International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 4 Issue 10, October - 2020, Pages: 23-30

Breast Cancer Diagnosis and Survival Prediction Using JNN

Mohammed Ziyad Abu Shawarib, Ahmed Essam Abdel Latif, Bashir Essam El-Din Al-Zatmah, Samy S. Abu-Naser

Department of Information Technology,
Faculty of Engineering and Information Technology,
Al-Azhar University, Gaza, Palestine

***Abstract**: Breast cancer is reported to be the most common cancer type among women worldwide and it is the second highest women fatality rate amongst all cancer types. Notwithstanding all the progresses made in prevention and early intervention, early prognosis and survival prediction rates are still not sufficient. In this paper, we propose an ANN model which outperforms all the previous supervised learning methods by reaching 99.57 in terms of accuracy in Wisconsin Breast Cancer dataset. Experimental results on Haberman's Breast Cancer Survival dataset show the superiority of proposed method by reaching 88.24 % in terms of accuracy. The results are the best reported ones obtained from Artificial Neural Network using JNN environment without any preprocessing of the dataset.*

Keywords: Prediction, JNN, ANN, Breast Cancer

1. INTRODUCTION

Breast Cancer is the second most dangerous cancer after Lung Cancer which is classified to the number one dangerous cancer. Breast cancer constitutes 12% of new cancer cases approximately out of which close to 25% are women [1]. People visit an oncologist, in case of any sign or symptom of cancer. The oncologist can diagnose and detect breast cancer through Mammograms, Magnetic resonance imaging (MRI) of breast, ultrasound of X-ray of the breast, tissue biopsy etc. [2]. Once breast cancer is confirmed, sentinel node biopsy of the patient is done regularly which helps to detect cancerous cells in lymph nodes. Machine Learning techniques are also used for the classification of benign and malignant tumors. The early detection of Breast Cancer can enhance the prediction and survival rate of the patients [3]. This will help the patients to take necessary treatments at the right time [4]. For benign tumors the patients can avoid unnecessary treatments. Artificial Neural Networks technique when applied in the medical field can help in prediction of various outcomes, cost minimization and upgrade the healthcare value to rescue lives of people. The process of classifying tumors can be done by Artificial Neural Networks technique.

In information technology (IT), an artificial neural network (ANN) is a system of hardware and/or software patterned after the operation of neurons in the human brain. ANNs -- also called, simply, neural networks -- are a variety of deep learning technology, which also falls under the umbrella of artificial intelligence (AI)[5].

Commercial applications of these technologies generally focus on solving complex signal processing or pattern recognition problems. Examples of significant commercial applications since 2000 include handwriting recognition for check processing, speech-to-text transcription, oil-exploration data analysis, weather prediction and facial recognition [6].

An ANN usually involves a large number of processors operating in parallel and arranged in tiers. The first tier receives the raw input information -- analogous to optic nerves in human visual processing. Each successive tier receives the output from the tier preceding it, rather than from the raw input -- in the same way neurons further from the optic nerve receive signals from those closer to it. The last tier produces the output of the system [7-8].

Each processing node has its own small sphere of knowledge, including what it has seen and any rules it was originally programmed with or developed for itself. The tiers are highly interconnected, which means each node in tier *n* will be connected to many nodes in tier *n-1* -- its inputs -- and in tier *n+1*, which provides input data for those nodes. There may be one or multiple nodes in the output layer, from which the answer it produces can be read [9].

Artificial neural networks are notable for being adaptive, which means they modify themselves as they learn from initial training and subsequent runs provide more information about the world. The most basic learning model is centered on weighting the input streams, which is how each node weights the importance of input data from each of its predecessors. Inputs that contribute to getting right answers are weighted higher [10].

Typically, an ANN is initially trained or fed large amounts of data [11]. Training consists of providing input and telling the network what the output should be. For example, to build a network that identifies the faces of actors, the initial training might be a series of pictures, including actors, non-actors, masks, statuary and animal faces. Each input is accompanied by the matching identification, such as actors' names, "not actor" or "not human" information. Providing the answers allows the model to adjust its internal weightings to learn how to do its job better [12].

