International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 11, November - 2020, Pages: 8-15

Tumor Classification Using Artificial Neural Networks

Jamal Khamis El-Mahelawi, Jinan Usama Abu-Daqah, Rasha Ibrahim Abu-Latifa, Bassem S. Abu-Nasser, Samy S. Abu-Naser

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

Abstract: *Tumor is a group of diseases that involve abnormal increases in the number of cells, with the potential to invade or spread to other parts of the body. Not all tumors or lumps are cancerous; benign tumors are not classified as being cancer because they do not spread to other parts of the body. There are over 100 different known Tumors that affect humans. Tumors are often described by the body part that they originated in. However, some body parts contain multiple types of tissue, so for greater precision, tumors are additionally classified by the type of cell that the tumor cells originated from. The aim of this study is to propose an Artificial Neural Network model for the classification of tumor types. Some of important features in the classification of the tumors are age, sex, histologic-type, degree-of-diffe, status of bone, bone-marrow, lung, pleura, peritoneum, liver, brain, skin, neck, supraclavicular, axillar, mediastinum, and abdominal. They were used as input variables for the ANN model. A model based on the Multilayer Perceptron topology was created and trained using "primary tumor" dataset which was collected from the University Medical Centre, Institute of Oncology, Ljubljana, Evaluation of the ANN model showed that the ANN model is able to correctly classify the tumor type with 79.65 % accuracy rate.*

Keywords: Artificial Neural Networks, Tumor type, Cancer, JNN, Medicine, Classification.

1. Introduction

The main objective of this study is to determine tumor category for patients based on attributes which are set of tests for the patient body. Specifically the study seeks to explore the possibility of using an Artificial Neural Network model to predict the category of a tumor. The category of a tumor may be certain type of function with a number of factors. However, it seems that it will be difficult to find a mathematical model that effectively models these factors relationship. A useful approach to deal with this type of problem is to apply common regression analysis in which historical data are the best fitted to some function. The result is an equation in which each of the inputs xj is multiplied by a weight wj; the sum of all such products and a constant θ, gives an output $y = \Sigma\ wj\ xj + \theta$, where j=0..n[1].

Such studies face problems with the complexity of selecting an appropriate function fit to capture all forms of data associations as well as automatically adjusts output in case of additional information, because of the performance of a candidate is controlled by a number of factors, and this control is not going to be any straightforward well-known regression model[2].

Artificial neural network emulates humans' brain in solving problems; it is a common approach that can tackle that kind of problems. Therefore, the attempt to build an adaptive system such as Artificial Neural Network to predict a tumor's category based on the consequence of such factors[3].

The objectives of this study are:

- To identify some suitable factors that affects tumor classification,
- To convert these factors into forms appropriate for an adaptive system coding, and
- To model an Artificial Neural Network that can be used to predict the tumor category based on some predetermined data for a given patient.

2. Artificial neural networks

An Artificial Neural Network (ANN) is a branch of Artificial Intelligence [4]. It is a mathematical model that is encouraged by the organization and/or functional feature of biological neural networks. A neural network has a connected set of artificial neurons, and it processes information using a connectionist form to computation. Generally, an ANN is an adaptive system that fine-tunes its organization based on external or internal information that runs through the network during the learning process.

Latest neural networks are non-linear numerical data modeling tools. They usually used to model sophisticated relationships among inputs and outputs or to uncover patterns in data. ANN has been applied in various applications with considerable fulfillment [5]. For example, ANN has been applied effectively in the area of prediction, handwritten character recognition, evaluating prices of housing [6].

Neurons often grouped into layers. Layers are groups of neurons that implement similar tasks. There are three types of layers. The input layer is the layer of neurons that receive input from the user program. The output layer is the layer of neurons that send data to the user program. And Between of them there are hidden layers. The Hidden layer neurons are connected only to other neurons and never directly interact with the user program. Every neuron in a neural network has the opportunity to affect processing which can occur at any layer in the neural network. In neural networks, the hidden layers are optional. The input and output layers are essential, however it is possible to have on layer that act as an input and output layer [7].

