women were excluded from the study because their BMI’s were higher than 40 kg/m2 or because values of one or more of the quantitative variables necessary for the study were not available for them [11].

2.2 Artificial Neural Network

The ANN is a network of neurons that are connected to one the other as shown in Figure 1. These neurons are designed to operate like the neurons of the human brain do. Usually, the ANN is trained on a task to produce output from certain inputs. An initial weight is assigned to each neuron. This weight changes during training of the network in order to adapt with the network inputs. The training algorithms of the ANN are broadly categorized into unsupervised and supervised training algorithms. In this study, a supervised, feed-forward, back-propagation neural network is used (Figure 1) [19].

In this kind of ANN, the data are trained by applying feed-forward, back-propagation artificial intelligence models. Several network parameters need first to be calibrated and the ‘trial and error’ approach is often employed to identify the optimum network configuration, which includes the number of input layers, the number of hidden layers, the training algorithm, and the error target. In this study, performance of the back-propagation training algorithm was evaluated with a gradient descent algorithm and an adaptive learning rate.

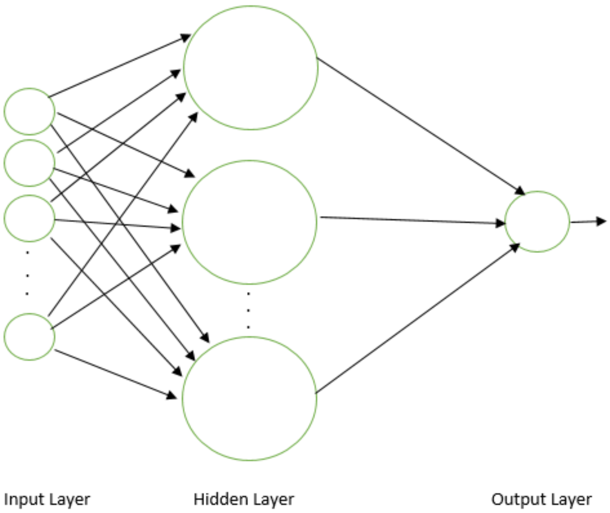

Figure 1: Artificial neural network

The feed-forward, back-propagation artificial intelligence algorithm has an input, hidden, and output layers. The back-propagation learning algorithm was used in the present study for training these networks. During training, calculations were performed from the input layer to the output layer, and the associated error values were propagated to the previous layers. The study used a three-layer network with the sigmoid function (Equation 1) hidden layer that was trained using the scaled conjugate gradient back-propagation algorithm [20].

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

2.3 Decision Tree

Approaches using the decision tree, as a machine learning method of feature selection, proved to be successful in varying practical applications. Quinlan's Interactive Dichotomizer 3 (ID3) and its variations were implemented to numerous prediction problems due to the ease of apply and understandability of the concomitant rules, i.e., decision trees [21]. The ID3 is based on the idea of using training samplesto get the classifying features. The ID3 algorithm identifies the most discriminatory feature in the data upon an entropy criterion and divides the data samples used for training into segments. This process is recursively repeated for every sub-set till it encircles the examples of only a single class. After network training, a decision tree is generated in which every node is the most discriminatory feature and every branch is a test associated with that sub-set of training samples which satisfies the test. The generated decision tree is actually a set of rules [22]. However, the ID3-based algorithmcan not model the domains which have alot of features which has continuous-valued numbers like the back-propagation algorithm [23]. Hence, the ID3 algorithm is employed in this study as an identifier of the parameters that affect breast cancer prediction the most.

2.4 The Proposed Method

In this research, the feed-forward, back-propagation neural network was applied on nine parameters (the BMI, age, insulin, glucose, leptin, HOMA, MCP.1, adiponectin, and resistin) for breast cancer prediction. A decision tree was then used to determine the parameters that are most effective in the prediction process so as to eliminate the non-related and the least-related parameters. The decision tree uncovered that only seven parameters are effective predictors of breast cancer. These are the BMI, age, insulin, glucose, resistin, adiponectin, and leptin. It turned out that MCP.1 and HOMA are not good predictors of breast cancer. In view of the decision tree outputs, the researcher re-run the feed-forward, back-propagation ANN (Figure 2) with the seven most meaningful parameters as inputs in an effort to gain improved breast cancer prediction accuracy.

