Training Neural networks for Experimental models: Classifying Biomedical Datasets for Sickle Cell Disease

Mohammed Khalaf[1], Abir Jaafar Hussain[1], Dhiya Al-Jumeily[1], Robert Keight[1], Russell Keenan[2], Paul Fergus[1], Haya Al-Askar[3], Andy Shaw[1], Ibrahim Olatunji Idowu[1]

[1] Faculty of Engineering and Technology, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK
[2] Liverpool Paediatric Haemophilia Centre, Haematology Treatment Centre, Alder Hey Children's Hospital, Eaton Road, West Derby, Liverpool L12 2AP, UK
[3] College of Computer Engineering and Science, Sattam Bin Abdulaziz University, KSA

M.I.Khalaf@2014.ljmu.ac.uk, {a.hussain, d.aljumeily, p.fergus, a.shaw}@ljmu.ac.uk, Russell.keenan@alderhey.nhs.uk, R.Keight@2015.ljmu.ac.uk,

Abstract. This paper presents the use of various type of neural network architectures for the classification of medical data. Extensive research has indicated that neural networks generate significant improvements when used for the pre-processing of medical time-series data signals and have assisted in obtaining high accuracy in the classification of medical data. Up to date, most of hospitals and healthcare sectors in the United Kingdom are using manual approach for analysing patient input for sickle cell disease, which depends on clinician's experience that can lead to time consuming and stress to patents. The results obtained from a range of models during our experiments have shown that the proposed Back-propagation trained feed-forward neural network classifier generated significantly better outcomes over the other range of classifiers. Using the ROC curve, experiments results showed the following outcomes for our models, in order of best to worst: Back-propagation trained feed-forward neural net classifier: 0.989, Functional Link neural Network: 0.972, in comparison to the Radial basis neural Network Classifiers with areas of 0.875, and the Voted Perception classifier: 0.766. A Linear Neural Network was used as baseline classifier to illustrate the importance of the previous models, producing an area of 0.849, followed by a random guessing model with an area of 0.524.

Keywords: Sickle Cell Disease, Neural network Architectures, Real-time data, Receiver Operating Characteristic Curve, The Area under Curve, e-Health

1 Introduction

Sickle cell disease (SCD) is considered long-term disease that facing medical sector, affects human from early childhood. It has a high severe impact on the life expectancy and the patient's quality of life due to red blood cell (RBCs) abnormality. This disease affects patients who suffers from genetic blood disorders, where the haemoglobin behaves abnormal in the blood. SCD is phenotypically complex, with various medical

outcomes ranging from early childhood mortality to a nearly unrecognised condition [1]. With respect to sickle cell disease, the most significant symptoms that could show effects on patients are fatigue, shortness of breath, dizziness, and headaches. The World health Organisation (WHO) reported that 5% of the population around the world carries the haemoglobin disorder, mostly, sickle cell disease and thalassemia [2].

These days, the significance improvement in medical/science data have provide a new solutions to improve healthcare outcomes and services. Recent research has demonstrated the positive effects of a drug called hydroxyurea/hydroxycarbamide in terms of modifying the disease phenotype [3]. The clinical procedure to manage SCD modifying therapy is time consuming and difficult for medical staff. In order to curtail the significant medical variability presented by such difficult crisis, healthcare consultants need to improve adherence to therapy, which is regularly poor and subsequently results in fewer benefits and elevated risks to patients. Up to the present, the new trend of machine learning is becoming essential for the analysis of data within the medical domain. Machine learning (ML) provides a number of services for diagnostic and prognostic problems in a variety of clinical societies. ML techniques are being utilized for the analysis of the significance of healthcare parameters in association with their integrations for prognosis, for instance, overall patient management, providing therapy and support, and predict the disease progression [4].

