International Journal of Academic Information Systems Research (IJAISR)
ISSN: 2643-9026
Vol. 4, Issue 9, September – 2020, Pages: 66-73

Evolving Efficient Classification Patterns in Lymphography Using EasyNN

Ahmed Suhail Jaber, Ahmed Khalil Humid, Mohammed Ahmed Hussein, Samy S. Abu-Naser

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

Abstract: *A neural network exploits the non-linearity of a problem to define a set of desired inputs. Neural networks are important in realizing a better way for classification in machine learning and finds application in various fields such as data mining, pattern recognition, forensics etc. In this paper, our focus is to classify of patient records obtained from clinical data. Feature selection is a supervised method that attempts to select a subset of the predictor features based on the information gain. The Lymphography dataset comprises of 18 attributes and 148 instances with the class label having four distinct values. This paper highlights the accuracy of EasyNN backbrapagation calssification algorithm in classifying predictor attributes and highlights its performance on Lymphography dataset. The accuracy we have reached is 97.78 percent in classification accuracy with the predictor feature.*

Keywords: EaysNN. Feature Selection, Classification, Lymphography Data

1. INTRODUCTION

Machine learning is concerned with the design and development of algorithms that enable a system to automatically learn to identify complicated patterns and make intelligent decisions based on the available data. However the enormous size of available data poses a major impediment in recognizing patterns. To handle the vast collection of data, a tool called EasyNN was used to classify a collection of records. Classification is a supervised technique that designates items in a collection to target categories or classes. The main aim of classification is to precisely predict the target class for each unknown case in the data. Multiclass Classification, also called Multinomial classification assigns the given set of input data to one of many categories [1].

A lymph node [2] is an oval-shaped organ of the immune system, distributed widely throughout the body. They tend to expand in size for diverse reasons, indicating health complications that scale from trivial, to life-threatening ailments such as cancers. In the latter, the condition of lymph nodes is so significant that it is used to accurately sense the stage in Cancer progression, which decides the treatment to be adopted. Lymphography [3] is a medical imaging technique in which a radio contrast agent is injected, and then an X-ray picture is taken to visualize structures of the lymphatic system, including lymph nodes, lymph ducts, lymphatic tissues, lymph capillaries and lymph vessels. This data is necessary to decide on whether the clinical details acquired from a Lymphograph pertains to a normal or abnormal finding. Additionally the existing state of the lymph nodes could also suggest the possibility of occurrence of cancer [4]. Though the procedure for performing Lymphography involves potential hurdles, the data from the images facilitate accurate and precise determination of the state of the lymph nodes, ducts and capillaries. Hence proper classification and determination of credential attributes could simplify the process of disease prediction and evoke deterrent measures. Since cancer is a leading cause of death round the globe, crafting an efficient classifier for an oncogenic database has been the rationale for our research.

Our research work mainly focuses on recognizing a suitable classification algorithm for the Lymphography dataset from the UCI Machine Learning repository. We realize this by executing classification algorithms on the dataset for a comparative analysis.

2. 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.

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[5].

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[6].

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.

In the field of Lymphography such as: Authors in [5] made a study of four clustering techniques and reviewed the most representative off-line clustering techniques: K-means clustering, Fuzzy Cmeans clustering, Mountain clustering, and Subtractive clustering. The techniques are implemented and tested against a medical problem of heart disease diagnosis. Performance and accuracy of the four techniques were presented and compared-Means achieved a clustering accuracy of 80% while Fuzzy Cmeans and Mountain clustering achieved an accuracy of 78%. Subtractive clustering offered the least accuracy of 75%.

Authors in [6] presented a cluster analysis method for multidimensional time-series data on clinical laboratory examinations. Their method represented the time series of test results as trajectories in multidimensional space, and compared their structural similarity by using the multi-scale comparison technique. It enabled identification of the part-to-part correspondences between two trajectories, taking into account the relationships between different tests. The resultant dissimilarity could be further used with clustering algorithms for finding the groups of similar cases. The method was applied to the cluster analysis of Albumin-Platelet data in the chronic hepatitis dataset. The results demonstrated that it could form interesting groups of cases that had high correspondence to the fibrotic stages.

Authors in [7] revealed the means of effectively using a number of validation sets obtained from the original training dataset to improve the performance of a classifier. The proposed validation boosting algorithm was illustrated with a support vector machine in Lymphography classification. A number of runs with the algorithm was generated to show its robustness as well as to generate consensus results. At each run, a number of validation datasets were generated by randomly picking a portion of the original training dataset. At each iteration the trained classifier was used to classify the current validation dataset. Experimental results on the Lymphography dataset showed that the proposed method with validation boosting could achieve much better generalization performance with a testing set than the case without validation boosting.

