© Universiti Tun Hussein Onn Malaysia Publisher's Office

JSCDM

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/jscdm

Journal of Soft Computing and Data Mining

Classification of Breast Cancer Patients Using Neural Network Technique

Putri Marhida Badarudin[1], Rozaida Ghazali[1], Abdullah Alahdal[2*], N.A.M. Alduais[1], Salama A Mostafa[1]

[1]Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor, MALAYSIA

[2]Department of Computer Engineering, Faculty of Computer Science and Engineering,
Hodeidah University, YEMEN

*Corresponding Author

DOI: https://doi.org/10.30880/jscdm.2021.02.01.002
Received 05 November 2020; Accepted 01 April 2021; Available online 15 April 2021

Abstract: This work develops an Artificial Neural Network (ANN) model for performing Breast Cancer (BC) classification tasks. The design of the model considers studying different ANN architectures from the literature and chooses the one with the best performance. This ANN model aims to classify BC cases more systematically and more quickly. It provides facilities in the field of medicine to detect breast cancer among women. The ANN classification model is able to achieve an average accuracy of 98.88 % with an average run time of 0.182 seconds. Using this model, the classification of BC can be carried out much more faster than manual diagnosis and with good enough accuracy.

Keywords: Artificial neural network (ANN), multi-layer, classification, breast cancer

1. Introduction

Breast cancer is considered to be one of the common cancer types in women and the second most common cancer-causing lung cancer. Breast cancer is the rapid and unstable division of breast cells, resulting in breast lumps. Then, it passes through the lymph nodes to reach other parts of the body [1]. This type of disease may occur in breast ducts, glandular tissues, or other breast tissues. It is recommended that all women regularly check for signs of breast cancer, and if there are any lumps or tissue changes in the breasts, please consult a doctor. It should also be noted that most breast lumps are benign. Numerous diagnostic procedures can be used to diagnose breast cancer, but the most effective method helps diagnose the body and the least damage to the body [2].

Neural Network or Artificial Neural Network (ANN) is an information processing method that has the same characteristics as biological neural networks. This neural network has been developed as a general conclusion for human thought's mathematical model [3]. Based on research on different types of human disease diagnosis systems such as brain tumor, Parkinson's disease, liver cancer, and breast cancer (BC), ANNs play a significant role in improving the results of such systems [4].

In 2017, Mirajkar et al. [3] proposed an expectation strategy for breast cancer that can offer assistance doctors identify a breast cancer determination based on patients' clinical data (classes were a kind tumor). The domestic dataset was used in Lagos, which included more than 1700 cases. Eleven attributes were picked, such as cell estimate, cell shape, and rank. To recognize information on breast cancer, three directed learning algorithms were utilized. C4.5, WEKA's toolkit was utilized to study Multilayer Perceptron (MLP) and Naïve Bayes.

**Corresponding author: ab.alahdal@gmail.com*

penerbit.uthm.edu.my/ojs/index.php/ijie

In 2018, Osman et al. [5] proposed programmed breast cancer symptomatic approach utilizing a two-step clustering strategy and bolster vector machine calculation. The crossover approach aims to make strides in conclusion exactness and the restorative misdiagnosis recognizable proof to understand breast-tumor-related screening problems. The two-step strategy and SVM calculation were combined and utilized to recognize the approaching tumors in order to decide the covered-up characteristics of a kind and malignant cancer. When analyzed within the UCI-WDBC, the evolved crossover strategy makes strides in its exactness by 99.1%. In 2018, Yue et al. [6] tested machine learning (ML) methods employed for the diagnosis and prognostics of breast cancer.

This work proposes an ANN model to classify BC cases into two specific malignant and benign classes that show if the person has BC. The application of this work is developed using Matlab software that can classify cancer patients quickly and accurately. The objective of this application is to develop an application that can classify breast cancer patients as well as perform the classification of breast cancer patients. For this work, data on breast cancer patients are obtained from William H. Wolberg at the University of Wisconsin Hospitals, Madison, Wisconsin, USA. This dataset has 699 classes/cases. It is downloaded from the Machine Learning Repository website. This data is intended to classify and predict whether or not breast cancer patients are infected. The number of data obtained is data.

