Nigerian Journal of Technology (NIJOTECH)
Vol. 40, No. 5 September, 2021, pp.954 –965
www.nijotech.com

Print ISSN: 0331-8443
Electronic ISSN: 2467-8821
http://dx.doi.org/10.4314/njt.v40i5.21

Application of Neural Network Algorithm for Schizophrenia Diagnosis

J. E. Zaccheus[1], O.P. Fidelis[2,*], E.O. Nwoye[3]

[1]*Department of Biomedical Engineering, Afe Babalola University, Ado-Ekiti, Ekiti State, NIGERIA*
[2]*Department of Biomedical Technology, Federal University of Technology, Akure, Ondo State, NIGERIA*
[3]*Department of Biomedical Engineering, University of Lagos, Lagos State, Nigeria*

Abstract
Schizophrenia is a prolonged mental condition that affects functional impairment in work, interpersonal relationships, and self-care. This research was aimed at developing a neural network model to diagnose schizophrenia using text data acquired from patients' records. The model was developed from datasets obtained from Neuropsychiatric Hospital in Yaba and the Lagos University Teaching Hospital, both in Lagos, Nigeria, using Python programming language and is provided with significant features from data sets to learn patterns within the training data and perform classification on the test data. The results show that the model produced a test accuracy of 85%, specificity of 95% and a precision of 93%. These results indicate that the model can be used for effective computer-aided diagnosis of schizophrenia.

Keywords: Artificial intelligence; medical records; mental health; neural network; schizophrenia

1.0 INTRODUCTION

Schizophrenia is a chronic and severe mental disorder affecting 20 million people worldwide and typified by delusions, hallucinations and other cognitive difficulties [1]. It is a mental disorder that frequently emerges in late adolescence or early adulthood and may affect educational and occupational performances [2]. The definitive cause of schizophrenia is unknown; however, a combination of environmental and psychosocial factors may be responsible. The diagnosis of schizophrenia is complex and it takes a number of processes to arrive at the right decision. Psychiatrists have to watch that the observed symptoms are persistent for at least a month as contained in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) classification details and that those symptoms are not as a result of other related diseases nor a behavioral activity (e.g. alcohol consumption) nor a result of a pharmacological side effects. These requirements have made the process of diagnosis lengthy and complex. Psychiatrists need a quicker, effective, and knowledge-based technique to diagnose Schizophrenia for early detection, prompt treatment and effective management of the disease [3].

***Corresponding author **(Tel:** +234 (0) 8038193799**)**
Email addresses: zaccheusje@abuad.edu.ng (Zaccheus E. Jesuloluwa), opfidelis@futa.edu.ng (Fidelis P. Obinna), enwoye@unilag.edu.ng (Nwoye O. Ephraim)

For diagnosis to be verified, a combined effort between the psychiatrist, patient and patient's relation or neighbor is required. The process of obtaining information is sometimes demanding and time-consuming because some patients may not recall some details of their past experiences. The DSM-V has outlined some criteria that can be used to diagnose schizophrenia vis-a-viz delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior, negative symptoms (reduced emotional expression or lack of motivation) – two or more of which must be present during a 1-month period [4]. In addition, the International Classification of Diseases, Eleventh Edition (ICD-11) outlines some criteria for diagnosing schizophrenia including positive symptoms such as persistent delusions, persistent hallucinations, disorganized thinking (typically manifest as disorganized speech), grossly disorganized behavior, and experiences of passivity and control, and negative symptoms such as blunted or flat affect and avolition, and psychomotor disturbances [5]. Diagnosis is not only time consuming but also risky for the patient. Therefore, Psychiatrists and Doctors are desperate for scientists to develop quicker ways of diagnosing the condition as early diagnosis can reduce the cost of treatment [6].

Currently, detecting schizophrenia includes the subjective evaluation of a patient's test results and mental health record; although symptoms overlay with other mental illnesses can take place [7], increasing the

probability of misdiagnosis. The condition has no well-established and standard biomarker, though studies [8, 9] indicate that Magnetic Resonance Imaging (MRI) may be an efficient biomarker for Schizophrenia. The challenge with utilizing MRI data to diagnose schizophrenia based on structural changes is the overlap in structural change brought on by factors related with schizophrenia such as alcoholism and anti-psychosis medication [10]. Aside medical imaging, computer-aided diagnosis which is powered by deep learning is gaining popularity in modern medicine. Deep learning which is categorized as a subfield of machine learning can isolate features and carry out independent classification [11], and has multiple range of application including applications in medicine and healthcare. It can be used for prognosis and optimizing treatment and represents a possible solution for improved diagnosis and management of schizophrenia.