Authors in [8] proposed a novel hybrid classification system based on C4.5 decision tree classifier and one-against-all approach to classify the multi-class problems including dermatology, image segmentation, and Lymphography datasets taken from UCI (University of California Irvine) machine learning database. In their work, initially C4.5 decision tree was executed for all the classes of datasets and they reported 84.48%, 88.79%, and 80.11% classification accuracy for dermatology, image segmentation, and Lymphography datasets using 10-fold cross validation, respectively. The proposed method based on C4.5 decision tree classifier and one-against-all approach obtained 96.71%, 95.18%, and 87.95% for the above datasets, respectively.

3. Methodology

3.1 Lymphography Dataset

This Lymphography dataset was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia [9]. The dataset comprises of a target class that can have four distinct values and the number of predictor attributes sums up to eighteen. This data provides 148 cases to train the classifier. The details of the attributes, their possible values and the associated Attribute ID are clearly listed in Table 1.

Table 1. Description of the Attributes in the Lymphography dataset

Attribute	Possible Values	Assigned values	Attribute ID
Lymphatics	Normal, arched, deformed, displaced	1-4	1
Block of afferent	No, Yes	1,2	2
Block of lymph.c (superior and inferior flaps)	No, Yes	1,2	3
Block of lymph.s (lazy incision)	No, Yes	1,2	4
Bypass	No, Yes	1,2	5
Extravasates (force out of lymph)	No, Yes	1,2	6
Regeneration	No, Yes	1,2	7
Early uptake in	No, Yes	1,2	8
Lymph nodes diminish	0-3	0-3	9
Lymph nodes enlarge	1-4	1-4	10
Changes in lymph	Bean, oval, round	1-3	11
Defect in node	No, lacunar, lacunar marginal, lacunar central	1-4	12
Changes in node	No, lacunar, lacunar marginal, lacunar central	1-4	13
Changes in structure	no, grainy, drop-like, coarse, diluted, reticular, stripped, faint	1-8	14
Special forms	No, Chalices, vesicles	1-3	15
Dislocation	No, Yes	1-2	16
Exclusion of no.	No, Yes	1-2	17
Number .of nodes in	0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, >=70	1-8	18
Target Class	Normal , metastases, malign lymph, fibrosis	1-4	19

3.2 Data Pre-processing

The Lymphography dataset was obtained from the UCI Machine Learning Repository website (UCI, SGI MLC++) [9] and saved as a text file. This file is then imported into Excel spreadsheet and the values are saved with the corresponding attributes as column headers. The Excel file is then uploaded into EasyNN [10], a data classification tool and the uploaded data is visualized to ensure that the precise values are inserted. The predictor and the target attributes are specified. In order to apply classification algorithm, the textual data needs to be stored as a Comma Separated Version (.CSV) file and the attribute selection must be categorical.

3.3 Back propagation Algorithm

Backpropagation, short for "backward propagation of errors," is an algorithm for supervised learning of artificial neural networks using gradient descent. Given an artificial neural network and an error function, the method calculates the gradient of the error function with respect to the neural network's weights. It is a generalization of the delta rule for perceptrons to multilayer feedforward neural networks.