ANN learning can be directed or undirected. Directed training means giving the neural network a set of sample data alongside the predicted outputs from each of these samples. Directed training is the most common form of neural network training. As directed training continues, the neural network goes through several iterations, or epochs, until the actual output of the neural network equals the predicted output, with a reasonably small error rate. Each iteration is one pass through the training samples. Undirected training is similar to the directed one but no predicted outputs are provided. Undirected training usually occurs when the neural network tends to classify the inputs into several groups. The training progresses through many epochs, just as in directed training. As training progresses, the neural network discovers the classification groups [8]. Training is the process by which these connection weights are assigned. Most training algorithms begin by assigning random numbers to the weight matrix. Then the validity of the neural network is inspected. Next, the weights are tuned based on how valid the neural network done. This process is repetitive until the validation error is within an acceptable limit [8].

Validation of the system is done once a neural network has been trained and it must be assessed to tell if it is ready for actual use. This final step is important so that it can be determined if additional training is required. To correctly validate a neural network, validation data records must be completely separated from the training data records [9]. About 80% of the total sample data was used for network training in this paper. About 20% of the total sample data used for validation of the system.

3. Literature Review

Artificial Neural Networks have been used many fields. In Education such as: Predicting Student Performance in the Faculty of Engineering and Information Technology using ANN, Prediction of the Academic Warning of Students in the Faculty of Engineering and Information Technology in Al-Azhar University-Gaza using ANN, Arabic Text Summarization Using AraBERT Model Using Extractive Text Summarization Approach[6].

In the field of Health such as: Parkinson's Disease Prediction, Classification Prediction of SBRCTs Cancers Using ANN, Predicting Medical Expenses Using ANN, Predicting Antibiotic Susceptibility Using Artificial Neural Network, Predicting Liver Patients using Artificial Neural Network, Blood Donation Prediction using Artificial Neural Network, Predicting DNA Lung Cancer using Artificial Neural Network, Diagnosis of Hepatitis Virus Using Artificial Neural Network, COVID-19 Detection using Artificial Intelligence[7].

In the field of Agriculture: Plant Seedlings Classification Using Deep Learning, Prediction of Whether Mushroom is Edible or Poisonous Using Back-propagation Neural Network, Analyzing Types of Cherry Using Deep Learning, Banana Classification Using Deep Learning, Mango Classification Using Deep Learning, Type of Grapefruit Classification Using Deep Learning, Grape Type Classification Using Deep Learning, Classifying Nuts Types Using Convolutional Neural Network, Potato Classification Using Deep Learning, Age and Gender Prediction and Validation Through Single User Images Using CNN[8].

In other fields such as : Predicting Software Analysis Process Risks Using Linear Stepwise Discriminant Analysis: Statistical Methods, Predicting Overall Car Performance Using Artificial Neural Network, Glass Classification Using Artificial Neural Network, Tic-Tac-Toe Learning Using Artificial Neural Networks, Energy Efficiency Predicting using Artificial Neural Network, Predicting Titanic Survivors using Artificial Neural Network, Classification of Software Risks with Discriminant Analysis Techniques in Software planning Development Process, Handwritten Signature Verification using Deep Learning, Email Classification Using Artificial Neural Network, Predicting Temperature and Humidity in the Surrounding Environment Using Artificial Neural Network, English Alphabet Prediction Using Artificial Neural Networks[9-13].

4. Methodology

A data set refer to Igor Kononenko, and Bojan Cestnik [19] was used, it contains a number of factors that are considered to have an effect on the classification of a tumor. These factors were carefully studied and synchronized into a convenient number appropriate for computer coding within the environment of the ANN modeling. These factors were classified as input variables. The output variables represent the predicted tumor classification based on those inputs.

4.1 The Input Variables

Table 1: Input Data Transformation

#	Input	Domain	Transformed domain
1.	age	<30, 30-59, >=60	1, 2, 3
2.	sex	male, female	0,1
3.	histologic-type	epidermoid, adefalse, anaplastic	1,2,3
4.	degree-of- DIFFE	well, fairly, poorly	1,2,3
5.	bone	yes, no	1,0
6.	bone-marrow	yes, no	1,0
7.	lung	yes, no	1,0
8.	pleura	yes, no	1,0
9.	peritoneum	yes, no	1,0
10.	liver	yes, no	1,0
11.	brain	yes, no	1,0
12.	skin	yes, no	1,0
13.	neck	yes, no	1,0
14.	supraclavicular	yes, no	1,0
15.	axilla	yes, no	1,0
16.	mediastinum	yes, no	1,0
17.	abdominal	yes, no	1,0

These factors were converted into a format suitable for neural network analysis as shown in Table1.