One study [12] utilized data mining approaches for Genome-wide Association of Mood Disorders. Six classifiers namely Bayesian Network, Support Vector Machine, Logistic Regression, Random Forest, Radial Basis Function and Polygenic Scoring Approach were compared. It was identified that simple polygenic score classifier performed better than others and it was also found that all the classifiers performed worse with small number of Single Nucleotide Polymorphisms in the brain expressed set compared to whole genome set.

In another study [13] a neural network model with an accuracy of 82.35% for predicting the likelihood of developing psychological conditions such as anxiety, behavioral disorders, depression and post-traumatic stress disorders was developed. A similar study [14] proposed a mental health diagnostic expert system to assist the Psychologists in diagnosing and treating their mental patients. Three artificial intelligence techniques: Rule-Based Reasoning, Fuzzy Logic and Fuzzy-Genetic Algorithm were used for diagnosis and suggestion of treatment plans. The study in [15] developed a Bayesian network (BN) decision model for the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. The BN model was taken into account because it is appropriate for indicating uncertainty and causality. A supervised learning algorithm was used to assess the network using a dataset of real clinical cases. Sensitivity analysis was used to evaluate the model and it stands out when compared to most of the other well-known classifiers.

The authors of a previous study [16] investigated potential linguistic markers of schizophrenia using the tweets of self-identified schizophrenia patients, and illustrate various natural language processing methods to analyze the language. The study observed how these signals compare with the commonly used linguistic inquiry and word count categories for understanding mental health and provide preliminary evidence of additional linguistic signals that may aid in identifying and getting help to people suffering from schizophrenia.

In [17] the authors state that prominent formal thought disorder, expressed as unusual language in speech and writing, is often a central feature of schizophrenia. Thirty-six patients with DSM-V criteria chronic schizophrenia provided a page of writing (300-500 words) on a designated topic. Writing was examined by automated text categorization and compared with non-psychiatrically ill individuals, investigating any differences with regards to lexical and syntactical features. Computerized methods used included extracting relevant text features, and using Machine Learning techniques to induce mathematical models distinguishing between texts belonging to different categories. Observations indicated that automated methods distinguish schizophrenia writing with 83.3% accuracy.

Results reflect underlying impaired processes including semantic deficit, independently establishing connection between primary pathology and language. The studies which had been surveyed indicate that there are different machine learning techniques utilized by researchers for the diagnosis of schizophrenia using different forms of data. The present study is aimed at developing a model that is optimized for schizophrenia diagnosis using text data and stands out among other previously developed networks.7

2.0 MATERIALS AND METHOD

2.1 Materials

The following items were deployed for the development of the model.

- Anaconda Environment – It is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing and predictive analytics). Anaconda houses numerous environments, some examples include: Jupyter notebook, Spyder, Jupyter Lab, Orange 3, Qt Console, Visual Studio Code and R Studio.
- Spyder IDE was used for writing and compiling code during this research.
- Python 3.7 – Python is an interpreted, high-level, general-purpose programming language.
- Medical records were obtained from Lagos University Teaching Hospital (LUTH) and Federal Neuropsychiatric Hospital Yaba, Lagos State.

2.2 Data Acquisition

Data used for this study were collected from the

psychiatric clinic of the Lagos University Teaching Hospital, Lagos (Approval number: ADM/DCST/HREC/APP/3263) and the Federal Neuropsychiatric Hospital, Yaba, Lagos (Approval number: FNPH/HREC/19/20). The various features from the patient case files were collated in a Microsoft excel file. One hundred and ninety-eight (198) health records were acquired for the study. The dataset consists of health records of patients reported between 2013 and 2019 and includes patients diagnosed with schizophrenia as well as those diagnosed with related illnesses. The inaccessibility of electronic health record was a huge challenge because data entry is a tedious and time consuming task. The dataset has 38 attributes including the CLASS column. However, only attributes based on DSM-V and ICD-11 specifications for diagnosis of schizophrenia were used in training and validating the model and as shown in Table 1.