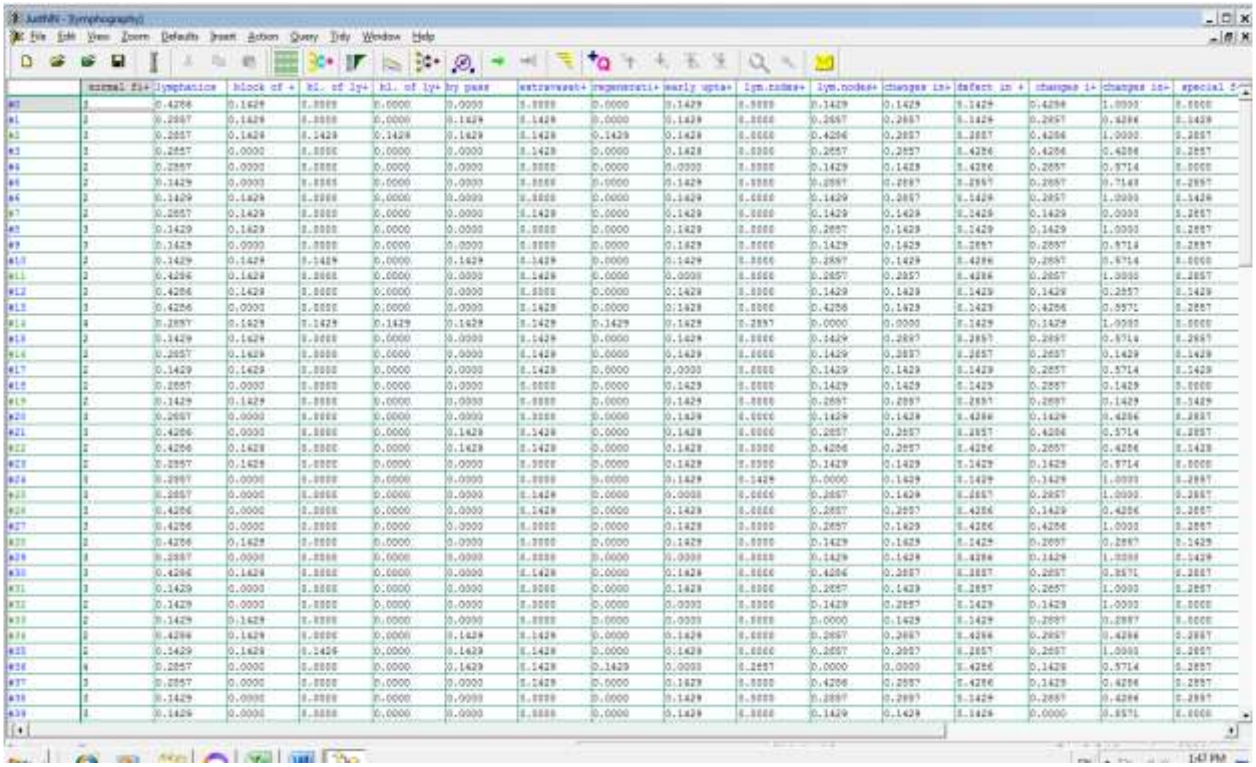
The "backwards" part of the name stems from the fact that calculation of the gradient proceeds backwards through the network, with the gradient of the final layer of weights being calculated first and the gradient of the first layer of weights being calculated last. Partial computations of the gradient from one layer are reused in the computation of the gradient for the previous layer. This backwards flow of the error information allows for efficient computation of the gradient at each layer versus the naive approach of calculating the gradient of each layer separately.

Backpropagation's popularity has experienced a recent resurgence given the widespread adoption of deep neural networks for image recognition and speech recognition. It is considered an efficient algorithm, and modern implementations take advantage of specialized GPUs to further improve performance.

4. PERFORMANCE EVALUATION

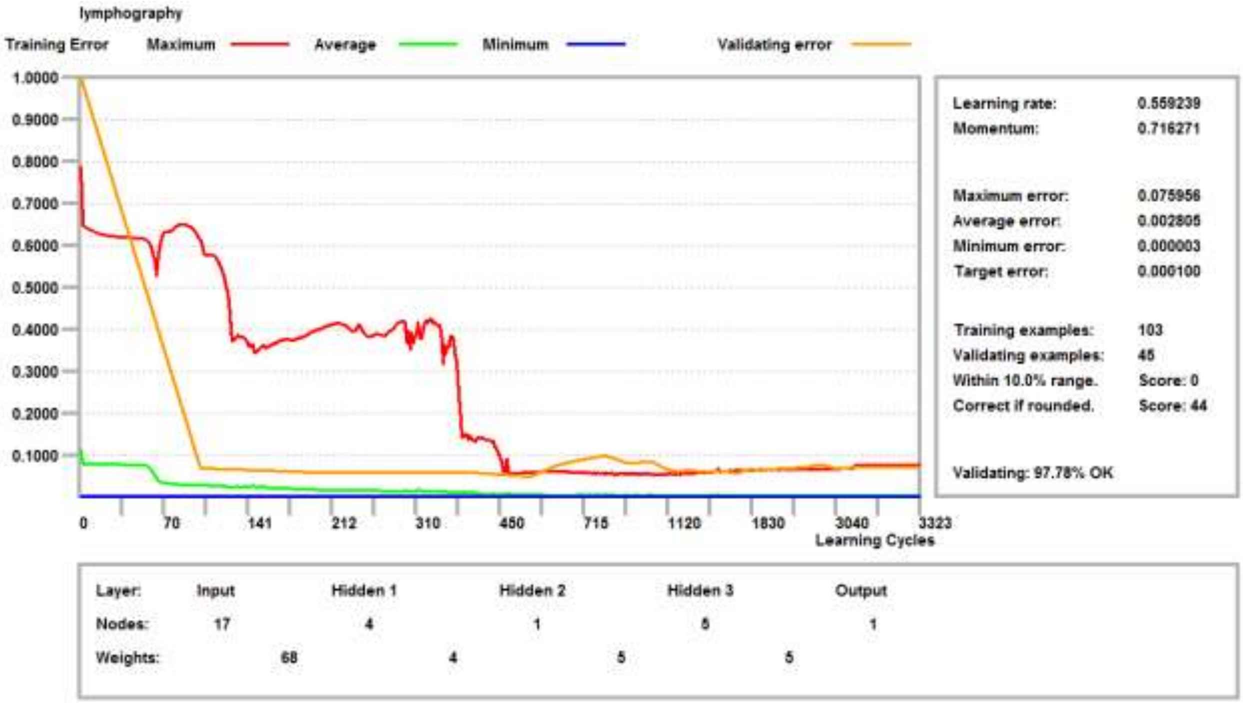
The classification algorithms are ranked based on its testing accuracy and less computational complexity. Accuracy of a classifier is measured in terms of how precisely the classifier places the input datasets under the correct category. This is denoted as the Misclassification rate which is computed as 1- Accuracy(C) where C denotes Classifier.