2. Methods and Materials

2.1 Methodology

The Cross Industry Standard Process for Data Mining (CRISP-DM) methodology is carried out to find out the development requirements for the ANN model to be developed. The CRISP-DM methodology is a complete set of guidelines that contains models, facilities, and specific techniques that need to be followed when doing particular research [7]. This methodology is then documented and can be referenced for the use of the organization. The main methods are the ANN, BC dataset, and evaluation criteria. Among the core phases of the CRISP-DM is the selection of variables, data understanding, data preparation, modeling, evaluation, and implementation [8], [9]. The advantage of using this model is that problems can be identified, and the system can be improved to meet all the criteria and needs of the user as well as the system itself. Figure 1 shows the CRISP-DM methodology.

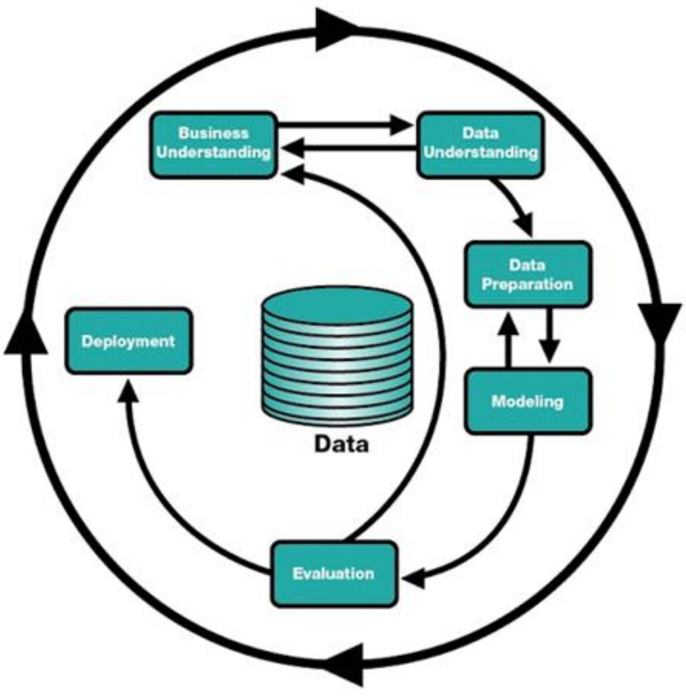


Fig. 1 - The methodology of CRISP-DM [7].

A literature review was conducted on the concept of neural networks and the classification of breast cancer patients. This literature review aims to gather more detailed information about the study to be conducted. Based on the information available, better idea generation will be made. This variable selection phase is performed to identify the type of ANN models and the model to be developed. In this phase, data selection and understanding of the data to be obtained should be made to develop The ANN model for diagnosing the BC. The related data is obtained from breast cancer patients from the University of Wisconsin Hospitals, Madison, Wisconsin, USA, was taken from the Machine Learning Repository website. In the modeling phase, the ANN and the interface were developed using MatLab R2010a. During the implementation and evaluation phases, it is important to determine performance weaknesses that can be corrected before the next phase is performed. This phase is done to identify whether the developed model successfully performs the BC diagnosis and the quality or correctness of the performs.

2.1 Neural Networks

The Artificial Neural Network (ANN), since its introduction around 1940 has been widely implemented in various scientific pawns. ANN is widely used to make predictions or predictions. Different researchers have examined the use of ANN with back-propagation training algorithms to predict cancer cases [4]. ANN with the type of feed-forward network or back-propagation used in this study has been shown to provide good results and results for prediction purposes [3]. Neural networks have two several neural network architectures, namely single-level architecture and multi-level architecture. Figure 2 shows a single-level architecture with an input layer, an output layer, and even a weighting layer. The weighting for this one-level architecture consists of only one weight.