3. RESULTS AND DISCUSSION

This paper proposes a model for breast cancer prediction that is based on decision tree and the feed-forward, back-propagation neural network. This model was run on nine potential predictors of breast cancer: the BMI, age, insulin, glucose, leptin, HOMA, MCP.1, resistin, and adiponectin. Data pertaining to 116 subjects were used as an input for the neural network. About half of them did not have breast cancer. These data were processed using the Matlab software. Accuracy of predicting breast cancer was 93.1%, which is a promising result. A decision tree was then employed to identify the most predictive parameters. It disclosed that only seven parameters are related to breast cancer (all the investigated parameters except HOMA and MCP.1). By applying feed-forward, back-propagation neural network to these seven parameters, higher prediction accuracy (94%) than before was achieved. However, despite having good predictions of breast cancer, the researcher recommends development of methods that can give yet better and more accurate results.

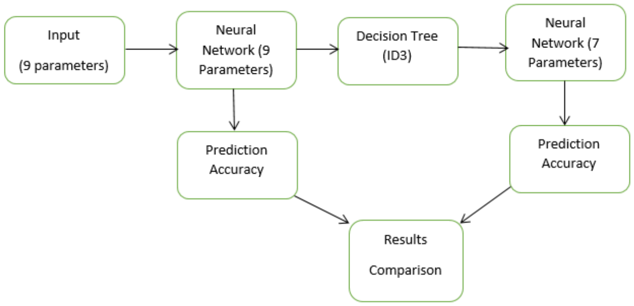

Figure 2: The proposed method

3.1 Neural Network for Nine Parameters

In the first experiment, all the aforementioned nine parameters were input to a feed-forward, back-propagation three-layer neural network that consists of one input layer, one hidden layer, and an output layer as shown in Figure 3.

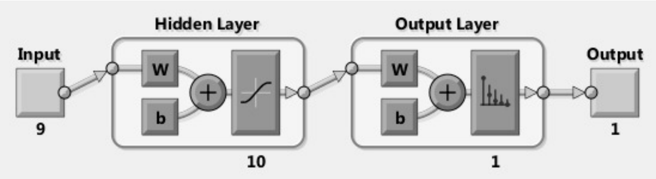

Figure 3: Architecture of the artificial neural network designed for the nine parameters

After training the neural network for several times, a prediction accuracy of 93.1% was achieved. In the case of the training sub-set, the achieved accuracy was 95.1%, corresponding to 40.2% of the subjects predicted to be having breast cancer and 54.9% of the subjects predicted not to be having it. As to the validation sub-set, the realized prediction accuracy was 94.1%, with 41.2% of the sample subjects classified as not having breast cancer and 52.9% classified as having it. Regarding the testing sub-set, the prediction accuracy was 82.4%, corresponding to 52.9% and 41.2% of the sample subjects categorized as having and not having breast cancer, respectively. It was observed that the sample data were mixed well (Figure 4).

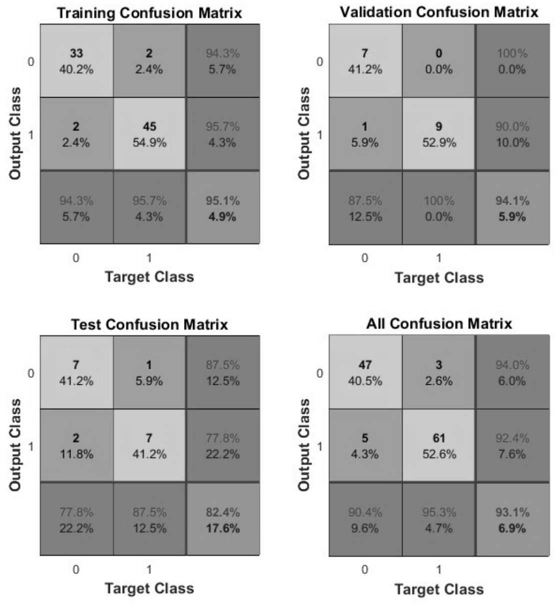

Figure 4: Confusion matrix of the model using data on the nine investigated parameters