We have 148 samples in the dataset. We divide it into 103 training sample and 45 validating sample then we imported the dataset in Easy Neural Network (ENN) environment (as shown in Figure 1). We then trained, validated the ENN model (as seen in Figure 2). We found the most important attributes contributing to the ENN model as shown in Figure 3. The detail of ANN model is shown in Figure 4. The Architecture of the ANN model we used consists of 5 layers: one Input layer, three hidden layers, and one output layer as shown in Figure 5. The hidden layers consist of (4 x 1 x 8) nodes. The controls of the parameters of the model are shown in Figure 6. The accuracy of the ANN model was 97.78%.


JustNN - [lymphography]
File Edit View Zoom Defaults Insert Action Query Tidy Window Help
1:47 PM

	normal fl+	lymphatics	block of +	bl. of ly+	bl. of ly+	by pass	extravasat+	regenerati+	early upta+	lym.nodes+	lym.nodes+	changes in+	defect in +	changes i+	changes in+	special f+
#0	3	0.4286	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.1429	0.1429	0.4286	1.0000	0.0000
#1	2	0.2857	0.1429	0.0000	0.0000	0.1429	0.1429	0.0000	0.1429	0.0000	0.2857	0.2857	0.1429	0.2857	0.4286	0.1429
#2	3	0.2857	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.0000	0.4286	0.2857	0.2857	0.4286	1.0000	0.2857
#3	3	0.2857	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.0000	0.2857	0.2857	0.4286	0.4286	0.4286	0.2857
#4	2	0.2857	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.1429	0.4286	0.2857	0.5714	0.0000
#5	2	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.2857	0.2857	0.2857	0.2857	0.7143	0.2857
#6	2	0.1429	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.2857	0.1429	0.2857	1.0000	0.1429
#7	2	0.2857	0.1429	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.0000	0.1429	0.1429	0.1429	0.1429	0.0000	0.2857
#8	3	0.1429	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.2857	0.1429	0.1429	0.1429	1.0000	0.2857
#9	3	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.1429	0.2857	0.2857	0.5714	0.2857
#10	2	0.1429	0.1429	0.1429	0.0000	0.1429	0.1429	0.0000	0.1429	0.0000	0.2857	0.1429	0.4286	0.2857	0.5714	0.0000
#11	2	0.4286	0.1429	0.0000	0.0000	0.0000	0.1429	0.0000	0.0000	0.0000	0.2857	0.2857	0.4286	0.2857	1.0000	0.2857
#12	2	0.4286	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.1429	0.1429	0.1429	0.2857	0.1429
#13	3	0.4286	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.0000	0.4286	0.1429	0.1429	0.4286	0.8571	0.2857
#14	4	0.2857	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.2857	0.0000	0.0000	0.1429	0.1429	1.0000	0.0000
#15	3	0.1429	0.1429	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.0000	0.1429	0.2857	0.2857	0.2857	0.5714	0.2857
#16	2	0.2857	0.1429	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.0000	0.1429	0.2857	0.2857	0.2857	0.1429	0.1429