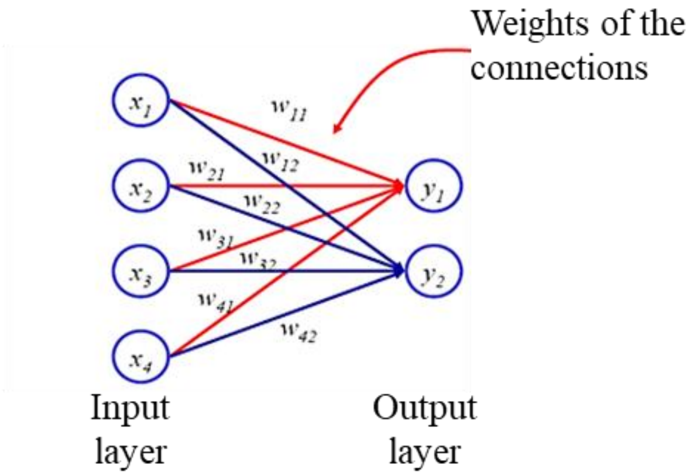


Fig. 2 - One-level neural network architecture.

Figure 3 shows a multi-level architecture where it has input layers, output layers, and even hidden layers. This architecture consists of two levels of weights, where it is in the middle between the input layer and the hidden layer and in the middle - the hidden layer and the output layer. The learning algorithm for neural networks is divided into two, namely supervised and unregulated. Supervised neural networks are networks that learn based on predefined targets as well as successful learning if the error is minimal. Examples of supervised networks are backbone networks and radial-based functions. For non-regulatory networks, it is a learning network by collecting all the same input patterns in one group, and no targets are given. Examples of non-regulatory networks are the Konohen network and the adaptive resonance theory [4]. In this research, the propagator network algorithm is chosen because the propagator network's goal is to balance the ability to respond to input patterns during learning and respond to other input patterns that are very similar to the patterns used in learning [5].

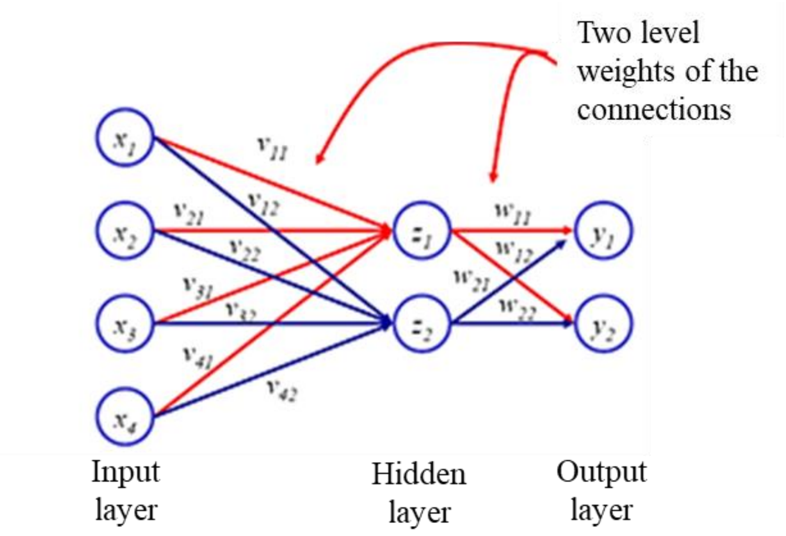


Fig. 3 - Multi-layer neural network architecture.