4.2 The Output Variable

The output variable is the Tumor Class, and its domain is: Lung, Head & neck, Esophagus, Thyroid, Stomach, Duodenum & sm.int, Colon, Rectum, Anus, Salivary glands, Pancreas, Gallbladder, Liver, Kidney, Bladder, Testis, Prostate, Ovary, Corpus uteri, Cervix uteri, Vagina, Breast

4.3 Network Architecture

Humans and other animals process information with neural networks. These are formed from trillions of neurons (nerve cells) exchanging brief electrical pulses called action potentials. Computer algorithms that mimic these biological structures are formally called artificial neural networks to distinguish them from the squishy things inside of animals. However, most scientists and engineers are not this formal and use the term neural network to include both biological and nonbiological systems [14-16].

Neural network research is motivated by two desires: to obtain a better understanding of the human brain, and to develop computers that can deal with abstract and poorly defined problems. For example, conventional computers have trouble understanding speech and recognizing people's faces. In comparison, humans do extremely well at these tasks [17]. The network is a multilayer perceptron neural network using the linear sigmoid activation function.
The ANN model consists of three layers: 1 input layer, 2 hidden layers, and 1 output layer (as seen in Figure. 1).

4.4 The Back-propagation Algorithm

Algorithm 1 The basic backpropagation algorithm [18]

1: Initialize weights randomly

2: Initialize err, $threshold$, and $maxEpochs$ 3: **while** $epoch < maxEpoch$ **and** $err > threshold$ **do**
4: **for** each example (x,y) in the training set **do**

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 11, November - 2020, Pages: 8-15

```
/* Propagate the inputs forward to compute the outputs */
for each node i in the input layer do
    a_i ← x_i
end for
for £ = 2 to L do
    for each node j in layer £ do
        in_j ← Σ_i w_i,j a_i
        a_j ← g(in_j)
    end for
end for
/* Propagate deltas backward from output layer to input layer */
for each node j in the output layer do
    Δ[j] ← g'(in_j) × (y_j − a_j)
end for
for £ = L − 1 to 1 do
    for each node i in layer £ do
        Δ[i] ← g'(in_i) Σ_j w_i,j Δ[j]
    end for
end for
/* Update each weight using deltas */
for each weight w_i,j do
    w_i,j ← w_i,j + α × a_i × Δ[j]
end for
end while
```

4.5 Training and Validation of the model

The collected dataset was preprocessed and then imported into JNN environment as shown in Figure2. After setting the control parameter of the proposed ANN model, as can be seen in Figure 3, we have split the data set into 2 groups (training and validation) randomly using the JNN environment. The ration of split is 2/3 of the dataset training and 1/3 is for validation.

Then, we trained the ANN model for 27680 cycles. The Validation accuracy rate was 79.65 (as in Figure 4). We have determined the most influential feature of the dataset as in Figure 5. The details of the proposed ANN model as seen in Figure 6.

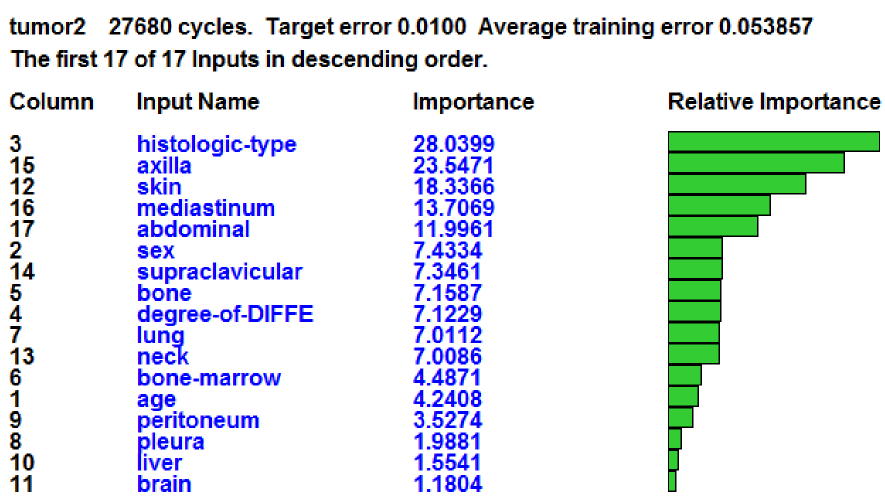

Figure 5: Shows the relative importance of the input attributes.