Artificial Neural Network (ANN) is widely-used classification methods for medical domains [5]. This type of machine learning approach can identify groups of attributes in order to recognise the specific kind of diseases and illnesses. The main reason behind that is due to their characteristics of parallel processing, self-organization, non-linearity, and the most important point is self-learning [6]. In this study, the applications of neural networks approaches for health datasets classification will be discussed. The performance of four types of neural network approaches, including the feed-forward neural net classifier (BPXNC), the Functional Link neural Network (FLNN), Radial basis neural Network Classifiers (RBNC), and Voted Perception classifier (VPC), will be considered. The networks will be applied to classify the amount of medication dosage for those who suffer from SCD, the results of which will be presented and discussed as a case study.

The rest of this paper is structured as follows. Section 2 will discuss various neural network architectures while section 3 shows the clinical care pathway. The methodology for classifying sickle cell datasets is shown in section 4, while section 5 shows the result section. Section 6 illustrates the conclusion.

2 Artificial Neural networks (ANN)

Artificial Neural networks (ANN) are considered a form of biologically inspired algorithm (BIA), based on the constellations of connected elements observed in the biological brain, namely networks of specialised cells called neurons [7]. ANN are inspired software programs that were designed in order to simulate the way that mankind brain processes any type of information. ANN attempts to model actual systems depending on the information provided to it. This kind of machine learning involves hundreds of single unit artificial neurons and is considered a powerful computational and mathematical data model that is cable of representing complex input and output connections. The main motivation behind developing an artificial neural network is the capacity to perform intelligent tasks, which are performed in a manner similar to the operation of the human brain. Furthermore, the power of ANN comes from connecting the neurons in a one particular network that have ability to represent non-linear and linear relationships.

Artificial neural networks for computational modelling need a number of neurons so that they can be connected together to form a network. In this context, neurons are organised in layers and have processing units, which takes one or more inputs to generate an output. In this case, at each neurone, all inputs have to be connected with a weight that modifies the strength of each input. As a result, neurons will simply collect all the inputs together to calculate an output. The weights in each ANN are trained using different types of learning algorithm, for example supervised and unsupervised learning. In order that the network promotes the most important features within training process, learning algorithms are utilised to update the weights of the ANN using mathematical equations.

The main backbone of using weights is to check the output if it is too high, then the weights should be lowered by with certain amount to be fit to the output for all the input instance. On the other hand, if the predicated output is too low, then the weights need to be increment by the set amount. The hidden layer learns to provide a representation for the inputs. In each ANN, one or more hidden layers can be applied. The process of constructing such an architecture is referred to as learning in ANN [8]. There are four difference neural networks classifiers have been used in this experiments:

A. **The back-propagation trained feed-forward neural network classifier (BPXNC):** This type of neural networks is trained to set a number of input data through making adjustment of the complete weights [9]. In this context, the information t collected from imputes is fed forward to improve the weights between neurons. This classifier then will be able to read the output and inputs value within the training datasets in order to decrease the differentiation between observed and predicted values [10].

B. **Functional Link Neural networks (FLNN):** FLNN has single layer of neurons capable to handle non-separable classes through by increasing the dimensionality of the whole input spaces. It is called a flat classifier. This feature provides such an excellent support that makes the learning algorithm in the network less complex. We have selected this model in the purpose of enhancing the performance of network.

C. **Voted Perception classifier (VPC):** This algorithm is based on the perceptron models of frank and Rosenblatt. The classifiers can work with the input data that is considered linearly separable. It uses kernel functions to display very high dimensional spaces. The VPC is much more efficient and simple to implement. VPC is trained using supervised learning algorithms. The training process implements many full sweeps by the training data. In this model, the classification are performed for a new objective by permitting the ensemble of perceptron's to make vote in the neural network on the label of each test point [11].

D. **The Radial basis neural Network Classifiers (RBNC):** RBNC is considered one of the most popular models in neural network architectures, which is often utilised in classification problem and complex pattern recognition. Typically, this models has an input layer, a one hidden layer but without activation function, and output layer. The mapping procedure can be modified for RBNC through adjusted the weights values in the output layer [9].