2.3 Breast Cancer Dataset

To precisely distinguish cancer as being generous or malignant, a collection of examples the proposed strategy tried based on four sorts of the breast cancer UCI store [10]. There are: Wisconsin Information set for Symptomatic Breast Cancer (WDBC). Features are computed from a digitized picture of a fine needle suction (FNA) of a breast mass. They describe characteristics of the cell cores show within the image [11], [12]. The fine needle aspirator (FNA) characteristics in the breast mass are calculated from digitalized pictures. Some of the pictures are distinguished by Bennett. Appropriate characteristics were nominated using a comprehensive look within the space of (1-4) characteristics and (1-3) separating planes.

3. Classification of Breast Cancer

3.1 Implementation

This classification of breast cancer is an application that is a good aid for doctors to make clinical decisions and detect whether patients are infected with cancer or not [13], [14]. The application also has specific algorithms for classification statistics that allow classification to be made in this application. The application results are the results of the classification of breast cancer patients into two different malignant and benign groups. The application contains an interface that requires the user to enter some input as input and output graphs and results in the section provided. Figure 4 shows an interface designed to meet the user's needs, and the application itself' needs, where the user must enter the value of the learning rate, node input, and momentum to run this application.

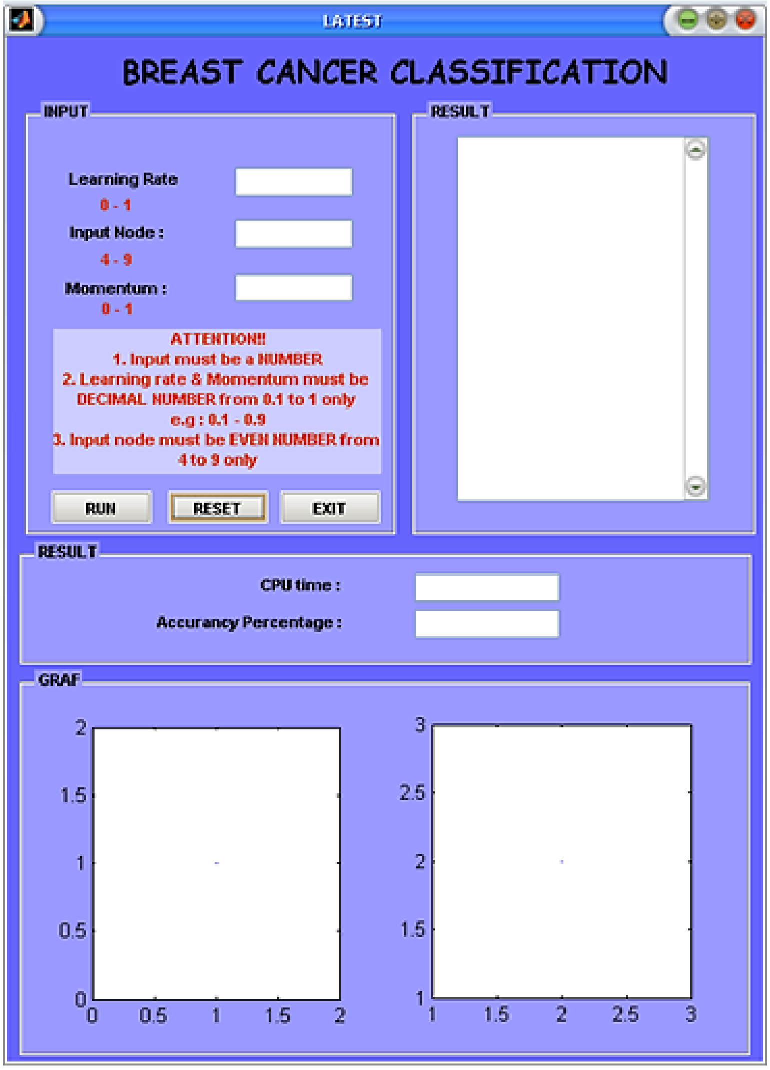
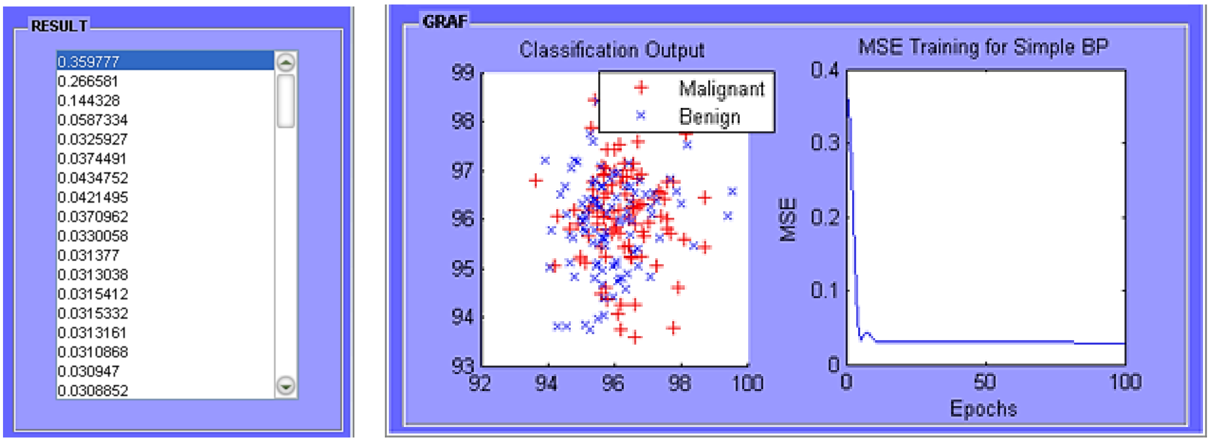