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 11, November - 2020, Pages: 8-15

tumor2

	Tumor Class	age	sex	histologic+	degree-of-+	bone	bone-marrow	lung	pleura	peritoneum	liver	brain	skin
#0	0	1.0000	1.0000	1.0000	3.0000	2.0000	2.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#1	0	1.0000	1.0000	1.0000	3.0000	2.0000	2.0000	2.0000	2.0000	2.0000	1.0000	2.0000	2.0000
#2	0	1.0000	2.0000	2.0000	3.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#3	0	1.0000	2.0000	1.0000	3.0000	1.0000	2.0000	1.0000	1.0000	2.0000	2.0000	2.0000	2.0000
#4	0	1.0000	2.0000	1.0000	3.0000	1.0000	2.0000	1.0000	1.0000	2.0000	2.0000	2.0000	2.0000
#5	0	1.0000	2.0000	1.0000	3.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	1.0000	2.0000
#6	0	2.0000	1.0000	1.0000	1.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	1.0000
#7	0	2.0000	1.0000	1.0000	1.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#8	0	2.0000	1.0000	1.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#9	0	2.0000	1.0000	1.0000	2.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	1.0000
#10	0	2.0000	1.0000	1.0000	3.0000	1.0000	2.0000	2.0000	1.0000	2.0000	2.0000	2.0000	2.0000
#11	0	2.0000	1.0000	1.0000	3.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	1.0000
#12	0	2.0000	1.0000	1.0000	3.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#13	0	2.0000	1.0000	1.0000	3.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#14	0	2.0000	1.0000	1.0000	3.0000	2.0000	2.0000	2.0000	2.0000	2.0000	1.0000	2.0000	2.0000
#15	0	2.0000	1.0000	1.0000	1.0000	2.0000	2.0000	2.0000	1.0000	2.0000	1.0000	2.0000	2.0000
#16	0	2.0000	1.0000	1.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#17	0	2.0000	1.0000	2.0000	2.0000	2.0000	2.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#18	0	2.0000	1.0000	2.0000	3.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	1.0000
#19	0	2.0000	1.0000	2.0000	3.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#20	0	2.0000	1.0000	2.0000	1.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#21	0	2.0000	1.0000	2.0000	1.0000	2.0000	2.0000	1.0000	2.0000	2.0000	1.0000	2.0000	2.0000
#22	0	2.0000	1.0000	2.0000	1.0000	2.0000	2.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#23	0	2.0000	1.0000	2.0000	1.0000	2.0000	2.0000	2.0000	1.0000	1.0000	2.0000	2.0000	2.0000
#24	0	2.0000	1.0000	2.0000	1.0000	2.0000	2.0000	2.0000	1.0000	2.0000	2.0000	2.0000	2.0000
#25	0	2.0000	1.0000	2.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#26	0	2.0000	1.0000	3.0000	3.0000	1.0000	1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
#27	0	2.0000	1.0000	3.0000	3.0000	2.0000	2.0000	1.0000	2.0000	1.0000	2.0000	2.0000	2.0000
#28	0	2.0000	1.0000	3.0000	3.0000	2.0000	2.0000	2.0000	1.0000	2.0000	1.0000	2.0000	2.0000
#29	0	2.0000	1.0000	3.0000	3.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	1.0000	2.0000
#30	0	2.0000	1.0000	1.0000	3.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	2.0000	2.0000
#31	0	2.0000	1.0000	1.0000	3.0000	1.0000	1.0000	2.0000	2.0000	2.0000	2.0000	1.0000	2.0000

