Journal of Applied Sciences, Information and Computing
Vol. 2, No. 1 (June, 2021)
School of Mathematics and Computing, Kampala International University

ISSN: 1813-3509 https://jasic.kiu.ac.ug

Performance Evaluation of Classifiers Created using Elman Back-Propagation and Cascade Feed-forward Neural Networks

Adeleke Raheem Ajiboye[1], Fatima Usman-Hamza[2], Muyideen AbdulRaheem[3], Chinecherem Umezuruike[4]

[1,2,3]Department of Computer Science, Faculty of Communication & Information Sciences, University of Ilorin, Ilorin, Nigeria.ajibraheem@unilorin.edu.ng usman-hamza.fe@unilorin.edu.ng muyideen@unilorin.edu.ng
[4]Department of Information Technology, School of Mathematics & Computing, Kampala International University, Kampala, Uganda. *cumezuruike@kiu.ac.ug*

Abstract

Huge data being captured in our day-to-day activities are mostly imbalance. Such data therefore, calls for fast, accurate and robust techniques, through which they could be analyzed in order to fast-track early decision making. A Cascade Feed-forward Neural Networks and Elman Backpropagation are known techniques in neural network domain and their efficacies is therefore tested on separable data in this study. The objective of this study is to evaluate the performance of these techniques in solving a linear classification problem. The linear classification of data involves, splitting of separable data into two distinct clusters. In order to achieve the goal of this study, linear classifiers were created using the two aforementioned techniques. Both network structures were exposed to the same dataset and similar parameter configurations were set for each technique. The model that emanates through each technique was simulated using the set of untrained data. In order to determine the accuracy of each model created through each technique, their Mean Absolute Errors (MAE) was computed. The performance of each model was determined based on the value of MAE. The error computation for the simulated output reveals that, cascade neural network gives an error of 0.0928; while the model created using Elman Backpropagation network gives a relatively lower error of 0.0661. It can be inferred from this study, that, both techniques are capable of fitting accurate classifiers from dataset and specifically, both techniques are very suitable for binary classification of separable data.

Keywords: Linear classification, Elman backpropagation, Cascade feed-forward, Machine learning. Neural networks, Mean Absolute Error.

---

1. Introduction

In several applications that occur in real time, huge amount of data is generated; some of these data are usually imbalance due to their skewed distribution. A data set is said to be highly skewed if sample from one class is in higher number than other (Wang & Yao, 2012). The Datasets is imbalance if a particular class has more number of instances, this is referred to as a major class, while the one having relatively less number of instances are referred to as minor class (Chawla et al., 2004). This is a clear case of dataset that require the classification algorithm to achieve a distinct class.

Classification problems can be solved using machine learning techniques. Sometimes, there may be need to classify data into binary for decision making purposes, this is otherwise referred to as linear classification problems. There are some well-known linear classification methods that have been applied to text categorization problems, typical examples of techniques in these categories include linear least, squares fit, logistic regression, and support vector machines (Zhang & Oles, 2001). SVMs belong to the general category of kernel methods. A kernel method is an algorithm that depends on the data only through dot-products (Walker, 2010).

Artificial Neural Networks (ANNs), have been found to be an outstanding tool that is capable of fitting a generalizable model from datasets (Taud & Mas, 2018). ANNs are massively parallel computing systems that comprised of a very large number of simple processors with several interconnections. In this study, our focus is on two selected techniques in the neural networks domain: Elman back-propagation and Cascade feed-forward neural networks. Both networks are capable of creating a classification model for solving complex and non-complex problems, in other words, both neural network architectures are very suitable for linearly separable data.

Elman neural network (ENN) is a typical example of a recurrent neural networks (Ren et al., 2018).

In the course of implementing neural network structure for data classification, it involves establishing the relationship between the input variables and the corresponding target variables; this is otherwise referred to as supervised learning approach. In the process, weight is attached to each input and a number of neuron, including the bias. A bias is added to the weighted sum of the inputs and the neuron, prior to passing through the chosen transfer function. The bias determines the level of incoming activations that are required in order for the neuron to fire.

The motive of this study is to compare the results of evaluating the accuracies of the models created using the two selected state-of-the-art neural network architectures, Elman and Cascade. At the end of the experiment, the results of the evaluation reveals that, both techniques are very suitable for the linear classification of separable data, however, the model created using Elman back-propagation neural network shows a much lower error.