Figure 5 is the error histogram. It unveils that small errors were associated with the training, testing, and validation sub-sets. Meanwhile, Figure 6 shows the improvement in performance brought about by network training in terms of the mean-squared error (MSE). The best network performance (lowest error (0.16995)) was associated with the validation sub-set at epoch 58

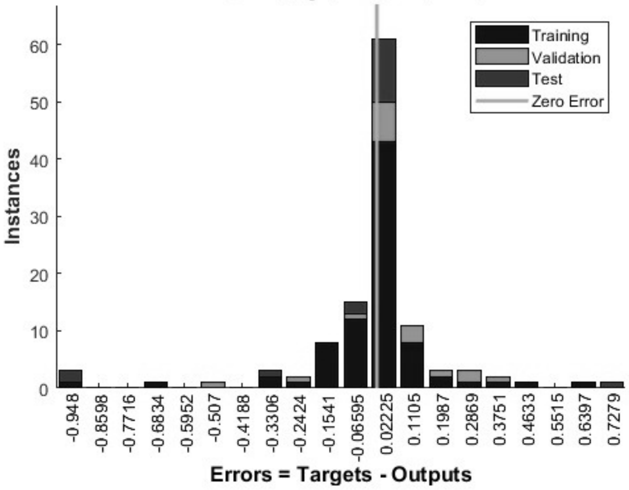

Figure 5: Error histogram

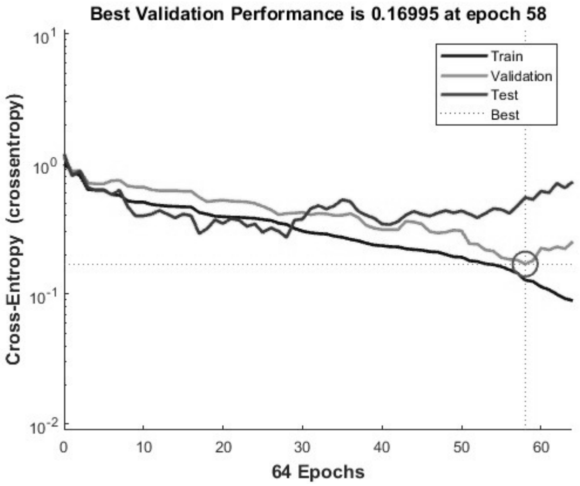

Figure 6: The MSE values associated with the training, validation, and testing sub-sets

Figure 7 is a plot of the receiver operating characteristic (ROC) curve, which gives an indication of the extent to which the neural network fit the data. The ROC curve is a plot of the false positive ratio on the x-axis against the true positive ratio on the y-axis. Since the lines for the training, validation, and testing sub-sets extend from the bottom left corner to the top right corner except for few samples (Figure 7), it is concluded that the breast cancer predictions associated with this neural network are highly accurate.

3.2 Decision Tree and Neural Network for Seven Parameters

To identify the particular parameters that influence breast cancer prediction the most, a decision tree was applied to the original nine parameters. The results depicted in Figure 8 point out that only seven of the nine parameters are highly related to breast cancer. These parameters are age, BMI, glucose, insulin, leptin, adiponectin and resistin. The HOMA and MCP.1 turned out not to be strong predictors of breast cancer [24].

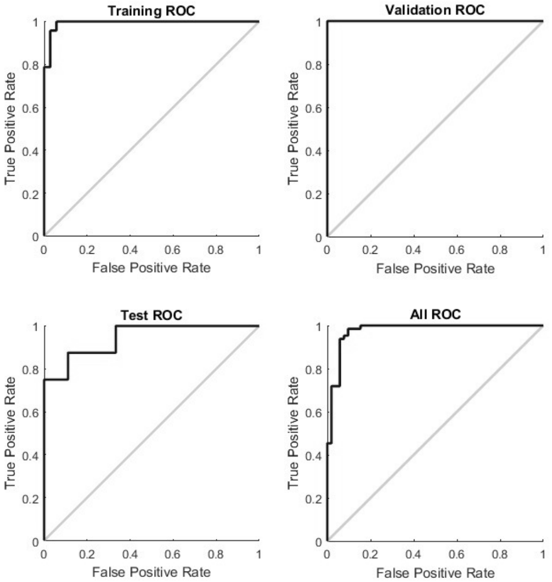

Figure 7: Receiver operating characteristic curve for the training, validation, and testing sub-sets