Figure 2: Imported dataset into JNN environment

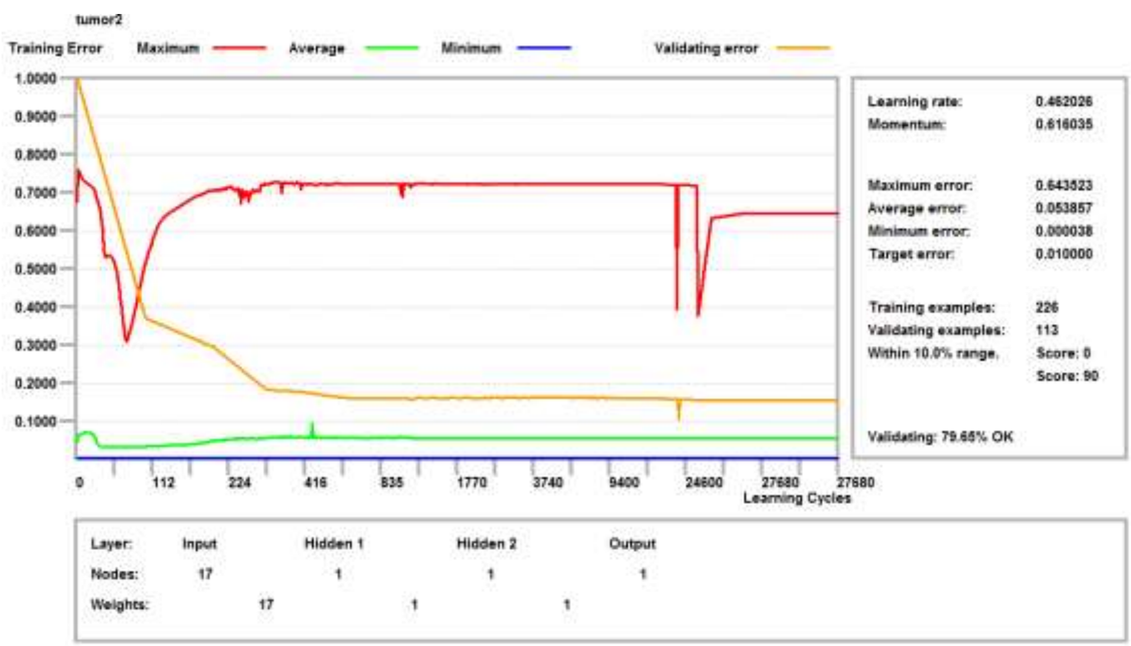

Figure 4: Shows the Training, error, and validation of the data set.

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 11, November - 2020, Pages: 8-15

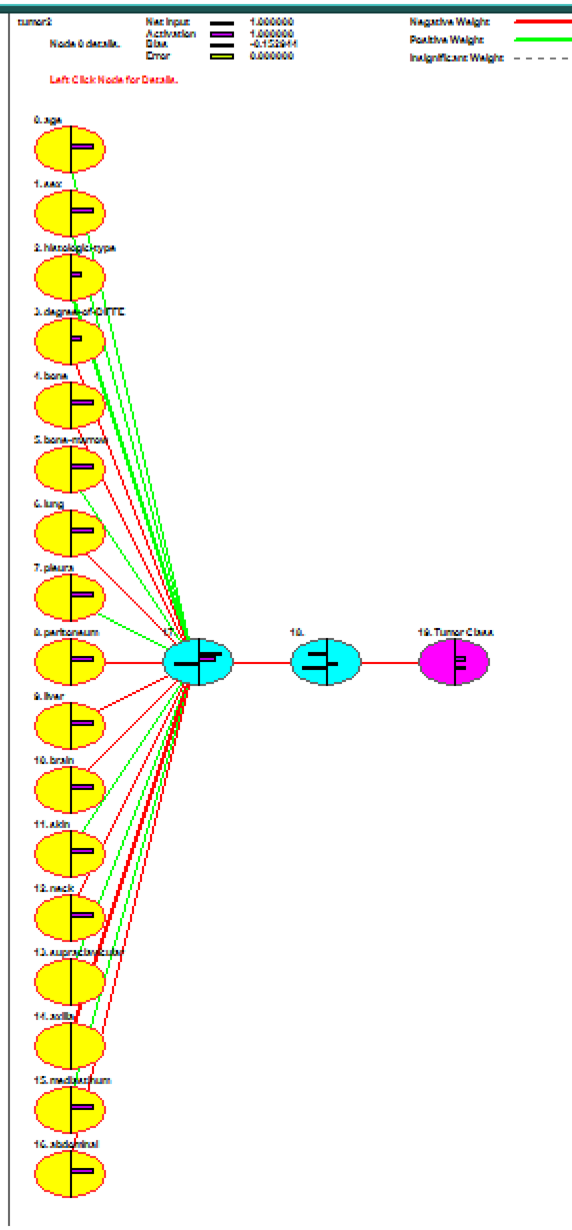

Figure 1: Shows the Design of the Neural Networks

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 11, November - 2020, Pages: 8-15