Table 1: Shows the representation of some of the features in the dataset, description and values

No	FEATURE	DESCRIPTION	VALUES
1	YEAR	Year patient reported in hospital	2013 - 2019
2	AGE	Age of patient	16 - 78
3	SEX	Sex of patient	MALE, FEMALE
4	OCCUP_HX	Occupation History	Unemployed, Occupation
5	MAR_STA	Marital Status	Married, Single, divorced, Widow
6	DUR_EPIS	Episode Duration(length of time the patient has suffered symptoms)	Time in months
7	P_PXY_HX	Past psychiatric history	No, Yes, Rape, Mental illness, Grief
8	P_MED_HX	Past medical history	No, Disease suffered
9	FAM_P_HX	Family Psychiatric History	Yes, No
10	P_SOC_HX	Past social history	Yes, No
11	P_SEX_HX	Past sexual history	Normal, Masturbation
12	FOR_HX	Forensic history	Yes, No
13	PREMOB_HX	Premorbid History	Normal, Introvert, Extrovert, Melancholic
14	MSE	Mental State Examination	Kempt, Unkempt, Poor eye contact, Restless
15	SPEECH	Speech Status	Normal, Reduced volume, Mute, Slurred, Decreased tone, Irrelevant, Incoherent
16	TH_CONTENT	Thought content at time of report	Persecutory delusion, Auditory hallucination, Grandiose delusion.

2.3 Data preprocessing

Data acquired from hospitals are mostly unstructured and require preprocessing. The first step of preprocessing the acquired dataset is sorting out empty entries found in the dataset. Missing values in dataset can occur during data extraction or collection and can lead to wrong prediction or classification. The data set contains more of text data, and therefore the method of finding the mean and mode of the column would not suffice in this case. The missing values are replaced with values that have the highest frequencies. Next, text data is converted to numeric format to be processed by the model. This is where 'label encoding' is utilized. Label Encoder is a class in the *Sci-kit-learn* library that simply converts values in a specified column into numbers. The *fit_transform* class is used to fit the encoded columns into the original dataset. Scaling features in the dataset before feeding into the model is necessary. Feature Scaling is used to scale the features in a dataset to a range which is centered on zero. The Standard Scaler class in the Sci-Kit Learn library transforms data to have a mean of zero and a standard deviation of 1. This is done so that the variance of the features is in the same range. If a feature's variance is orders of magnitude more than the variance of other features, that particular feature might dominate other features in the dataset, which will reduce the accuracy in the model. Using the *Sci-kit-learn* library, the entire dataset is split into the test and training data at different test/train ratio (60/40, 50/50, 65/35, 40/60); the training data was used for training the samples in the data set, while the test data was used to evaluate the accuracy of the trained model.

2.4 Feature Selection

In prediction or classification models, features that are relevant to the diagnosis case are to be used for training and testing the model. Feature selection improves the model accuracy by removing irrelevant and redundant features. All features are first fed into the network for trials and then relevant features for the model were selected. If

feature selection is not done, the resulting model will likely become more complex and less accurate. Features were selected based on factors that are important in the diagnosis of schizophrenia based on the DSM-V diagnostic criteria for schizophrenia diagnosis.

2.5 Classifier Development

The schizophrenia classifier has 1 input layer, 2 hidden layers and 1 output layer. The programming language used in developing the deep learning model is Python. The input layer is also called the source nodes because it supplies the input data into the network. The number of nodes in the input layer is determined by the number of input features in the dataset. The hidden layer is the layer between the output and the input layers; the hidden layer produces an output based on the activation function used in each hidden layer. The Rectified Linear Unit (ReLU) activation function is used in the hidden layers of the schizophrenia classifier. The adam optimizer was used with the default learning rate.

2.6 RELU Activation Function

The RELU activation function adapted from a previous study [18] can be represented using this *if statement*:

$$\left\{\begin{array}{c} if\ input\ value > 0 \\ return\ input\ value \\ else: \\ return\ 0 \end{array}\right\}$$

This indicates that the input value is returned if it is greater than zero, while the input value is returned as zero if otherwise. The RELU activation function provides a quick computing of values which reduces the time used in training.