Fig. 4 - The implementation of the ANN for BC classification.

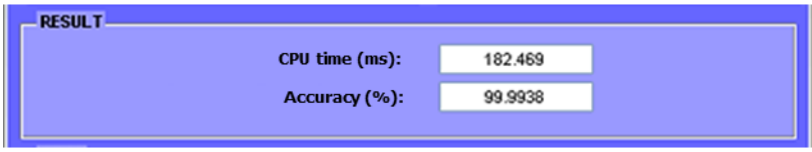
The learning rate, better known as the learning rate, will be entered by the user based on 0 and 1. This is because the learning rate is too large will cause the neural network cannot learn completely, and if the value is too small, it causes the neural network to slow down. The input node is the user input that the user can fill according to the level of complexity of the problem or data provided as well as suitability and trial methods where if the input node is too high or too low, it will give an ineffective value to the output [11]. The momentum constant is to increase the concentration rate to achieve a better solution. The value of the momentum constant lies between 0 and 1. If the momentum constant is too small, the weight change movement is too slow for it to slow down. However, too many features or values can cause a delay. A small number of values might possibly cause some solutions not taken into account.

3.2 Results and Discussion

Figure 5 (a) displays the results of the ANN based on the Mean Square Error (MSE). These results help to identify whether the ANN model diagnosis the BC with an acceptable error rate. The second graph of Figure 5 (b) is related to MSE against Epoch. in this context, the epoch represents a complete pass through the data's training set. This graph shows the BC classification outputs of malignant and benign. The graph explains that if the graph approaches zero value, the ANN meets the value to be classified or predicted. The results that are shown in Figure 5 (c) represent the average calculation of CPU run time and performance accuracy. These results help in identifying that the ANN model achieves a high-quality performance.



(a) MSE results (b) Classification Graph and MSE Graph against Epochs



(c) Execution time and accuracy results

Fig. 5 - Classification results.

The classification of breast cancer is divided into two classes, namely benign and malignant. Benign represents patients who do not have BC, while malignant represents patients who do not have BC. This ANN classification model is able to achieve an average accuracy of 98.88 %.%. The excursion speed of the model is 0.182 seconds on average.