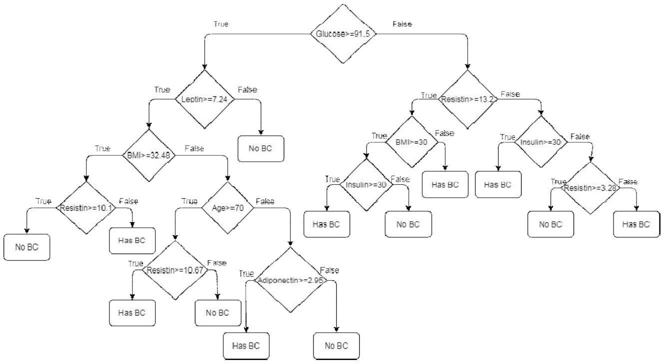

Figure 8: Decision tree

The seven parameters were used as inputs to the feed-forward, back-propagation, three-layer ANN (Figure 9) and a second run was made so as to verify the results of the decision tree. This ANN consisted of one input layer with seven inputs, one hidden layer with 18 neurons, and one output layer (Figure 9).

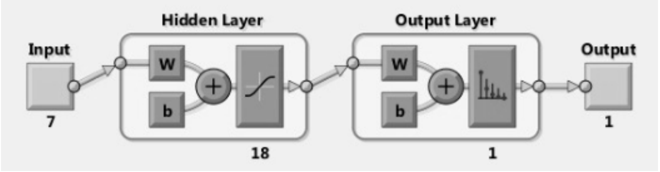

Figure 9: Architecture of the ANN designed for the seven parameters

After training the neural network for several times, an overall prediction accuracy of 94% was achieved. With respect to the training sub-set, the achieved accuracy was 92.7%, corresponding to 51.2% of the subjects identified as having breast cancer and 41.5% of the subjects categorized as not having it. As regards the validation sub-set, the realized prediction accuracy was 100%, with 35.3% of the sample subjects predicted not to be having breast cancer and 64.7% predicted to be having it. In the case of the testing sub-set, the prediction accuracy was 94.1%, corresponding to 47.1% and 52.9% of the sample subjects categorized as having and not having breast cancer, respectively. It was observed that the sample data were mixed well (Figure 10).

Figure 11 is the error histogram. It points out that small errors were associated with the training, testing, and validation sub-sets. Figure 12 illustrates the improvement in performance resulting from network training in terms of the MSE. The best network performance, that is, lowest error (0.12254), was concomitant with the validation sub-set at epoch 25.