2.7 Sigmoid Activation Function

The sigmoid activation function is used in the output layer because it ranges from 0 to 1. Therefore, it is useful in a model like this that is used for predicting the probability of an output. During back-propagation, the derivative of loss with respect to the parameter is calculated. The derivative of the sigmoid function in equation (1) is shown in equation (2) below as adapted from a previous study [19]:

$$f(x) = \frac{1}{1+e^{-x}} \tag{1}$$

$$f'(x) = \frac{\partial f(x)}{\partial x} = -\frac{1}{(1+e^{-x})^2}(-e^{-x}) \tag{2}$$

$$= \frac{1}{1+e^{-x}}\left(1-\frac{1}{1+e^{-x}}\right) \tag{3}$$

$$= y(1-y) \tag{4}$$

$$where\ y = \frac{1}{1+e^{-x}} \tag{5}$$

2.8 Epoch

The epoch represents the point when the entire dataset is passed forward and backward through the neural network only once. The epochs used in training the network were 100, 250 and 500 epochs. The number of epochs was selected randomly, but 500 epochs was the most preferred option.

2.9 Batch Size

Passing the entire dataset into the network at once is inefficient; therefore, the data set can be divided into smaller batches. The batch size indicates the total number of samples present in a single batch. The batch size is selected based on the amount of data samples present in the dataset. The batch sizes used in this research are 16 and 128.

2.10 Dropout

Dropout is a regularization method where randomly selected neurons or nodes are dropped out depending on the probability value selected. The probability $p = 0.2$ was selected in the model. Dropout technique is used to prevent overfitting in deep learning models [20]. The model structure of the classifier is shown in Table 2.

2.11 Mathematical Formulation of the Algorithm

Below is the detailed sample network architecture with a typical 2 inputs, 2 hidden layers and one output layer for the mathematical analysis of the computing processes of network. An overview of the inputs and weights in the layers of the neural network classifier as well as a possible output, y, is shown in Figure 1.
From Figure 1,

$$y = f.(\sum_{j=0}^{n} w_{kj}x_j + b) \tag{6}$$

Where f = activation function
x_j = input features (e.g. age, occupation etc.)
w_{kj} = weights (the connections between each node)
b = bias

The input of each neuron is calculated by multiplying each input feature by the corresponding weight.

Table 2: Model Structure

LAYER	DESCRIPTION
Input layer	The neurons in this layer contain 16 features fed into the network
First Hidden Layer	This layer has 8 neurons with a RELU activation function
Dropout Layer 1	This layer includes a dropout probability value of $p = 0.2$
Second Hidden Layer	This layer also has 8 neurons with a RELU activation function.
Dropout Layer 2	This layer includes a dropout probability value of $p = 0.2$.
Output Layer	The output layer has a sigmoid activation function for classification.