#17	2	0.1429	0.1429	0.0000	0.0000	0.0000	0.1429	0.0000	0.0000	0.0000	0.1429	0.1429	0.1429	0.2857	0.5714	0.1429
#18	2	0.2857	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.1429	0.1429	0.2857	0.1429	0.0000
#19	2	0.1429	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.2857	0.2857	0.2857	0.2857	0.1429	0.1429
#20	3	0.2857	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.1429	0.4286	0.1429	0.4286	0.2857
#21	3	0.4286	0.0000	0.0000	0.0000	0.1429	0.1429	0.0000	0.1429	0.0000	0.2857	0.2857	0.2857	0.4286	0.5714	0.2857
#22	2	0.4286	0.1429	0.0000	0.0000	0.1429	0.1429	0.0000	0.1429	0.0000	0.4286	0.2857	0.4286	0.2857	0.4286	0.1429
#23	2	0.2857	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.1429	0.1429	0.1429	0.5714	0.0000
#24	1	0.2857	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.1429	0.0000	0.1429	0.1429	0.1429	1.0000	0.2857
#25	3	0.2857	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.0000	0.0000	0.2857	0.1429	0.2857	0.2857	1.0000	0.2857
#26	3	0.4286	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.0000	0.2857	0.2857	0.4286	0.1429	0.4286	0.2857
#27	3	0.4286	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.2857	0.1429	0.4286	0.4286	1.0000	0.2857
#28	2	0.4286	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.1429	0.1429	0.2857	0.2857	0.1429
#29	3	0.2857	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.2857	0.4286	0.1429	1.0000	0.1429
#30	3	0.4286	0.1429	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.0000	0.4286	0.2857	0.2857	0.2857	0.8571	0.2857
#31	3	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.2857	0.1429	0.2857	0.2857	1.0000	0.2857
#32	2	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.2857	0.1429	0.1429	1.0000	0.0000
#33	2	0.1429	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.1429	0.2857	0.2857	0.0000
#34	2	0.4286	0.1429	0.0000	0.0000	0.1429	0.1429	0.0000	0.1429	0.0000	0.2857	0.2857	0.4286	0.2857	0.4286	0.2857
#35	2	0.1429	0.1429	0.1429	0.0000	0.1429	0.1429	0.0000	0.1429	0.0000	0.2857	0.2857	0.2857	0.2857	1.0000	0.2857
#36	4	0.2857	0.0000	0.0000	0.0000	0.1429	0.1429	0.1429	0.0000	0.2857	0.0000	0.0000	0.4286	0.1429	0.5714	0.2857
#37	3	0.2857	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.0000	0.4286	0.2857	0.4286	0.1429	0.4286	0.2857
#38	2	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.2857	0.2857	0.1429	0.2857	0.4286	0.2857
#39	3	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1429	0.0000	0.1429	0.1429	0.1429	0.0000	0.8571	0.0000

Figure 1: Imported the data set in ENN environment

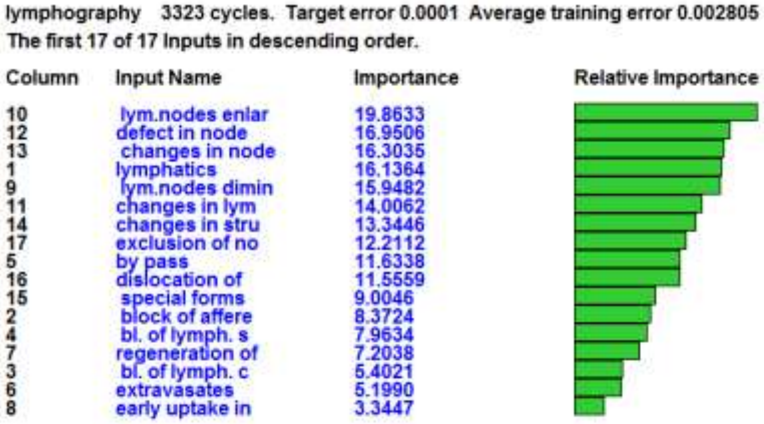


Learning rate:	0.559239
Momentum:	0.716271
Maximum error:	0.075956
Average error:	0.002805
Minimum error:	0.000003
Target error:	0.000100
Training examples:	103
Validating examples:	45
Within 10.0% range.	Score: 0
Correct if rounded.	Score: 44