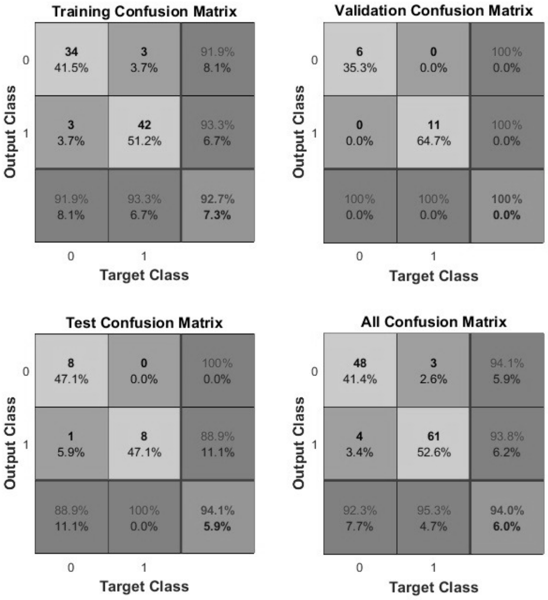

Figure 10: Confusion matrix of the model using data on the seven parameters

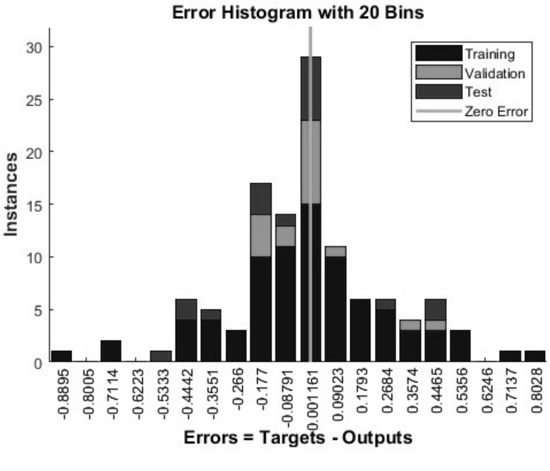

Figure 11: Error histogram

Figure 13 presents the ROC curve, which gives indication of the extent to which the designed neural network fits the data. Owing to that the lines for the training, validation, and testing sub-sets extend from the lower left corner to the top right corner except for few cases, it is concluded that the breast cancer predictions associated with this neural network are highly accurate.

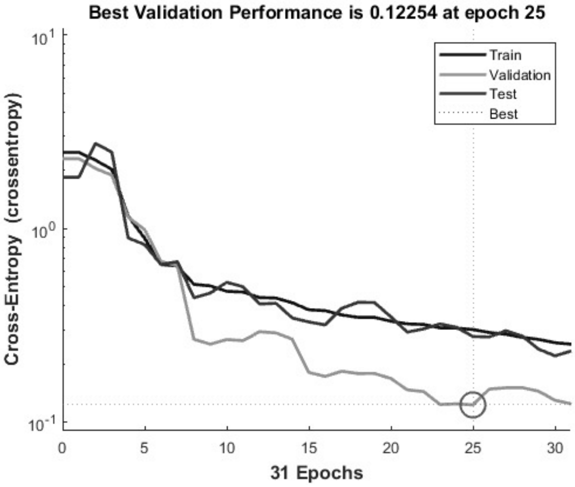

Figure 12: The MSE values associated with the training, validation, and testing sub-sets

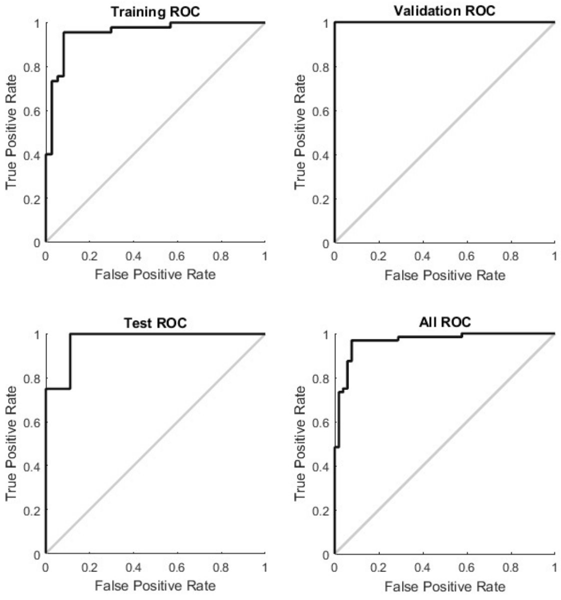

Figure 13: Receiver operating characteristic curve for the training, validation, and testing sub-sets

3.3 Comparison with Related Previous Studies

The researcher compares the performance results of the present study with those of a related previous study [11] in terms of specificity and sensitivity, Specificity denotes no misdiagnosis of the disease, that is, a low false positive ratio. The higher the sensitivity, the lower is the false negative ratio [25]. In the current study, specificity and sensitivity were calculated according to standard methods. The calculation results are listed in Table 1.