3 Clinical Care Pathway

Clinical care pathways also known as integrated care pathways are considered one of the major methods to manage and provide the best quality in clinical concerning the standardisation of care procedures. It is obvious that their implementation improves outcomes and decreases the variability in medical practice [12]. In this context, this type of processes refer to healthcare guidelines and how the medical consultants can treat the SCD patients according to their conditions. It promotes efficient and organised patient care depend on practice. The major goal of this study is to explore and develop methodology for the datasets that are collected from the local hospital, which can increase the efficiency of the data analysis procedures. Healthcare professional provides 50 mg of hydroxyurea drug/liquid until 1700 mg according to the patient's condition. Figure 1 demonstrates the full flowchart that healthcare professionals use to analyse patients' blood test and provide accurate amount of medication.
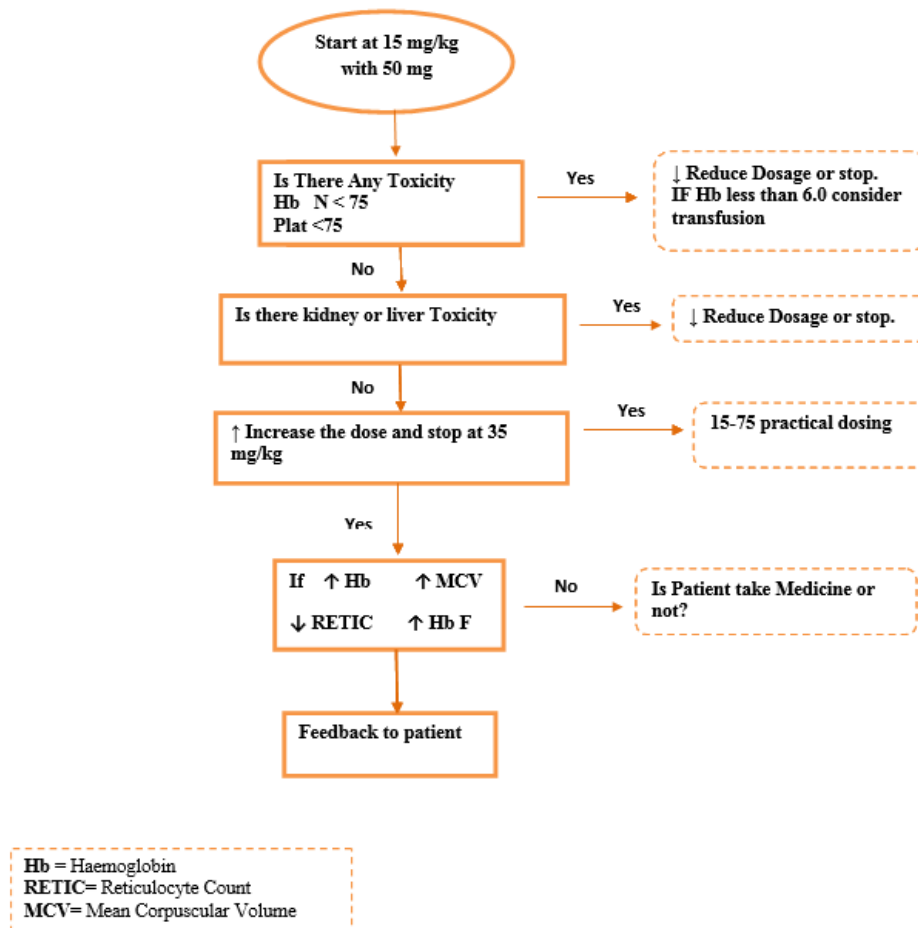
Fig. 1. Clinical Care Pathway