Validating: 97.78% OK

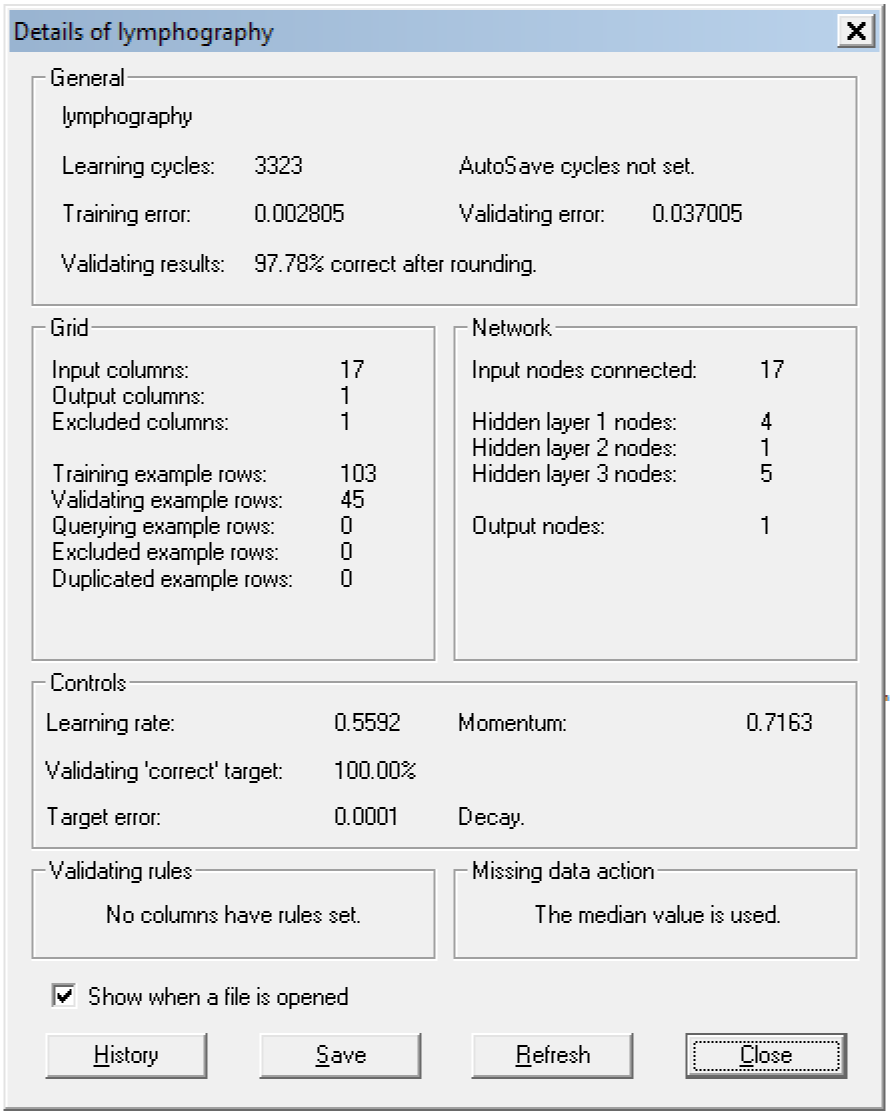
Layer:	Input		Hidden 1		Hidden 2		Hidden 3		Output
Nodes:	17		4		1		5		1
Weights:		68		4		5		5	

Figure 2: Training and validating of the ANN model in ENN environment


lymphography 3323 cycles. Target error 0.0001 Average training error 0.002805
The first 17 of 17 Inputs in descending order.

Column	Input Name	Importance	Relative Importance
10	lym.nodes enlar	19.8633	🟩
12	defect in node	16.9506	🟩
13	changes in node	16.3035	🟩
1	lymphatics	16.1364	🟩
9	lym.nodes dimin	15.9482	🟩
11	changes in lym	14.0062	🟩
14	changes in stru	13.3446	🟩
17	exclusion of no	12.2112	🟩
5	by pass	11.6338	🟩
16	dislocation of	11.5559	🟩
15	special forms	9.0046	🟩
2	block of affere	8.3724	🟩
4	bl. of lymph. s	7.9634	🟩
7	regeneration of	7.2038	🟩
3	bl. of lymph. c	5.4021	🟩
6	extravasates	5.1990	🟩
8	early uptake in	3.3447	🟩

Figure 3: Most important attributes of the ANN model



Details of lymphography

General

lymphography

Learning cycles: 3323 AutoSave cycles not set.

Training error: 0.002805 Validating error: 0.037005

Validating results: 97.78% correct after rounding.

Grid

Input columns:	17
Output columns:	1
Excluded columns:	1
Training example rows:	103
Validating example rows:	45
Querying example rows:	0
Excluded example rows:	0
Duplicated example rows:	0

Network

Input nodes connected:	17
Hidden layer 1 nodes:	4
Hidden layer 2 nodes:	1
Hidden layer 3 nodes:	5
Output nodes:	1

Controls

Learning rate: 0.5592 Momentum: 0.7163

Validating 'correct' target: 100.00%

Target error: 0.0001 Decay.