4. Conclusion

In this research work, breast cancer (BC) classification is performed by proposing a Machine Learning solution. To ensure a quick and accurate solution, an artificial neural network (ANN) model has been proposed. The model considers studying different architectures of the ANN from the literature and chooses the one with the best performance. This breast cancer patient data is taken through William H. Wolberg, who is at the University of Wisconsin Hospitals, Madison, Wisconsin, USA, from the Machine Learning Repository website. This data is intended to classify and predict breast cancer cases with two classes: malignant and benign. Using this model, the classification of BC can be carried out much more faster than carrying out manual classification and with good enough accuracy. The ANN model achieves an average classification accuracy of 98.88 %. The excursion speed of the model is 0.182 seconds on average. The future work will consider testing other machine learning techniques such as random forest and deep learning techniques such as deep belief networks.

Acknowledgement

The authors would like to thank the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor and Department of Computer Engineering, Faculty of Computer Science and Engineering, Hodeidah University, Yemen.

References

[1] Mohammed, M. A., Al-Khateeb, B., Rashid, A. N., Ibrahim, D. A., Abd Ghani, M. K., & Mostafa, S. A. (2018). Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Computers & Electrical Engineering*, *70*, 871-882

[2] Osman, A. H., & Aljahdali, H. M. A. (2020). An effective of ensemble boosting learning method for breast cancer virtual screening using neural network model. IEEE Access, 8, 39165-39174

[3] Mirajkar, P., & Lakshmi, P. (2017). Prediction of cancer risk in perspective of symptoms using Naïve Bayes classifier. International Journal of Engineering Research in Computer Science and Engineering, 4(9), 145-149

[4] Mahmood, M., Al-Khateeb, B., & Alwash, W. M. (2020). A review on neural networks approach on classifying cancers. *IAES International Journal of Artificial Intelligence*, *9*(2), 317

[5] Osman, A. H., & Alzahrani, A. A. (2018). New approach for automated epileptic disease diagnosis using an integrated self-organization map and radial basis function neural network algorithm. IEEE Access, 7, 4741-4747.

[6] Yue, W., Wang, Z., Chen, H., Payne, A., & Liu, X. (2018). Machine learning with applications in breast cancer diagnosis and prognosis. Designs, 2(2), 13

[7] Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (pp. 29-39). London, UK: Springer-Verlag

[8] Razali, N., Mustapha, A., Abd Wahab, M. H., Mostafa, S. A., & Rostam, S. K. (2020, April). A Data Mining Approach to Prediction of Liver Diseases. In Journal of Physics: Conference Series (Vol. 1529, No. 3, p. 032002). IOP Publishing

[9] Razali, N., Mostafa, S. A., Mustapha, A., Abd Wahab, M. H., & Ibrahim, N. A. (2020, April). Risk Factors of Cervical Cancer using Classification in Data Mining. In Journal of Physics: Conference Series (Vol. 1529, No. 2, p. 022102). IOP Publishing

[10] Wolberg W H., Street W. N. & Mangasarian O. L. (1995). Breast Cancer Wisconsin (Diagnostic) Data Set, Retrieved 1 January 2018, from https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

[11] Agarap, A. F. M. (2018, February). On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In *Proceedings of the 2nd international conference on machine learning and soft computing* (pp. 5-9)

[12] Ak, M. F. (2020, June). A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications. In *Healthcare* (Vol. 8, No. 2, p. 111). Multidisciplinary Digital Publishing Institute

[13] Mohammed, M. A., Abd Ghani, M. K., Arunkumar, N. A., Hamed, R. I., Mostafa, S. A., Abdullah, M. K., & Burhanuddin, M. A. (2020). Decision support system for nasopharyngeal carcinoma discrimination from endoscopic images using artificial neural network. The Journal of Supercomputing, 76(2), 1086-1104

[14] Mostafa, S. A., Mustapha, A., Mohammed, M. A., Hamed, R. I., Arunkumar, N., Abd Ghani, M. K., ... & Khaleefah, S. H. (2019). Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson’s disease. *Cognitive Systems Research*, *54*, 90-99