4 Methodology

In this paper, the previous work in the field of machine learning have been forced to check and predict the severe crises of SCD, rather than using optimal predication techniques to give correct amounts of hydroxyurea in modifying the disease phenotype [9]. Currently, there is no standardisation of disease modifying therapy management. Through using the proposed computerised comprehensive system, the goal is to provide an optimised and reproducible standard of care in various medical

settings across the UK. The main aim of this research is to use recent advanced neural network approaches, in order to provide an excellent support to the clinicians by providing specific medication for each individual conditions. In this context, and due to the pattern of the SCD dataset, we attempt to propose classification of the patient's datasets records at an earlier stage, according to how much of a dose the patient will need to take. This can potentially lower the costs, avoid unnecessary admission to hospitals, mitigate patient illness before proceed to the worse condition, improve patient welfare, and unnecessary interventions.

Neural networks approaches can be implemented to figure strong integrated classifiers, including previous patient cases that have been collected from the local hospital for SCD in Liverpool, UK, over the last ten years using 1168 patient records.

4.1 Data Collection

The original datasets that are used in our study for SCD were gathered within a ten-year period. Each patient sample covers 13 instances deemed vital factors for classifying the SCD datasets as showed in Table 1. In order to collect these datasets, the local hospital has supported this study with many patient records to obtain better accuracy and services. The resulting dataset comprised 1168 sample points, with a single target variable describing the hydroxyurea medication dosage in milligrams. To facilitate our classification study, the target dosage was discretised into 6 bins, denoted classes 1 through 6, formed through dividing the output range (in Milligrams) into membership intervals of equal size: Class 1: $[148 \leq Y < 410mg]$, Class 2: $[410 \leq Y < 659mg]$, Class 3: $[659 \leq Y \leq 919mg]$, Class 4: $[919 \leq Y < 1200mg]$, Class 5: $[1200 \leq Y < 1430mg]$, Class 6: $[1430 \leq Y \leq 1700mg]$.

Table 2. SCD Datasets Features.

No	Types of Attributes
1	Weight
2	Haemoglobin(Hb)
3	Mean Corpuscular Volume (MCV)
4	Platelets(PLTS)
5	Neutrophils (white blood cell NEUT)
6	Reticulocyte Count (RETIC A)
7	Reticulocyte Count (RETIC %)
8	Alanine aminotransferase (ALT)
9	Body Bio Blood (BIO)
10	Hb F
11	Bilirubin (BILI)
12	Lactate dehydrogenase (LDH)

4.2   Experimental Setup

The experimental setup covers the design of the test environment used in our experiments, the models tested, and the configuration of each model. Finally, the performance evaluation metrics utilised to measure the results of the neural network algorithms are presented for the SCD data set.

This study are composed of trained models using four types of integrated neural network approaches: Back-propagation trained feed-forward neural net classifier, Functional Link neural Network, the Radial basis neural Network Classifiers, and the Voted Perception classifier. These models are appropriate to act as comparators of high performance [13]. The linear model used includes a linear transformation function with a single layer neural network at each class output unit. To obtain performance estimates for the respective models, we ran each simulation 50 times and calculated the mean of the responses. The full set of models used in the experiments are described in Table 2. Finally, the random oracle model is used to establish random case performance through the assignment of random responses for each class.

Table 2. Classification Models Description.

Models	Description	Architecture	Training Algorithm	Role
ROM	Random Oracle Model	Pseudorandom number generator	N/A	Random Guessing Baseline
BPXNC	Feed-forward neural network algorithm	Context Units: One context unit for each output unit.	This classifier is trained properly to map a set of input data in order to make iterative modification for the whole weights.	Non-linear Comparison Model
FLNN	Functional link neural network	Units: 13-30-3, tansig activations.	Gradient descent with momentum and adaptive learning rate backpropagation	Test model
RBNC	Feed-forward neural net including N sigmoid neurons.	13 inputs, 3 outputs	The classifier has radial basis units with only 1 hidden layer	Non-linear Comparison Model
VP C	Combines an ensemble of perceptron's through voting procedure.	Units: 13-30-3, tansig activations	This Classifier performed by permitting the ensemble of perceptron's to vote in the neural network on the label of each test point.	Non-linear Comparison Model
LN N	Linear Combiner Network	Units: 13-3, linear activations	Batch training with weight and bias learning rules	Linear Comparison Model