Validating rules

No columns have rules set.

Missing data action

The median value is used.

☑ Show when a file is opened

Figure 4: Details of the ANN model

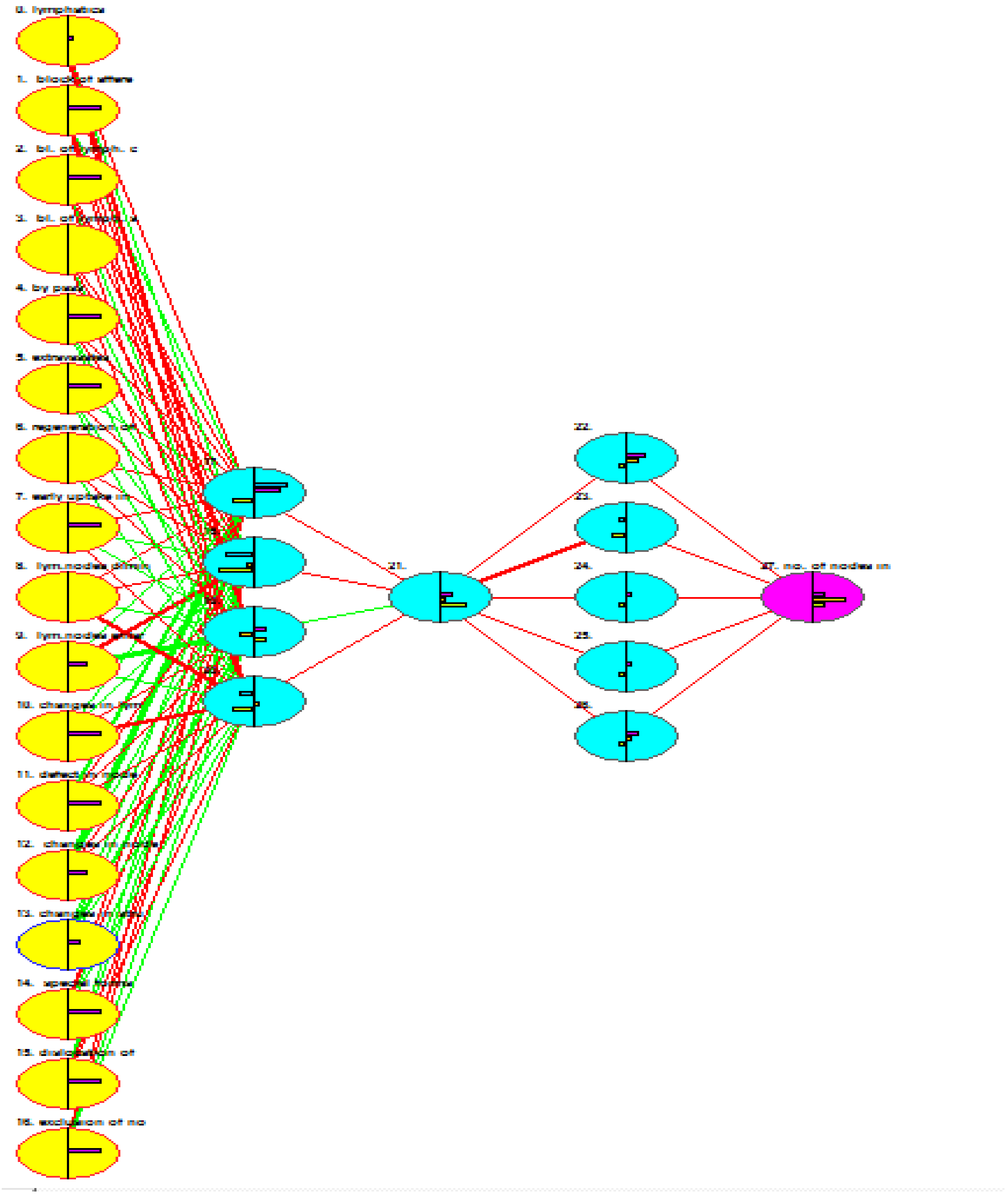
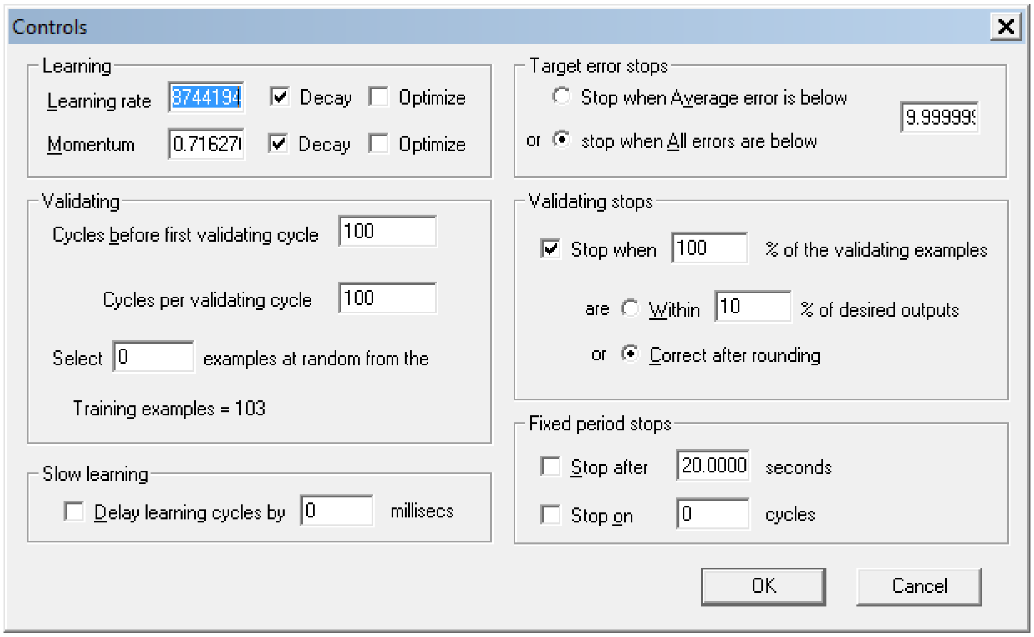