Table 1: Comparison of prediction accuracy, specificity, and sensitivity between the neural network using all nine variables and the neural network using seven variables as predictors of breast

Number of Variables	Prediction Accuracy	Specificity	Sensitivity
All nine variables	93.1%	90.3%	95.3%
Seven variables	94%	92.3%	95.3%

In [11], an approach was proposed that uses random forests, support vector machines, and logistic regression to predict breast cancer using different numbers of variables. Their results show that the prediction accuracy ranged from 87 to 91%, specificity ranged from 85 to 90%, and sensitivity fell in the range of 82-88%. In comparison with [11], both studies used the same dataset. Results of the present study indicate better prediction accuracy, both when using nine and seven variables as predictors of breast cancer. The best accuracy achieved in [11] is 91% while the best accuracy values obtained by the current study are 93.1% when using all nine variables and 94% when using seven variables as breast cancer predictors (Table 2). The best specificity value obtained in [11] was 90% while in the present study it is 90.3% when using all nine variables and 92.3% when using seven variables (Table 2). Meanwhile, the highest sensitivity value obtained by [11] was 88%. In comparison, the highest sensitivity value realized in this study is 95.3%, both when using nine and seven variables as predictors.

Table 2: A comparison between our proposed approaches and previous approach

Study	Prediction Accuracy	Specificity	Sensitivity
Patrício et al. [11]	87% – 91%	85% - 90%	82% - 88%
Present Study (9 predictors)	93.1%	90.3%	95.3%
Present Study (7 predictors)	94%	92.3%	95.3%

4. CONCLUSIONS

Regular breast screening is beneficial for early diagnosis of breast cancer. This allows for early detection and treatment of breast cancer, which is one the cancers with high recovery rates. In this research, potential of two parameters and seven blood tests to serve as reliable predictors of breast cancer was examined to help in non-invasive early detection of the presence of breast cancer. The two investigated parameters were BMI and age whereas the blood tests were the levels of insulin, glucose, leptin, HOMA, MCP-1, resistin, and adiponectin. Data on these parameters were collected by the Department of Gynecology at CHUC in Portugal in the period 2009-2013. They were taken from women recently diagnosed with breast cancer and volunteering healthy women. Overall, the data set consisted of 116 subjects, sixty four of them were diagnosed with breast cancer while the rest 52 subject were non-patient, volunteering women [10].

A feed-forward, back-propagation neural network was applied on the nine parameters to examine their potential use for successful prediction of breast cancer. Then, a decision tree was used to determine the particular parameters of these nine parameters that are most effective in the prediction process. The decision tree pointed out that only seven parameters affect the prediction process tangibly, which are the BMI, age, insulin, glucose, resistin, adiponectin, and leptin. Thus, breast cancer is not associated with HOMA and MCP.1. Applying feed-forward, back-propagation neural network on these seven parameters enhanced the prediction accuracy, which increased from 93.1% to 94%. These findings were compared with findings of a previous study ([11]) that used the same dataset. The comparison supports superiority of the present method to that of [11] because the current study employed the decision tree in order to identify and use the variables most predictive of breast cancer out of the original nine variables.