In defining the rules and making determinations -- that is, the decision of each node on what to send to the next tier based on inputs from the previous tier -- neural networks use several principles. These include gradient-based training, fuzzy logic, genetic algorithms and Bayesian methods. They may be given some basic rules about object relationships in the space being modeled [13].

Further, the assumptions people make when training algorithms causes neural networks to amplify cultural biases. Biased data sets are an ongoing challenge in training systems that find answers on their own by recognizing patterns in data. If the data feeding the algorithm isn't neutral -- and almost no data is -- the machine propagates bias [14].

Neural networks are sometimes described in terms of their depth, including how many layers they have between input and output, or the model's so-called hidden layers [15]. This is why the term neural network is used almost synonymously with deep learning. They can also be described by the number of hidden nodes the model has or in terms of how many inputs and outputs each node has. Variations on the classic neural network design allow various forms of forward and backward propagation of information among tiers [16].

Specific types of artificial neural networks include [17-20]:

- Feed-forward neural networks
- Recurrent neural networks
- Convolutional neural networks
- De-convolutional neural networks
- Modular neural networks

Feed-forward neural networks are one of the simplest variants of neural networks. They pass information in one direction, through various input nodes, until it makes it to the output node. The network may or may not have hidden node layers, making their functioning more interpretable. It is prepared to process large amounts of noise. This type of ANN computational model is used in technologies such as facial recognition and computer vision [21-22].

Recurrent neural networks (RNN) are more complex [23]. They save the output of processing nodes and feed the result back into the model. This is how the model is said to learn to predict the outcome of a layer. Each node in the RNN model acts as a memory cell, continuing the computation and implementation of operations. This neural network starts with the same front propagation as a feed-forward network, but then goes on to remember all processed information in order to reuse it in the future. If the network's prediction is incorrect, then the system self-learns and continues working towards the correct prediction during backpropagation. This type of ANN is frequently used in text-to-speech conversions [24].

Convolutional neural networks (CNN) are one of the most popular models used today. This neural network computational model uses a variation of multilayer perceptrons and contains one or more convolutional layers that can be either entirely connected or pooled[25]. These convolutional layers create feature maps that record a region of image which is ultimately broken into rectangles and sent out for nonlinear processing [48]. The CNN model is particularly popular in the realm of image recognition; it has been used in many of the most advanced applications of AI, including facial recognition, text digitization and natural language processing. Other uses include paraphrase detection, signal processing and image classification [25].

De-convolutional neural networks utilize a reversed CNN model process. They aim to find lost features or signals that may have originally been considered unimportant to the CNN system's task. This network model can be used in image synthesis and analysis [25].

Modular neural networks contain multiple neural networks working separately from one another[12-13]. The networks do not communicate or interfere with each other's activities during the computation process. Consequently, complex or big computational processes can be performed more efficiently [14].

Advantages of artificial neural networks include [15]:

- Parallel processing abilities mean the network can perform more than one job at a time.
- Information is stored on an entire network, not just a database.
- The ability to learn and model nonlinear, complex relationships helps model the real life relationships between input and output.
- Fault tolerance means the corruption of one or more cells of the ANN will not stop the generation of output.
- Gradual corruption means the network will slowly degrade over time, instead of a problem destroying the network instantly.
- The ability to produce output with incomplete knowledge with the loss of performance being based on how important the missing information is.
- No restrictions are placed on the input variables, such as how they should be distributed.

- Machine learning means the ANN can learn from events and make decisions based on the observations.
- The ability to learn hidden relationships in the data without commanding any fixed relationship means an ANN can better model highly volatile data and non-constant variance.
- The ability to generalize and infer unseen relationships on unseen data means ANNs can predict the output of unseen data.