We used a competing models to the classification task. In addition, to apply ROM that provide a baseline indicator to show the performance yielded by random guessing [14]. Furthermore, we presented a linear model to discover the variance in performance presented between this weak classifier and the non-linear classifiers, such as neural networks. The integration of random control baselines, weak and strong, offers an experimental frame through which to gauge the relative performance.

Table 3. Performance Metric Calculation.

Metric Name	Calculation
Sensitivity	TP/(TP+FN)
Specificity	TN/(TN+FP)
Precision	TP/(TP+FP)
F1 Score	2 * (Precision*Recall)/(Precision+Recall)
Youden's J statistic (J Score)	Sensitivity + Specificity − 1
Accuracy	(TP+TN)/(TP+FN+TN+FP)
Area Under ROC Curve (AUC)	0 <= Area under the ROC Curve <= 1

In our study, the classifier models evaluation involves of both in-sample (training), and out-of-sample (testing) diagnostics, involving sensitivity, specificity, precision, the F1 score, Youden's J statistic, and the classification accuracy calculated as shown in Table 3. Moreover, the classifiers were categorised using the Receiver Operating Characteristic (ROC) plots and the Area under the Curve (AUC) and, where the classification capability across all operating points was determined.

5 Results

In this part, we analyse the results from the various experiments that have been applies in this study as listed in Tables 4 and 5, presenting outcomes for training and testing of the models, respectively. We run further performance visualisations through the use of AUC comparison plots (Figures 2 and 3) and the ROC plots as shown in Figures 4 and 5. The results obtained from the experiments show that the chosen dataset exhibits significant non-linear relationships, presenting a challenge for the test models. In order to make compassion among classifiers, the BPXNC outperformed the rest of models as demonstrated in Table 5, illustrating capability both in generalising and also in fitting the training data to unseen examples. The calculated means of AUCs for the BPXNC model, obtained for six classes during training yielded an area of 0.997 (ideal), in comparison to 0.989 over the test sample. Classes one to six were found to illustrate optimal consistent generalisation and performance from the training to the test sets for this classifier. It was discovered that the FLNN model was able to yield an average AUC of 0.972, outperforming the RBNC classifier and showing an overall rank of second place. The RBNC model produced an average AUC of 0.875, ranking third overall, outperforming the lower ranking models by a reasonable margin. All three of the top performing models, the BPXNC, FLNN, and RBNC, obtained nearly ideal AUCs and represent viable candidates for future use. These models produced exceptional results in terms of both training and generalisation. We found that the VPC classifiers, constituted the next performance level obtained, with test AUCs of 0.766 respectively

The LNN was able to learn the non-linear components in the data and so produced such an excellent classification results. The average test of AUCs for this model ranged 0.849, which is seen to demonstrate performance significantly below that of the other models. The ROM random guessing model is as expected produce a weak classification as shown in the ROC plots for all classes (see Figures 1 and 2), illustrating by contrast the consequence of the results from the other trained classifiers. The average AUC for this model yield 0.524, which considered the lowest outcomes among other models. This process of guessing yields both train and test set AUCs lower than the LNN trained model baseline.

Table 4. The Performance of Classifiers (Training phase).

Model	Sensitivity	Specificity	Precision	F1	J	Accuracy	AUC
ROM	0.493	0.542	0.179	0.253	0.0351	0.543	0.488
BPXNC	0.983	0.988	0.93	0.954	0.97	0.986	0.997
VPC	0.743	0.73	0.337	0.45	0.474	0.733	0.781
RBNC	0.83	0.865	0.62	0.702	0.695	0.86	0.907
FLNN	0.954	0.964	0.809	0.873	0.918	0.963	0.985
LNN	0.854	0.827	0.524	0.642	0.68	0.835	0.87

Table 5. The Performance of Classifiers (Testing phase).