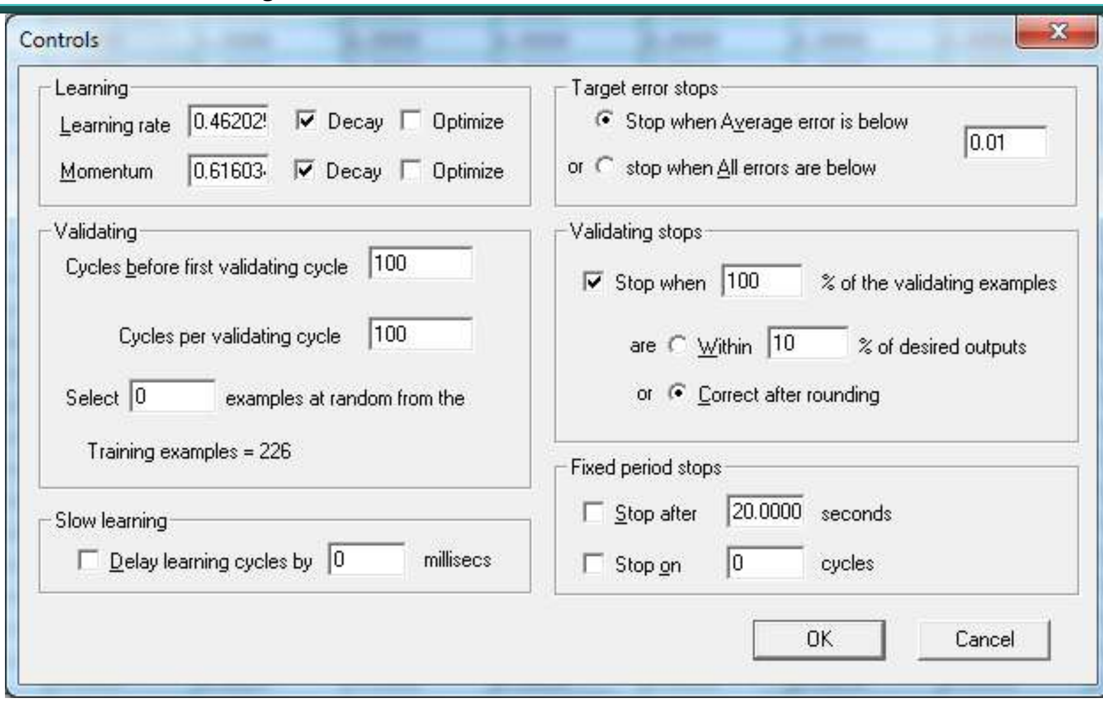

Figure 3: Controls of parameter of the ANN model

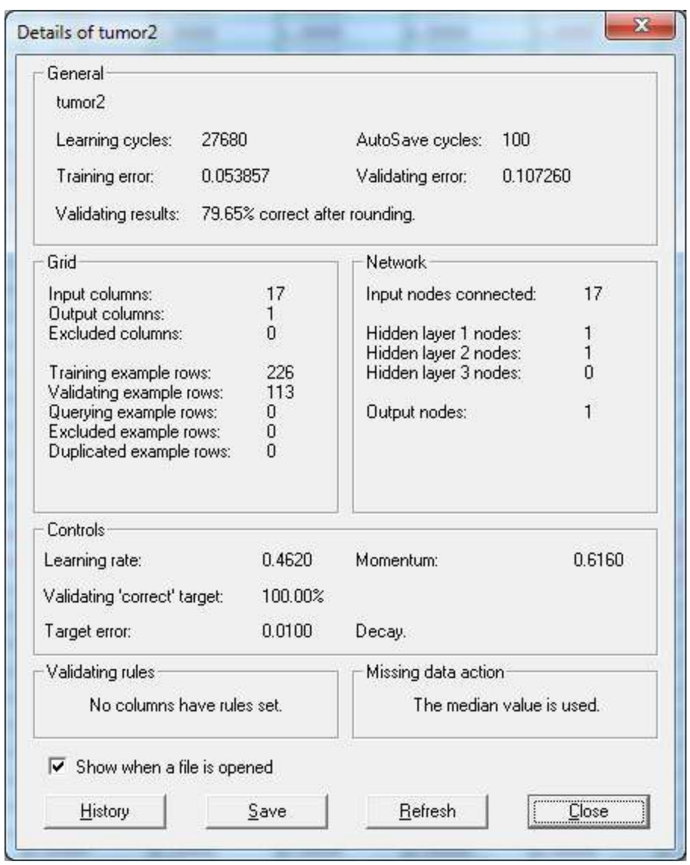

Figure 6: Shows the detail of the ANN model.