2. LITERATURE REVIEW

Many of breast cancer research have been reported in the literature of medical data analysis, and most of them turn up with good classification accuracies. They used Wisconsin Breast Cancer (original). Author in [12] proposed LS-SVM classifier algorithm for the diagnosis of breast cancer and achieved the classification accuracy of 98.53% using 10-fold cross validation. Author in [13] proposed a new method for the breast cancer diagnosis using support vector classification algorithm on the most predictive features and obtain the classification accuracies of 99.02% without cross-validation. Author in [14] presented an innovative technique for breast cancer detection, by using statistical methods in combination with swarm optimization and reported the accuracy of 98.71%. Authors in [15] proposed a new method AMMLP for the classification of breast cancer datasets by using an Artificial Neural Network over the biological met plasticity property and acquire classification accuracy of 99.26. Authors in [16] discussed the use of available technological advancements to develop prediction models for breast cancer. The manuscript used Naïve Bayes, RBF Network and J48 to develop prediction model by 10 fold cross-validation method for measuring the unbiased estimate of these models for performance comparison. Author in [17] used five classification algorithms Naïve bayes, SMO, REP Tree, J48 and MLP upon two data sets which are breast cancer and diabetes respectively, from the UCI ma-chine learning repository. Author in [18] used C4.5 decision tree for classification with a success of 94.74% was achieved. Author in [19] used fuzzy genetic algorithm with a success of 97.36% was obtained. Author in [20] achieved a 95% success using blurred neurons. Author in [21] used feedforward neural networks with a success of 98.1%. Author in [22] used perceptron neural network method and a success rate of 98.8% was obtained. Author in [23] used fuzzy clustering method with a success of 95.57% was achieved. Author in [24] used generalized regression neural networks and achieved a success of 98.8%. Authors in [25] used multilayer perceptron neural network, combined neural network (CNN), probabilistic neural network, recurrent neural network and support vector machine. The highest success was obtained using support vector machines with 99.54%. Authors in [22] achieved a success rate of 99.50% using the filter and wrapper methods. Author in [23] performed using support vector machines, a diagnostic success of 97.71% were obtained. Authors in [24] used an evolutionary artificial neural network approach based on the pareto-differential evolution algorithm augmented with local search for the prediction of breast cancer. The approach is named Memetic Pareto Artificial Neural Network. Since the early dates of the researches in the field of computational biomedicine, the cancer survivability prediction has been a challenging problem for many researchers [18]. In [19], artificial neural networks and decision trees along with logistic regression were used to develop the prediction models using 202,932 breast cancer patients records, which then pre-classified into two groups of “survived” (93,273) and “not survived” (109,659). The results of predicting the survivability were in the range of 93 % accuracy. Authors in[20] used a fully complex valued fast learning artificial neural network with GD activation function in the hidden layer. The comparison results showed that, FC-FLC provides a better classification performance comparing to the SRAN, MCFIS and ELM classifiers. Authors in [21] used the under-sampling C5 technique and bagging algorithm to deal with the imbalanced problem predictive models for breast cancer survivability.

3. METHODOLOGY

After getting the Breast Cancer Wisconsin dataset (Orgional) and figured out the best accuracy rate to be 99.57%[ ], We collected the Haberman's Survival Dataset from “the Center for Machine Learning and Intelligent Systems, University of California, Irvine”[26], we identified the input variables, output variables, upload the dataset, divided it to training and validating sets, determined the proper hidden layers. Then we trained and validated the sets to get the best accuracy.

3.1 The Input Variables

The input variables selected are those which can easily be obtained from Haberman's Survival Dataset. The input variables are: Sample code number, Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses, and Class. These factors were transformed into a format suitable for neural network analysis. The domain of the input variables used in this study shown in Table1.

Table 1: Input and out attributes

S.N.	Attribute name	Domain of Attribute	Attribute type
1	Age of patient at time of operation	numerical	Input
2	Patient's year of operation	year - 1900, numerical	Input

3	Number of positive axillary nodes detected	numerical	Input
4	Survival status (class attribute)	1 = the patient survived 5 years or longer 2 = the patient died within 5 year	Output

3.2 Output variable

The output variable is the class of whether the patient survived 5 years or longer or died within 5 year.

4.3 Neural Network

The neural network topology was built based on the Multilayer Perceptron with one input layer, two hidden layer (2x1 nodes) and one output layer as shown in Figure 1.

3.4 Evaluation of the study

First of all, for the evaluation of our study, we used a sample of 306 of Haberman's Survival Dataset representing whether the patient survived 5 years or longer or died within 5 year. We developed a model able to differentiate between whether the patient survived 5 years or longer or died within 5 year. Our model uses a neural network with one input layer, two hidden layers and one output layer. As input data for predicting the Validity of Haberman's Survival Dataset, we used attribute as shown in Figure 2.