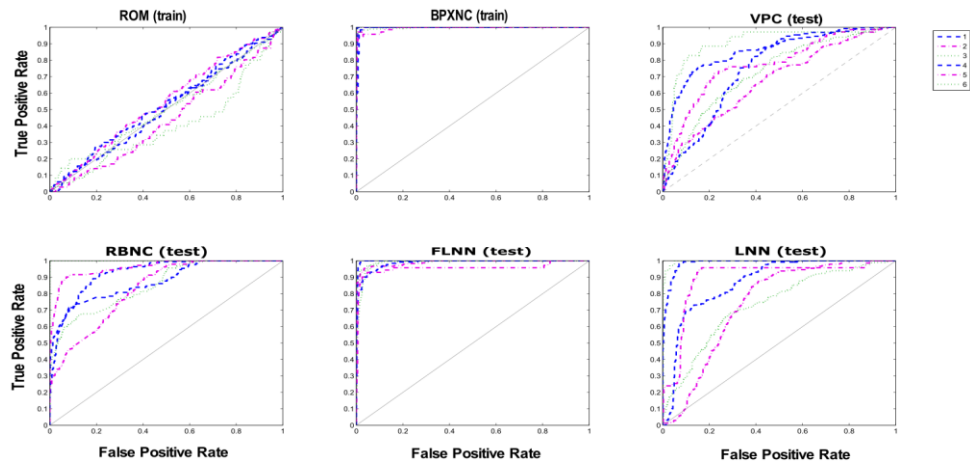| Model	Sensitivity	Specificity	Precision	F1	J	Accuracy	AUC
ROM	0.548	0.547	0.197	0.268	0.0949	0.539	0.524
BPXNC	0.978	0.98	0.91	0.942	0.958	0.98	0.989
VPC	0.729	0.745	0.383	0.496	0.475	0.749	0.766
RBNC	0.835	0.842	0.526	0.639	0.678	0.845	0.875
FLNN	0.95	0.966	0.803	0.862	0.916	0.962	0.972
LNN	0.824	0.848	0.534	0.641	0.672	0.84	0.849

Fig. 2. ROC curve (Train) For classifiers
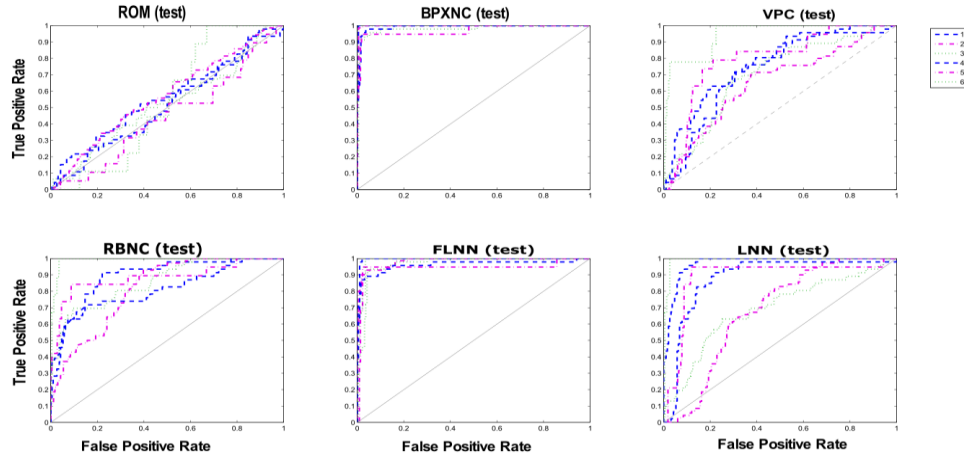
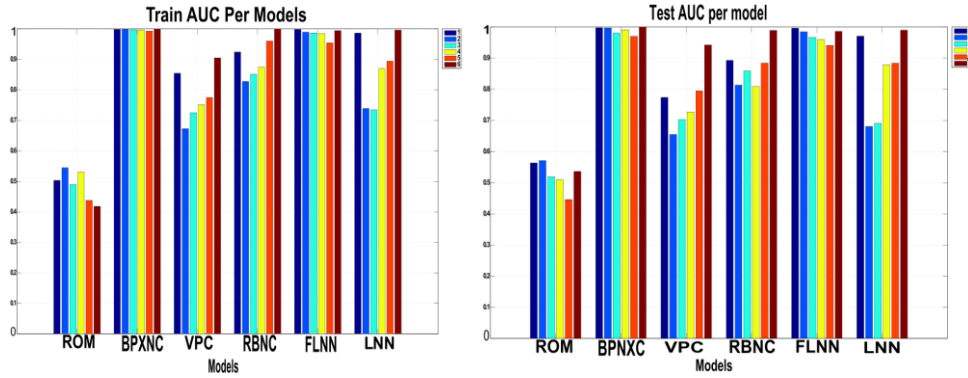**Fig. 3.** ROC curve (Testing) for classifiers