This paper is structured as follow: the next section discusses the concept of linear classification of data and what is meant by supervised learning. In Section 3, some of the studies related to the proposed study are briefly reviewed. In Section 4, the material and methods used in this study for the creation of the models for the classification of linearly separable data is discussed. Section 5 shows the experiment carried out based on the two techniques. The results of this study are shown and discussed in Section 6. This study is concluded in Section 7.

2. Data Classification Approaches and Supervised Learning

Data classification involves the analysis of structured or unstructured data and the process of organizing the data into categories that are based on file type, contents, and other metadata (Petters, 2020). In other words, *data classification* is the process of organizing data into categories for its most effective and efficient use. One of the benefits of classifying data is to help the organization to mitigate risk and manage governance policies.

In data classification problems, data can be divided into several types, such as commercial data, texts, DNAs and images (Hossin & Sulaiman, 2015). Some of the existing methods used for data classifications include: logistic regression, Naïve Bayes, Decision tree, ANNs, k-nearest neighbor, and several other techniques. These techniques are more traditional but still commonly used in the scientific community (Kowsari et al., 2019).

The classifiers that can effectively be used for data classification as reported in (Fernández-Delgado et al., 2014), belongs to a number of families such as: discriminant analysis, Bayesian, neural networks, support vector machines, decision trees, rule-based classifiers, boosting, bagging, stacking, random forests, generalized linear models, nearest-neighbors, partial least squares, principal component regression, logistic and multinomial regression, multiple adaptive regression splines and other ensembles.

At times, the class label may be known before the commencement of training ass in the case of the proposed study, this approach is referred to as supervised learning. In supervised learning, the training data comprised of input-output pairs. The learning algorithm makes use of the training data together with the corresponding target value for each training vector. The focus of the supervised learning algorithm was to create a prediction function using the training data that will generalize for unseen training vectors, in other to classify them correctly. Supervised learning approach has two phases: training and testing. Training builds a model using a large sample of historical data called a training set, and simulation involves trying out the model on new, previously unseen data to determine its accuracy and physical performance characteristics.

3. Review of Related Studies

Several classification techniques have been reported in the literature for the classification of imbalance text (Chawla et al., 2004), classification of audio signals (Dhanalakshmi et al., 2009), classification of schizophrenia data (Rampisela & Rustam, 2018), and lots of others. The use of neural network for classification purposes is identified in (Fernández-Delgado et al., 2014), as one of the techniques that can guaranteed accuracy. The linear classification of data using a Single Layer Perceptron (SLP) neural network (Fontenla-Romero et al., 2010; Khalid & Singh, 2017; Narayanan et al., 2004) and other techniques within the neural network domain are well reported in solving classification problems.

A number of studies have been proposed for the classification of linearly separable data. The study in (Deng et al., 2019), proposed a model of double descent for High-dimensional Binary Linear Classification. The study uses a logistic regression also known as generative model for the training of a linear classifier. The results from the study validate the theoretical predictions and unveil double-descent phenomena that complement similar recent findings in linear regression settings.

Also, the simulation of learning logical function which was reported in (Ahamad et al., 2020), used the structure of Single Layer Perceptron as a learning technique. The model created is considered as a binary classifier for the linearly separable set of inputs. The result of the study shows some strength of SLP in the separation of data linearly.

Although, Elman back propagation is mostly used for prediction (Kelo & Dudul, 2012) or forecasting (Jujie et al., 2014) and for other related purposes. In the study of Epilepsy classification proposed in (Rajaguru & Prabhakar, 2017), the researchers used the technique in conjunctions with multilayer perceptron to achieve an accurate classification of the task.

The use of Cascade neural network proposed in (Li et al., 2015), focused on using the technique for facial detection; this is also a classification problem and the study reported a better accuracy.

Feed-forward Cascade is also found to be suitable for disease classification as shown in the study proposed in (You et al., 2020). Also, in the study proposed and reported in (Cenate et al., 2016), the cascade feed-forward was optimized in order to give an efficient result in determining the defect classification in ultrasound images.

Although, the two learning techniques selected in the domain of neural network are well reported for solving forecasting and classification problems, however, the reviewed works failed to compare the performance of these techniques in solving classification tasks, such as linear problems, most especially when presented with the same dataset. This is the focus of the present study. The present study evaluates the accuracy of the classifier models created using the two selected architectures in the neural network domain.

4. Material and Methods

i. Data collection.