Our task was to predict the result based on the 3 input variables and one output. We conducted a series of tests in order to establish the number of hidden layers and the number of neurons in each hidden layer. Our tests give us that the best results are obtained with two hidden layers. We used a sample of (306 records): 204 training samples and 102 validating samples. The network structure was found on a trial and error basis (as seen in Figure 1). We started with a small network and gradually increased its size. Finally, we found that the best results are obtained for a network with the following structure: 3I-2H-1O, i.e. 3 input neurons, 2 hidden layers with (2x1) neurons, and an output layer with 1 neuron. For this study we used Just Neural Network (JNN)[78]. We trained the network for 31452 epochs (as shown in Figure 3) on a regular computer with 4 GB of RAM memory under the Windows 10 operating system. We got an accuracy of 88.24%. Figure 4 shows Parameters of the proposed ANN model. Figure 5 shows the factors, their importance and relative importance that affect the breast cancer artificial Neural Model using Just NN environment. Figure 6 outlines the detail of the proposed ANN model.

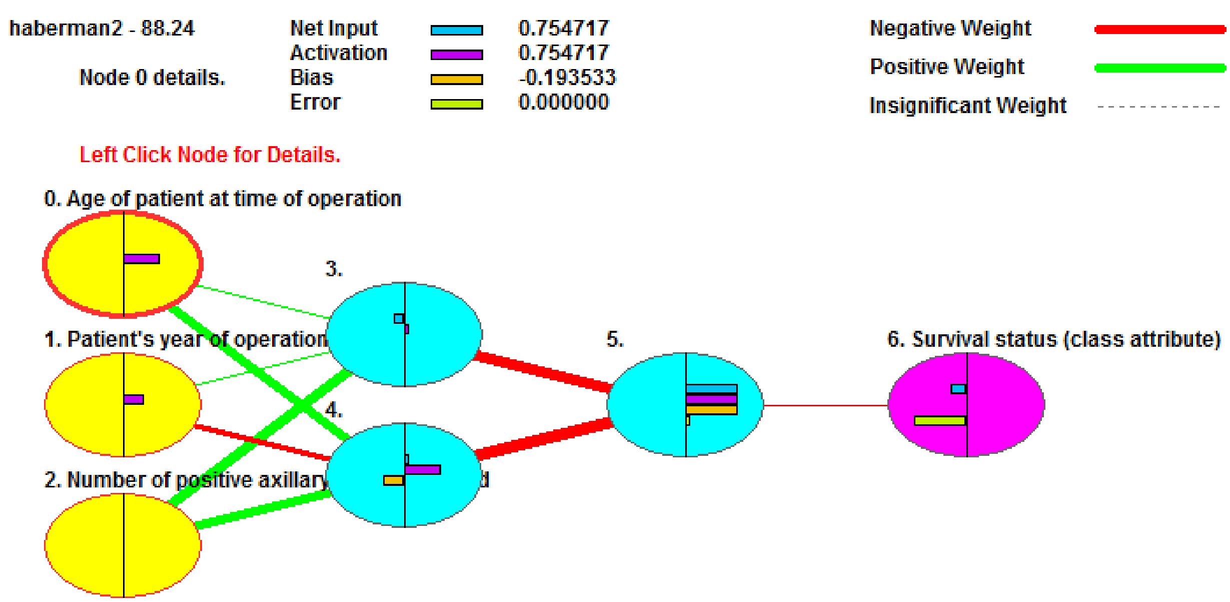
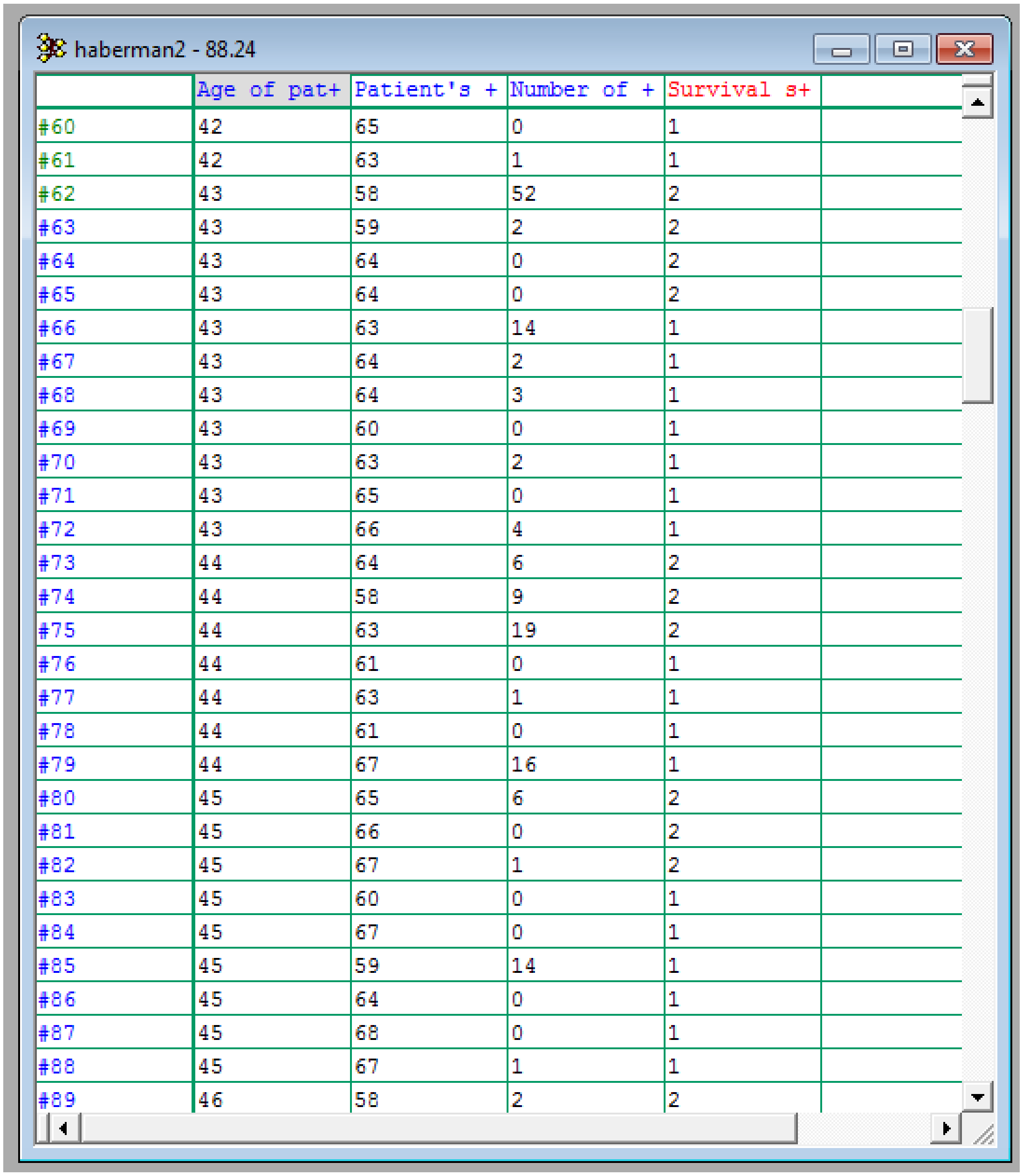