5. Evaluation of neural network

The purpose of this study was to classify the tumor type. Where we used patients test results, which provides the possibility to implement and test the neural network and its learning algorithm. Our neural network is designed to classify the tumor based on those test results.

After training and validation, the network was tested using test records and the following results were obtained. This involves inputting variable input data into the grid without output variable results. The output from the grid is then compared with the actual variable data. The neural network was successfully able to accurately classify 79.65 % of the data.

6. Conclusion

An artificial Neural Network model for classifying tumor type was proposed. The model used feed forward backpropagation algorithm for training. The features for the model were collected from dataset represents patients test results. The model was tested and the best result was 79.65%. This study showed that artificial neural network can classify tumor type successfully.

References

1. "Defining Cancer". National Cancer Institute. Retrieved 5 July 2020.
2. Varricchio, Claudette G. (2004). A cancer source book for nurses. Boston: Jones and Bartlett Publishers. p. 229.
3. American Cancer Society. "Melanoma Skin Cancer". American Cancer Society. American Cancer Society. Retrieved 5 July 2020.
4. American Cancer Society. "What is Testicular Cancer". American Cancer Society. American Cancer Society. Retrieved 5 July 2020.
5. A. Lofty and A. Benetton, "Using Probabilistic Neural Networks for Handwritten Digit Recognition", Journal of Artificial Intelligence, vol. 4, no. 4, (2011).
6. P. Khanate and S. Chitins, "Handwritten Devanagari Character Recognition using Artificial Neural Network", Journal of Artificial Intelligence, vol. 4, no. 1, (2011).
7. P. Erika and R. Udegbunam, "Application of neural network in evaluating prices of housing units in Nigeria: A preliminary investigation", J. of Artificial Intelligence, vol. 3, no. 1, (2010).
8. H. Martin and D. Howard, "Neural Network Design", 2nd Edition, Martin Hagan (2014).
9. Askarzadeh, A., and A. Rezazadeh. 2013. "Artificial Neural Network Training Using a New Efficient Optimization Algorithm." Applied Soft Computing, 13(2): 1206– 1213.
10. Bartlett, P.L. 1998. "The Sample Complexity of Pattern Classification with Neural Networks: The Size of the Weights is More Important than the Size of the Network." IEEE Transactions on Information Theory, 44(2): 525– 536.
11. Blackwell, W.J., and F.W. Chen. 2009. Neural Networks in Atmospheric Remote Sensing. Boston: Artech House.
12. Clair, T.A., and J.M. Ehrman. 1998. " Using Neural Networks to Assess the Influence of Changing Seasonal Climates in Modifying Discharge, Dissolved Organic Carbon, and Nitrogen Export in Eastern Canadian Rivers." Water Resource Research, 34(3): 447– 455.
13. Grape, D. 2013. Principles of Artificial Neural Networks. Hackensack, NJ: World Scientific Publishing.
14. Maier, H.R., and G.C. Dandy. 1996. "The Use of Artificial Neural Networks for the Prediction of Water Quality Parameters." Water Resources Research, 32(4): 1013– 1022.
15. Masters, T. 1993. Practical Neural Network Recipes in C++. New York: Academic Press.
16. Nuchitprasittichai, A., and S. Cremaschi. 2013. "An Algorithm to Determine Sample Sizes for Optimization with Artificial Neural Networks." AIChE Journal, 59(3): 805– 812.
17. Knur, A.S., N.H.M. Radial, and A.O. Ibrahim. 2014. "Artificial Neural Network Weight Optimization: A Review." TELKOMNIKA Indonesian Journal of Electrical Engineering, 12(9): 6897– 6902.
18. Pachepsky, Y.A., D. Timlin, and G. Varallyay. 1996. "Artificial Neural Networks to Estimate Soil Water Retention from Easily Measurable Data." Soil Science Society of America Journal, 60(3): 727– 733.
19. UCI Machine Learning repository (https://archive.ics.uci.edu/ml/datasets.html)
20. EasyNN Tool