The data used for the implementation of the classifiers created based on the two neural network techniques was retrieved from World Bank open repository (Databank, 2020). The dataset comprised of 224 samples, out of which 139 were trained and another 85 untrained set of data were used for simulation. The explored data is a summarized report of the World Bank development indicators that shows the mortality rate, neonatal (per 1,000 live births) in over 200 nations across the globe. The focus of this study is to classify the data linearly in order to show the most highly affected nations and the low affected nations as regards mortality rate within the data validity period. Using the *divide rand* function for the partitioning of the data being trained in Matlab environment; the divisions of the dataset for training, validation and testing is as shown in Figure 1.

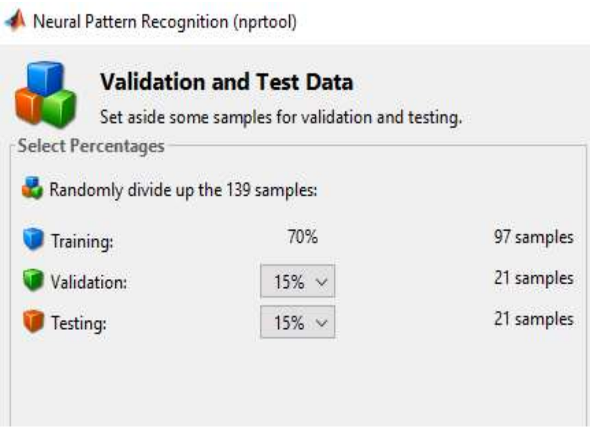

Figure 1: Randomly divided dataset for training

ii. Data pre-processing

The data retrieved from the open repository consist of some missing data. The preprocessing of the data got rid of the data anomalies; this was carried out in order to prepare the data for mining. Also, in the course of preprocessing the data, a threshold value was taken; the sum of the data recorded for a period of seven years that falls below the threshold value is classified as a nation with low mortality rate (0), and if otherwise, it is classified as high mortality rate (1).

All the data are found to be numeric type and the class label otherwise referred to as target variable or expected outputs comprised of 0 and 1 only.

iii. Implementation of the selected techniques

The techniques selected for implementation among the neural network techniques are: Elman back-propagation and Cascade feed-forward structure.

a. Creation of classification model using Cascade Feed-forward Backpropagation Network

Fitting of classifiers from the pre-processed dataset involves implementing the relevant algorithm on the dataset. The cascade feedforward neural network was used in this study to create the classifiers model in the Matlab environment. The architecture of the proposed network model is illustrated in Figure 2. It has seven inputs which comprised of the data recorded for a period of seven (7) years, 2013 to 2019 and a single expected output of either 0 or 1. The classification model created in Figure 2, was simulated using a set of untrained dataset. The errors generated were computed using the Mean Absolute Error as described in Section 4 (iiic.).

Also, a number of parameters were set, the configuration that gives the architecture in Figure 1, is as shown in Table 1.

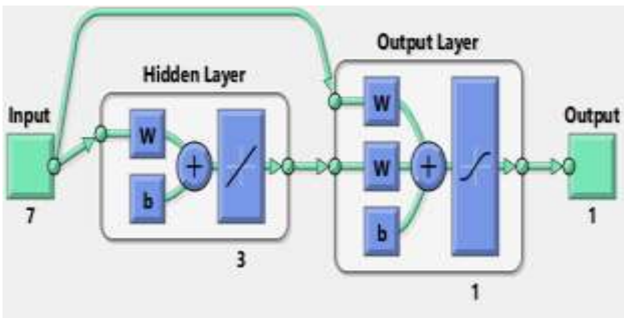

Figure 2: Cascade feedforward neural network architecture

Table 1: The Configuration settings for both cascade feedforward and Elman Backpropagation Neural Networks

Settings	Parameters
Training function	TrainLM
Learning function	LearnGDM
Performance function	Mean Square Error
Number of Layers	2
Number of Neuron	3
Transfer function	Purelin
Data division	Random

a. Creation of classification model using Elman Backpropagation Network

The last part of the implementations of the algorithm on the dataset has to do with using the Elman Back-propagation technique to maximally explore the dataset for binary classification. The implementation was carried out and repeated in order to observe patterns of the resulting output, which was found to be consistence. The architecture of Elman Back-propagation network is as shown in Figure 3. The structure comprised of 7 inputs and a single variable outputs, which is the label or expected outputs. The network output was automatically generated immediately after the training converges. The classifier created was later simulated using untrained data set to ascertain the accuracy of the model. The computation was done using Mean Absolute Error as shown in equation 1.