Figure 1: Structure of the proposed ANN model



	Age of pat+	Patient's +	Number of +	Survival s+
#60	42	65	0	1
#61	42	63	1	1
#62	43	58	52	2
#63	43	59	2	2
#64	43	64	0	2
#65	43	64	0	2
#66	43	63	14	1
#67	43	64	2	1
#68	43	64	3	1
#69	43	60	0	1
#70	43	63	2	1
#71	43	65	0	1
#72	43	66	4	1
#73	44	64	6	2
#74	44	58	9	2
#75	44	63	19	2
#76	44	61	0	1
#77	44	63	1	1
#78	44	61	0	1
#79	44	67	16	1
#80	45	65	6	2
#81	45	66	0	2
#82	45	67	1	2
#83	45	60	0	1
#84	45	67	0	1
#85	45	59	14	1
#86	45	64	0	1
#87	45	68	0	1
#88	45	67	1	1
#89	46	58	2	2

Figure 2: Imported dataset in JNN environment

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 4 Issue 10, October - 2020, Pages: 23-30

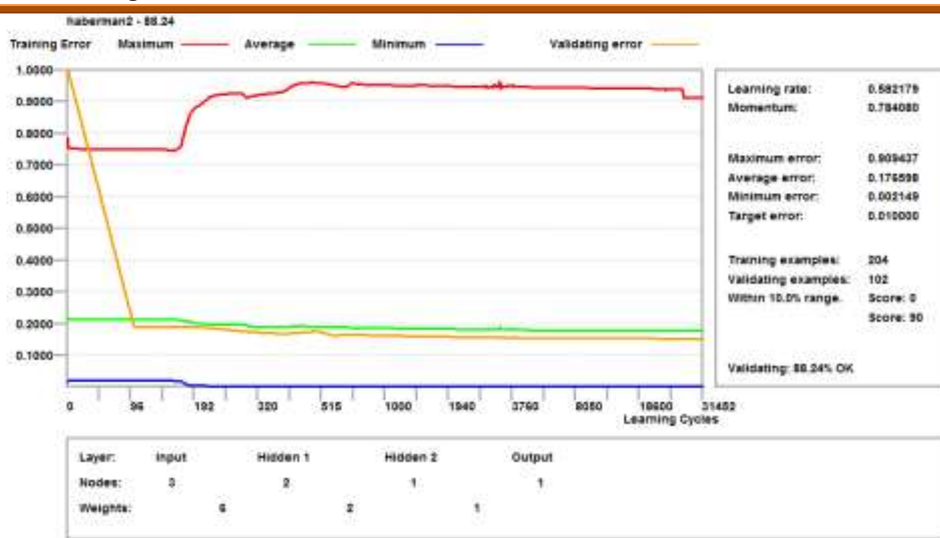


Figure 3: Training and validating the ANN model