REFERENCES

1. https://seer.cancer.govlast accessed 22September 2020.
2. Kerlikowske, K.; Grady, D.; Rubin, S.M.; Sandrock, C.; Ernster, V.L. **Efficacy of screening mammography: A meta-analysis.** JAMA 1995, 273, 149–154. [CrossRef]
3. Veronesi, U.; Paganelli, G.; Galimberti, V.; Viale, G.; Zurrida, S.; Bedoni, M.; Costa, A.; De Cicco, C.; Geraghty, J.G.; Luini, A. **Sentinel-node biopsy to avoid axillary dissection in breast cancer with clinically negative lymph-nodes.** Lancet 1997, 349, 1864–1867. [CrossRef]
4. Giunti, G.; Giunta, D.; Guisado-Fernandez, E.; Bender, J.; Fernández-Luque, L. **A biopsy of Breast Cancer mobile applications: State of the practice review.** Int. J. Med. Informat. 2018, 110, 1–9.
5. Vlahou, A.; Laronga, C.; Wilson, L.; Gregory, B.; Fournier, K.; McGaughey, D.; Perry, R.R.; Wright, G.L., Jr.; Semmes, O.J. **A novel approach toward development of a rapid blood test for breast cancer.**Clin. Breast Cancer 2003, 4, 203–209.
6. Jide E. T. Akinsola , Moruf A. Adeagbo , Ayomikun A. Awoseyi. **Breast Cancer Predictive Analytics Using Supervised Machine Learning Techniques.** International Journal of Advanced Trends in Computer Science and Engineering. Volume 8, No.6, November – December 2019.
7. O.C.ZachariaeChristianG.LarsenKoujiMatsushima. **Monocyte Chemoattractant Protein 1 (MCP-1).**Encyclopedia of Immunology (Second Edition). 1998, Pages 1748-1750.
8. Dalamaga M, et al. **Serum resistin: a biomarker of breast cancer in postmenopausal women? Association with clinicopathological characteristics, tumor markers, inflammatory and metabolic parameters.**ClinBiochem. 2013;46(7–8):584–90.
9. Assiri AM, Kamel HF. **Evaluation of diagnostic and predictive value of serum adipokines: Leptin, resistin and visfatin in postmenopausal breast cancer.**Obes Res ClinPract. 2015;10(4):442-53.
10. Provatopoulou X, et al. **Serum irisin levels are lower in patients with breast cancer:** association with disease diagnosis and tumor characteristics. BMC Cancer. 2015;15(1):898.
11. Patrício, M.; Pereira, J.; Crisóstomo, J.; Matafome, P.; Gomes, M.; Seiça, R.; Caramelo, F. **Using Resistin, glucose, age and BMI to predict the presence of breast cancer.** BMC Cancer 2018, 18, 29.
12. Zadeh, L.A. **A fuzzy-algorithmic approach to the definition of complex or imprecise concepts.** In Systems Theory in the Social Sciences; Springer: Berlin, Germany, 1976; pp. 202–282.
13. Pedrycz, W. **Fuzzy neural networks and neurocomputations.** Fuzzy Sets Syst. 1993, 56, 1–28.
14. Guimarães, A.J.; Araujo, V.J.S.; de Campos Souza, P.V.; Araujo, V.S.; Rezende, T.S. **Using Fuzzy Neural Networks to the Prediction of Improvement in Expert Systems for Treatment of Immunotherapy.** In Ibero-American Conference on Artificial Intelligence; Springer: Berlin, Germany, 2018; pp. 229–240.
15. Opstal-van Winden AW, et al. **A bead-based multiplexed immunoassay to evaluate breast cancer biomarkers for early detection in pre-diagnostic serum.**Int J Mol Sci. 2012;13(10):13587–604.
16. Zhu Q, et al. **Pathologic response prediction to neoadjuvant chemotherapy utilizing pretreatment near-infrared imaging parameters and tumor pathologic criteria.** Breast Cancer Res. 2014;16(5):456.
17. Chlebowski RT, et al. **Predicting risk of breast cancer in postmenopausal women by hormone receptor status.** J Natl Cancer Inst. 2007;99(22):1695–705.
18. Silva Araújo, V.J.; Guimarães, A.J.; de Campos Souza, P.V.; Rezende, T.S.; Araújo, V.S. **Using Resistin, Glucose, Age and BMI and Pruning Fuzzy Neural Network for the Construction of Expert Systems in the Prediction of Breast Cancer.** Mach. Learn. Knowl. Extr. 2019, 1, 466-482.
19. Mohammed Y. Alzahrani Ahmed H. Alahmadi. **Breast Cancer Image Classification Using the Convolution Neural Network.** International Journal of Advanced Trends in Computer Science and Engineering. Volume 8, No.6, November – December 2019.
20. M. Moller. **A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning.** Neural Networks 6 (4): 525-533 (1993).
21. Aich, A., Dutta, A., & Chakraborty, A. (2018). **A Scaled Conjugate Gradient Backpropagation Algorithm for Keyword Extraction.** Information Systems Design and Intelligent Applications, 674–684.
22. Chi, Z., Suters, M., & Yan, H. (1996). **Handwritten digit recognition using combined ID3-derived fuzzy rules and Markov chains.** Pattern Recognition, 29(11), 1821–1833.

23. Loh, W.Y. and Y.S. Shih. **Split Selection Methods for Classification Trees.** StatisticaSinica, Vol. 7, 1997, pp. 815–840.
24. Hitarthi Bhatt et al, . **Use of ID3 Decision Tree Algorithm for Placement Prediction.** (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4785-4789.
25. Methley, A.M., Campbell, S., Chew-Graham, C. et al. **PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews.** BMC Health Serv Res 14, 579 (2014). https://doi.org/10.1186/s12913-014-0579-0.