Fig. 4. Train AUC per model   **Fig. 5**. Testing AUC per model

In order to explore our datasets in 2- dimensional, the data representation was explored to investigate if there is any regularities that might be uncovered within its structure. Exploratory analysis is considered such a significant step in the machine learning approaches, permitting the human advisor to obtain an intuition of the data as well as the potential learnability of the data. The results from the whole data exploration can be utilised to lead the modelling point, since a key component learnability is known to be a task of the correspondence between the kind of demonstration it is supplied with and the learning algorithm. To undertake an exploration of the utilised data in these experiments, we computed summary statistics, followed by visualisation methods including t-distributed Stochastic Neighbourhood Embedding (tSNE) and Principal Component Analysis (PCA) and as reported in

Figures 6 and 7, respectively. Results from the visualisation procedures reveal that some discernible structure is present within the data. The tSNE plot demonstrates that the data can be geometrically separated when allowing different intervals of dosage level. The PCA plot indicates that there are potential clusters of values present within the data. Moreover, the exploratory data analysis illustrates no noticeable defects that might call into question the outcomes of consequent analysis.
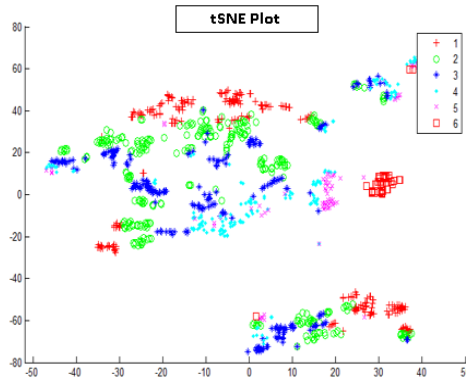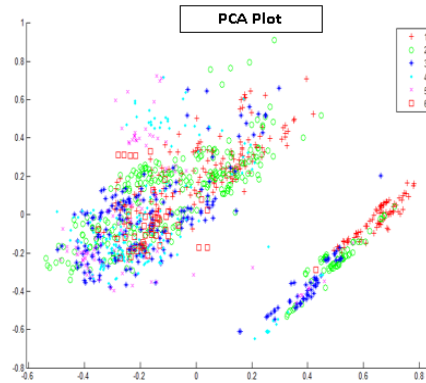