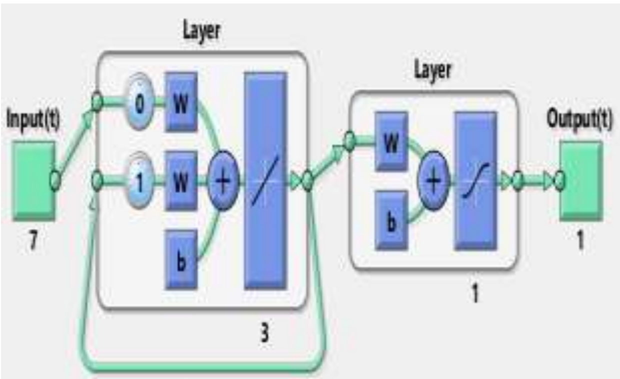

Figure 3. Elman Back-propagation neural network architecture

a. Evaluation of the generated outputs

In order to determine the reliability of the classifiers created in this study, the Mean Absolute Error (MAE) was computed for the generated results. This involves subtracting the network outputs from the expected outputs, adding up the differences, and divides the whole summation by the number of observation data. This is in compliance with the equation represented in equation 1.

MAE =

$$\frac{|p_1 - y_1| + \ldots + |p_n - y_n|}{n} \quad (1)$$

where p denotes the expected output, while y is the predicted output and n denotes the size of the data. At times, regardless of whether the differences in p and y are negative, the MSE of the differences would still be positive.

5. Discussion of Results

The result of creating a linear classification model with each of the technique implemented in this study is discussed here. The architecture in respect of each technique gives the graphical representation of the training process as illustrated in Figure 4. The training basically solves linear classification problem; In both network structures, a reduced number of iteration is recorded and within a short number of epochs, the networks converges. The three lines shown in each graph specifically represent: training, testing and validation data. The algorithm used for training is designed to bring about flexibility in the division of data for training. The word flexibility here implies that, data partitioned for training is not static. As the training progresses, the error values also continue to decrease as shown in Figures 4 and 5. Also as the iteration progresses, the value of MSE is being consistently checked by the validation data. The role of the validation data was to terminate the training process whenever the error begins to rise.

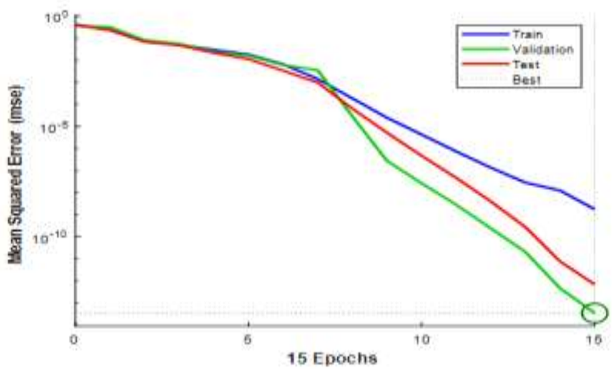

Figure 4: Graph showing the training process using Trainlm in Elman BP network.

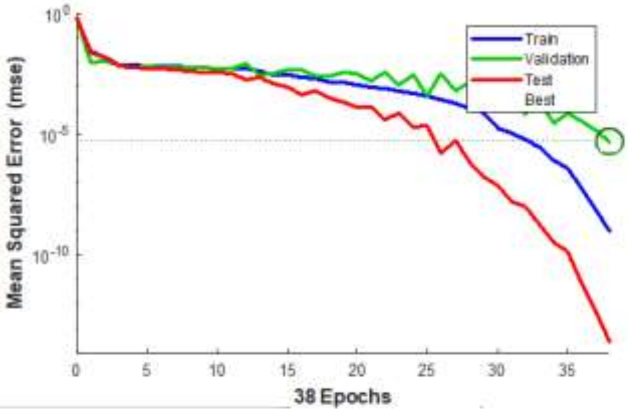

Figure 5: Graph showing the training process using Trainlm in Cascade forward BP network.

The graph in figure 5 shows that it took the validation data up to 35 iterations to detect that the error rate has started falling and it eventually converged.

The value of errors computed for each classifier based on equation 1, determines its accuracy. The classifier model created using Elman BP network technique has an error of 0.0661, while the classifier model created using Cascade forward BP network technique has an error of 0.0928. Each classifier model created based on the two selected techniques is of high accuracy.