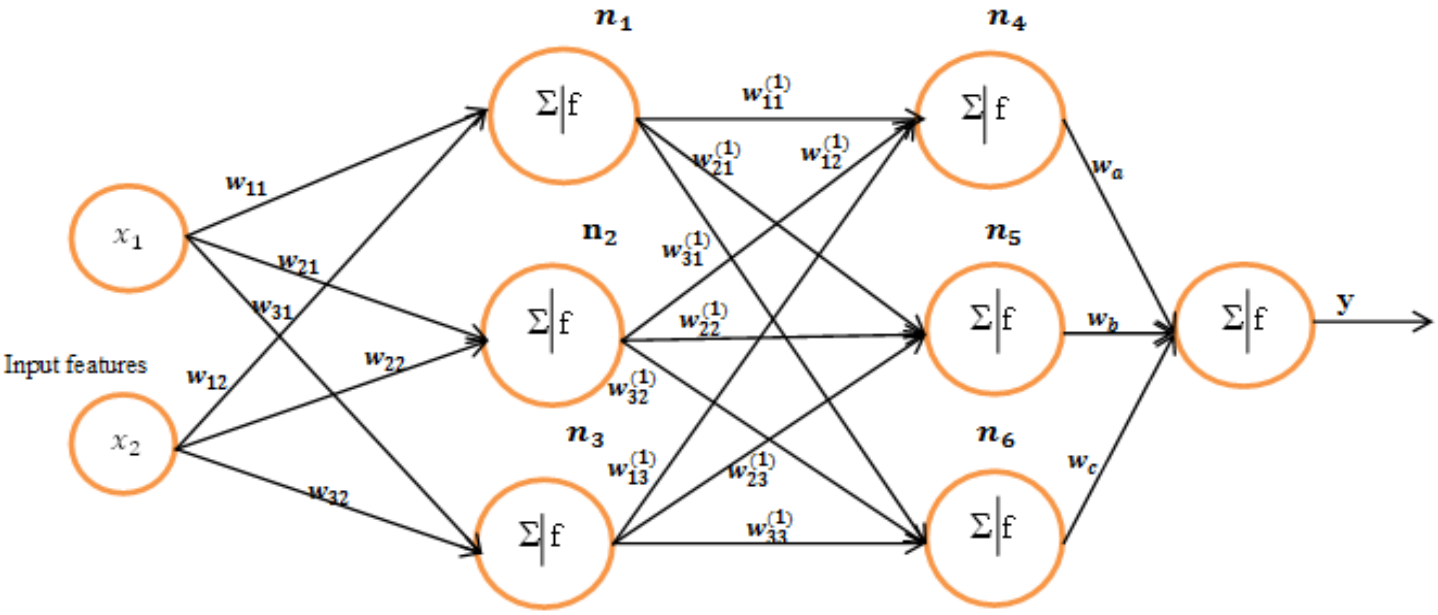

Figure 1: An overview of the inputs and weights in the layers of the neural network classifier

For the neuron n_1:

$$n_1 = f(w_{11}x_1 + w_{12}x_2 + b) \quad (7)$$

For the neuron n_2:

$$n_2 = f(w_{21}x_1 + w_{22}x_2 + b) \quad (8)$$

For the neuron n_3:

$$n_3 = f(w_{31}x_1 + w_{32}x_2 + b) \quad (9)$$

The activation function for the two hidden layers is a *ReLU* function; the *ReLU* function is represented mathematically as,

$$f(x) = \max(0, x) \quad (10)$$

Therefore, the output at n_1, n_2 and n_3 is given by:

$$n_1 = \max(0, (w_{11}x_1 + w_{12}x_2 + b) \quad (11)$$

$$n_2 = \max(0, (w_{21}x_1 + w_{22}x_2 + b)) \quad (12)$$

$$n_3 = \max(0, (w_{31}x_1 + w_{32}x_2 + b)) \quad (13)$$

For the neuron n_4:

$$n_4 = f\left(n_1 w_{11}^{(1)} + n_2 w_{12}^{(1)} + n_3 w_{13}^{(1)} + b\right) \quad (14)$$

For the neuron n_5:

$$n_5 = f\left(n_1 w_{21}^{(1)} + n_2 w_{22}^{(1)} + n_3 w_{23}^{(1)} + b\right) \quad (15)$$

For the neuron n_6:

$$n_6 = f(n_1 w_{31}^{(1)} + n_2 w_{32}^{(1)} + n_3 w_{33}^{(1)} + b) \quad (16)$$

Therefore, the output at n_4, n_5 and n_6 is given by:

$$n_4 = \max\left(0, (n_1 w_{11}^{(1)} + n_2 w_{12}^{(1)} + n_1 w_{13}^{(1)} + b)\right) \quad (17)$$

$$n_5 = \max\left(0, (n_1 w_{21}^{(1)} + n_2 w_{22}^{(1)} + n_3 w_{23}^{(1)} + b)\right) \quad (18)$$

$$n_6 = \max\left(0, (n_1 w_{31}^{(1)} + n_2 w_{32}^{(1)} + n_3 w_{33}^{(1)} + b)\right) \quad (19)$$

The total output of the model has an activation function (sigmoid) and is given by:

$$y = (n_4 w_a + n_5 w_b + n_6 w_c) \quad (20)$$

$$\sigma = \frac{1}{1 + e^{-y}} \quad (21)$$

$$y = \frac{1}{1 + e^{-(n_1 w_a + n_2 w_b + n_3 w_c)}} \quad (22)$$

2.12 Confusion Matrix

A confusion matrix was used to evaluate the performance of the model by calculating the accuracy, specificity, precision, recall and F1 score of the model from the matrix. The confusion matrix was obtained from the test dataset which is 40% of the total dataset.