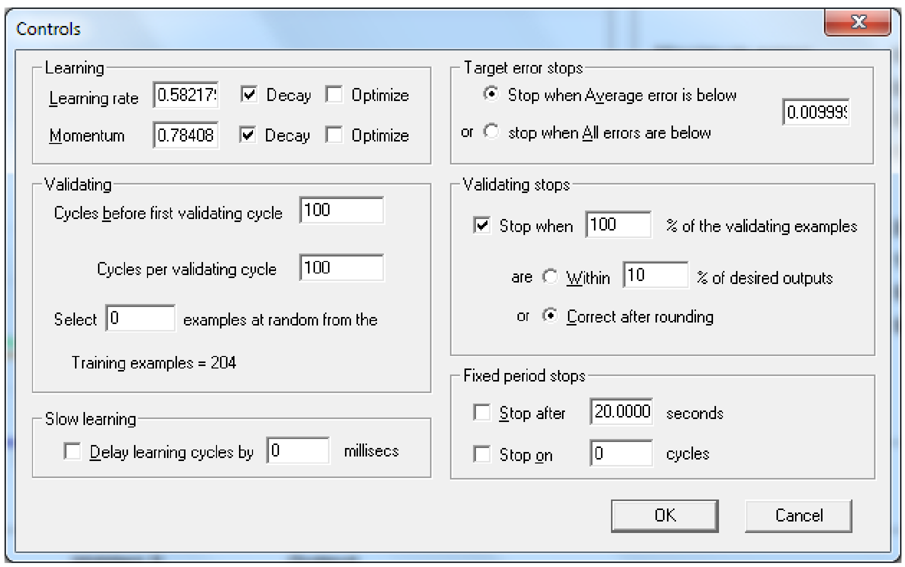


Figure 4: Parameters of the proposed ANN model

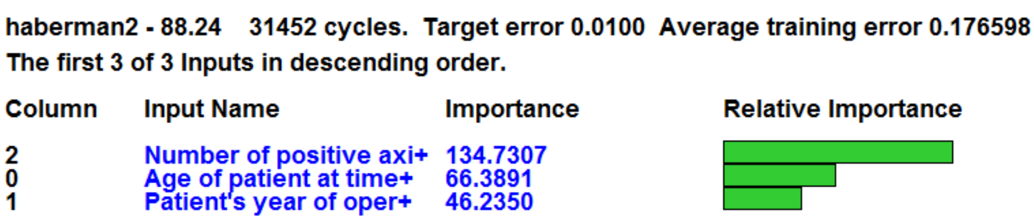


Figure 5: Most influential features in the dataset

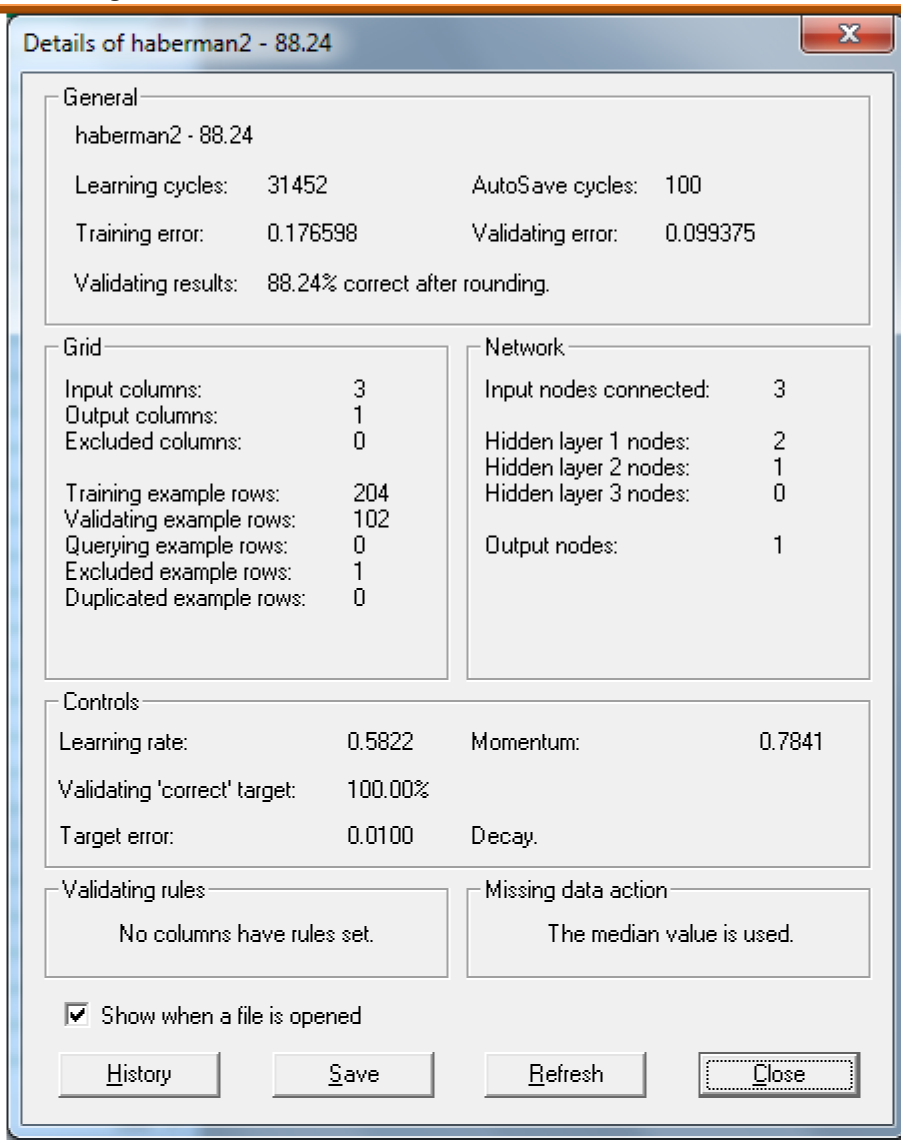


Figure 6: Details of the proposed ANN model

4. CONCLUSION

In This paper, we used the prediction power of a neural network to classify whether the patient survived 5 years or longer or died within 5 year. Our network achieved an accuracy of 88.24%. We used the JustNN environment for building the network that was a feed forward Multi-Layer Perceptron with one input layer, two hidden layers and one output layer. We have collected the Haberman's Survival Dataset from "the Center for Machine Learning and Intelligent Systems, University of California, Irvine".