Figure 5: Architecture of the ANN model



Learning

Learning rate 8744194 ☑ Decay □ Optimize

Momentum 0.71627 ☑ Decay □ Optimize

Target error stops

○ Stop when Average error is below 9.99999

or ● stop when All errors are below

Validating

Cycles before first validating cycle 100

Cycles per validating cycle 100

Select 0 examples at random from the

Training examples = 103

Validating stops

☑ Stop when 100 % of the validating examples

are ○ Within 10 % of desired outputs

or ● Correct after rounding

Slow learning

□ Delay learning cycles by 0 millisecs

Fixed period stops

□ Stop after 20.0000 seconds

□ Stop on 0 cycles

Figure 6: The controls of the parameters of the ANN model

5. CONCLUSION

In this paper we have used Easy Neural Network (ENN) to classify clinical data and proposed the design of a classifier that is trained on the Lymphography dataset from the UCI Machine Learning Repository to perform class categorization of clinical data. We have evaluated the performance of the classification algorithm on the dataset. We have achieved and accuracy of 97.78%. This research will aid in enhancing the current state of ailment prediction and classification in the field of clinical research.

References

1. Kotsiantis (2007), Supervised Machine Learning: A Review of Classification Techniques, Informatica (31), 249-268.
2. Mitchell, Tom M (1997), Machine Learning. The Mc-Graw-Hill Companies, Inc.
3. Warwick, Roger; Peter L. Williams (1973) . "Angiology (Chapter 6)". Gray's anatomy. Illustrated by Richard E. M. Moore (Thirty-fifth Ed.). London: Longman. pp. 588–785.
4. Guermazi et al. (2003). "Lymphography: an old technique retains its usefulness". Radiographics 23 (6): 1541–58; discussion 1559–60.
5. Hammouda, and F. Karay, A Comparative Study of Data Clustering Techniques, Course Project, 2000.
6. Shoji Hirano, and Shusaku Tsumoto, “Cluster Analysis of Time-series Medical Data Based on the Trajectory Representation and Multiscale Comparison Techniques, Proceedings of the Sixth International Conference on Data Mining (ICDM'06).
7. Tzu-cheng et al. (2007). Boosting Classification Accuracy With Samples Chosen From A Validation Set, ANNIE, Intelligent ` Engineering systems through artificial neural networks, St. Louis, MO, pp. 455-461.
8. Kemal P. and Salih G. (2009). A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems”, Expert Systems with Applications: An International Journal,Volume 36, Issue 2, Pergamon Press, Inc. Tarrytown, NY, USA
9. UCI Machine Learning repository (https://archive.ics.uci.edu/ml/datasets.html)
10. EasyNN Tool.