6. Conclusions

This study has demonstrated the use of two neural network techniques for the binary classification of separable data. The techniques used are Elman back-propagation and Cascade feed-forward neural network techniques. Although, both techniques have similar configuration settings, however, the evaluation results carried out shows that Elman back-propagation has a better accurate result of 0.0661 error, as against the classifier created using Cascade feed-forward neural network structure that has an error of 0.0928.

It can be concluded, therefore, that artificial neural network techniques is a very effective technique that is capable of fitting linear classifiers from a separable data. The study has also shown the right parameter settings that can give a classifier of better accuracy and a generic result that is free of overfitting.

This work could be extended to cover the implementation of other techniques within the neural network domain, in other to reveal their weaknesses or strength in the classification of separable data..

References

Ahamad, M. V., et al. (2020). Simulation of Learning Logical Functions Using Single-Layer Perceptron *Ambient Communications and Computer Systems* (pp. 121-133): Springer.

Cenate, C. T., et al. (2016). Optimization of the Cascade Feed Forward Back Propagation network for defect classification in ultrasonic images. *Russian Journal of Nondestructive Testing, 52*(10), 557-568.

Chawla, N. V., et al. (2004). Special issue on learning from imbalanced data sets. *ACM SIGKDD explorations newsletter, 6*(1), 1-6.

Databank. (2020). Development indicators showing the mortality rate, neonatal Retrieved January 16,, 2021, from https://databank.worldbank.org/reports.aspx?source=2&series=SH.DYN.NMRT#

Deng, Z., et al. (2019). A model of Double Descent for High-dimensional Binary Linear Classification. *Statistical Machine Learning*.

Dhanalakshmi, P., et al. (2009). Classification of audio signals using SVM and RBFNN. *Expert systems with applications, 36*(3), 6069-6075.

Fernández-Delgado, M., et al. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research, 15*(1), 3133-3181.

Fontenla-Romero, O., et al. (2010). A new convex objective function for the supervised learning of single-layer neural networks. *Pattern Recognition, 43*(5), 1984-1992.

Hossin, M., & Sulaiman, M. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process, 5*(2), 1.

Jujie, W., et al. (2014). Forecasting wind speed using empirical mode decomposition and Elman neural network. *Applied Soft Computing, 23*, 452-459.

Kelo, S., & Dudul, S. (2012). A wavelet Elman neural network for short-term electrical load prediction under the influence of temperature. *International Journal of Electrical Power & Energy Systems, 43*(1), 1063-1071.

Khalid, M., & Singh, J. (2017). Memristive crossbar circuits-based combinational logic classification using single layer perceptron learning rule. *Journal of Nanoelectronics and Optoelectronics, 12*(1), 47-58.

Kowsari, K., et al. (2019). Text classification algorithms: A survey. *Information*. doi: 10.3390/info10040150

Li, H., et al. (2015). *A convolutional neural network cascade for face detection.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Narayanan, A., et al. (2004). Single-layer artificial neural networks for gene expression analysis. *Neurocomputing, 61*, 217-240.

Petters, J. (2020). What is Data Classification? Guidelines and Process. Retrieved February 12, 2021, from https://www.varonis.com/blog/data-classification/

Rajaguru, H., & Prabhakar, S. K. (2017). *Hilbert transform with Elman backpropagation and multilayer perceptrons for epilepsy classification.* Paper presented at the 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA).

Rampisela, T. V., & Rustam, Z. (2018). *Classification of schizophrenia data using support vector machine (SVM).* Paper presented at the Journal of Physics: Conference Series.

Ren, G., et al. (2018). A modified Elman neural network with a new learning rate scheme. *Neurocomputing, 286*, 11-18.

Taud, H., & Mas, J. (2018). Multilayer perceptron (MLP) *Geomatic Approaches for Modeling Land Change Scenarios* (pp. 451-455): Springer.

Walker, J. M. (2010). *Data Mining Techniques for Life Science* (O. Carugo & F. Eisenhaber Eds.): Humana Press.

Wang, S., & Yao, X. (2012). Multiclass imbalance problems: Analysis and potential solutions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42*(4), 1119-1130.

You, Z., et al. (2020). Alzheimer's Disease Classification With a Cascade Neural Network. *Frontiers in Public Health, 8*.

Zhang, T., & Oles, F. J. (2001). Text categorization based on regularized linear classification methods. *Information retrieval, 4*(1), 5-31.