References

1. Jemal A, et al.(2005).. Cancer statistics, 2005. CA: a cancer journal for clinicians. 2005 Jan 1;55(1):10-30.
2. Polat K, Güneş S. Breast cancer diagnosis using least square support vector machine. Digital Signal Processing. 2007 Jul 1;17(4):694- 701.
3. Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications. 2009 Mar 1;36(2):3240-7.
4. Yeh W. et al.. (2009). A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method. Expert Systems with Applications. 36(4):8204-11.
5. Polat K, Güneş S. Breast cancer diagnosis using least square support vector machine. Digital Signal Processing. 2007 Jul 1;17(4):694- 701.
6. Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications. 2009 Mar 1;36(2):3240-7.
7. Yeh WC, Chang WW, Chung YY. A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method. Expert Systems with Applications. 2009 May 1;36(4):8204-11.
8. Marcano-Cedeño A, Quintanilla-Domínguez J, Andina D. WBCD breast cancer database classification applying artificial metaplasticity neural network. Expert Systems with Applications. 2011 Aug 1;38(8):9573-9.
9. Chaurasia V., Pal., S, Tiwari., BB.: Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms & Computational Technology, Vol. 12(2), pp. 119–126. DOI: http://dx.doi.org/10.1177/1748301818756225. (2018).
10. Verma, D., Mishra., N.:Analysis and Prediction of Breast cancer and Diabetes disease datasets using Data mining classification Techniques. In Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS), pp 533-538, (2017).
11. Quinlan, "Improved use of continuous attributes in C4. 5," Journal of artificial intelligence research, vol. 4, pp. 77-90, 1996.
12. Pena-Reyes and M. Sipper, "A fuzzy-genetic approach to breast cancer diagnosis," Artificial intelligence in medicine, vol. 17, pp. 131-155, 1999.
13. Nauck and R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data," Artificial intelligence in medicine, vol. 16, pp. 149-169, 1999.
14. Setiono, "Generating concise and accurate classification rules for breast cancer diagnosis," Artificial Intelligence in medicine, vol. 18, pp. 205-219, 2000.
15. Albrecht, G. Lappas, S. A. Vinterbo, C. Wong, and L. Ohno-Machado, "Two applications of the LSA machine," in Neural Information Processing, 2002. ICONIP'02. Proceedings of the 9th International Conference on, 2002, pp. 184-189.
16. Abonyi and F. Szeifert, "Supervised fuzzy clustering for the identification of fuzzy classifiers," Pattern Recognition Letters, vol. 24, pp. 2195-2207, 2003.
17. Kiyan and T. Yildirim, "Breast cancer diagnosis using statistical neural networks," IU-Journal of Electrical & Electronics Engineering, vol. 4, pp. 1149-1153, 2004.
18. Übeyli, "Implementing automated diagnostic systems for breast cancer detection," Expert systems with Applications, vol. 33, pp. 1054-1062, 2007.
19. Peng, Z. Wu, and J. Jiang, "A novel feature selection approach for biomedical data classification," Journal of Biomedical Informatics, vol. 43, pp. 15-23, 2010.
20. Salama, M. Abdelhalim, and M. A.-e. Zeid, "Breast cancer diagnosis on three different datasets using multi-classifiers," Breast Cancer (WDBC), vol. 32, p. 2, 2012.
21. Dursun D, Glenn W and Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artif Intell Med 2004; 34: 113–127.
22. Dursun, W. Glenn, K. Amit, 1 June 2005, Predicting breast cancer survivability: a comparison of three data mining methods Artificial intelligence in medicine (volume 34 issue 2 Pages 113-127.
23. Hussein A. Abbass, An Evolutionary Artificial Neural Networks Approach for Breast Cancer Diagnosis, School of Computer Science, University of New South Wales, Australian Defence Force Academy Campus.
24. Liu Y-Q, Wang C and Zhang L. Decision tree based predictive models for breast cancer survivability on imbalanced data. In: 3rd international conference on bioinformatics and biomedical engineering, 11-13 June 2009, Beijing, China, 2009.
25. Sivachitra, M., and S. Vijayachitra. "Classification of post-operative breast cancer patient information using complex valued neural classifiers." In Cognitive Computing and Information Processing (CCIP), 2015 International Conference on, pp. 1-4. IEEE, 2015.
26. UCI Machine Learning repository (https://archive.ics.uci.edu/ml/datasets.html)
27. EasyNN Tool.