Fig. 6. tSNE plot   **Fig. 7.** PCA plot

6 Conclusion

This study presents an empirical investigation into the use of various neural network algorithms to classify the level of dosage for SCD medication. In this paper, various model architectures are used for analysing the medical datasets obtained from SCD patients. The main purpose of this research is to examine the effectiveness of these models in terms of training and testing setting, investigating if such architectures could enhance classification results. It was found through experimental investigation, including the usage of SCD datasets and approaches such as BPXNC, the FLNN, RBNC, and VPC that the analysis of medical datasets is viable and produces precise results. The results gained from a range of models during our experiments have demonstrated that the proposed Back-propagation trained feed-forward neural network classifier produced the best results with the AUC 0.989 in compare with other models. It is found that the outcomes are based on a considerable data sample, containing more than 1100 sample examples, which supports the significance of the findings. It is recommended that further research will be used to make confirmation on our findings, where a large number of data could be utilised also to advance the performance of the results. In this circumstance, we recommend however that a machine learning algorithms, for instance, support vector machine and k-nearest neighbour classifiers could be used to increase the scale and scope of this research.

References

[1]   P. Sebastiani, M. F. Ramoni, V. Nolan, C. T. Baldwin, and M. H. Steinberg, "Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia," Nat Genet, vol. 37, pp. 435-440, 04//print 2005.

[2]   D. J. Weatherall, "The inherited diseases of hemoglobin are an emerging global health burden," Blood, 2010.

[3]   M. Kosaryan, H. Karami, M. Zafari, and N. Yaghobi, "Report on patients with non transfusion-dependent β-thalassemia major being treated with hydroxyurea attending the Thalassemia Research Center, Sari, Mazandaran Province, Islamic Republic of Iran in 2013," Hemoglobin, vol. 38, pp. 115-118, 2014.

[4]   G. D. Magoulas and A. Prentza, "Machine learning in medical applications," in Machine Learning and its applications, ed: Springer, pp. 300-307, 2001.

[5]   Q. K. Al-Shayea, "Artificial neural networks in medical diagnosis," International Journal of Computer Science Issues, vol. 8, pp. 150-154, 2011.

[6]   B. Liu, M. Wang, L. Yu, Z. Liu, and H. Yu, "Study of feature classification methods in BCI based on neural networks," in Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the, pp. 2932-2935,2006.

[7]   A. Brabazon and M. O'Neill, Biologically inspired algorithms for financial modelling: Springer Science & Business Media, 2006.

[8]   N. Karayiannis and A. N. Venetsanopoulos, Artificial neural networks: learning algorithms, performance evaluation, and applications vol. 209: Springer Science & Business Media, 2013.

[9]   I. O. Idowu, P. Fergus, A. Hussain, C. Dobbins, and H. A. Askar, "Advance Artificial Neural Network Classification Techniques Using EHG for Detecting Preterm Births," in Complex, Intelligent and Software Intensive Systems (CISIS), 2014 Eighth International Conference on,pp. 95-100, 2014.

[10]   P. Fergus, H. De-Shuang, and H. Hamdan, "Prediction of intrapartum hypoxia from cardiotocography data using machine learning," Applied Computing in Medicine and Health-Emerging Topics in Computer Science and Applied Computing, pp. 125-146, 2016.

[11]   A. J. Hussain, P. Fergus, H. Al-Askar, D. Al-Jumeily, and F. Jager, "Dynamic neural network architecture inspired by the immune algorithm to predict preterm deliveries in pregnant women," Neurocomputing, vol. 151, pp. 963-974, 2015.

[12]   M. Panella, S. Marchisio, and F. Di Stanislao, "Reducing clinical variations with clinical pathways: do pathways work?," International Journal for Quality in Health Care, vol. 15, pp. 509-521, 2003.

[13]   W. Duch, "Towards comprehensive foundations of computational intelligence," in Challenges for Computational Intelligence, ed: Springer, 2007, pp. 261-316.

[14]   X.-Y. Jia, B. Li, and Y.-M. Liu, "Random oracle model," Ruanjian Xuebao/Journal of Software, vol. 23, pp. 140-151, 2012.