3.0 RESULTS AND DISCUSSION

In this section, the train accuracy and test accuracy of the model is presented. This was performed on different train/test split ratios and different number of features. Graphs and tables are presented, explained and interpreted to understand how the datasets performed during the experiment using the developed model.

3.1 25 Features

The 25 features used in this case consist of 24 input features and one output feature. The 24 features do not include the following columns (YEAR, DIAGN, MEM_LT, FOR_HX, PREMORBID, INT_S_A_D, INT_PROV, PSE, ATTEN); the output feature is the CLASS column.

Table 3: Result of Train and Test Accuracy using 25 features from the dataset

Train/Test split	**65/35**	**35/65**	**50/50**	**60/40**	**40/60**
Train Accuracy	98.44 %	98.55 %	96.97 %	97.46 %	98.73 %
Test Accuracy	78.57 %	85.27 %	78.97 %	83.75 %	78.15 %

From Table 3, using 25 features from the data set the highest training accuracy attained was at the 40/60 split ratio while the highest test accuracy recorded was at the 35/65 split ratio. The lowest performance was recorded at the 40/60 split ratio. Also the highest percentage difference between the training and test accuracy was recorded at the 40/60 test/train split ratio The average accuracy of the model was not optimal when using 25 features due to overfitting; therefore experiments were carried out on fewer numbers of features.

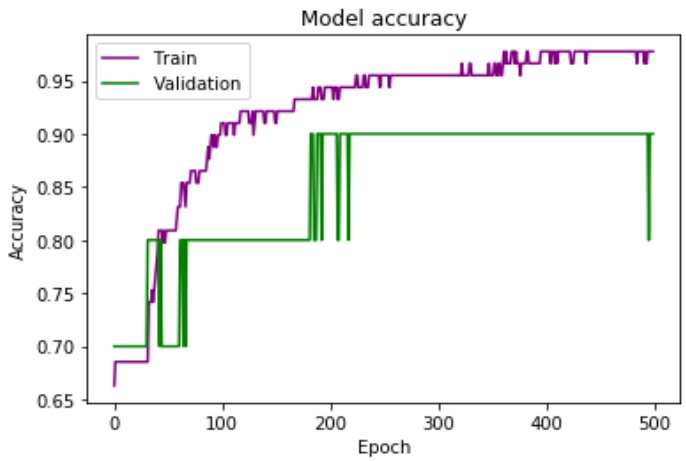

Figure 2: Model accuracy for 25 features at 50/50 Train-Test split.

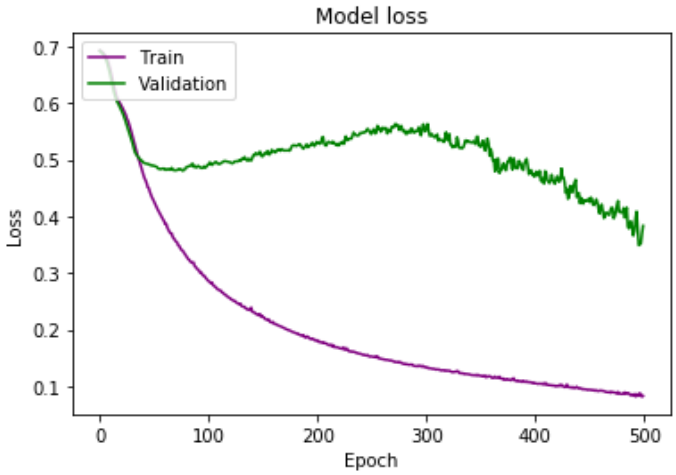

Figure 3: Model loss for 25 features at 50/50 Train-Test split.

Figures 2 and 3 show the training accuracy increased from 0.85 at 100 epochs to 0.97 after 500 epochs; the validation accuracy increased to 0.9 at 300 epochs but after 500 epochs it reduced to 0.8. The graphs of both the training and validation accuracy and the training and validation loss plotted indicate misclassification of the data samples.

From Figure 6 and 7, it is shown that the validation accuracy was constant at 0.92 over 500 epochs while the training accuracy increased from 0.9 to 0.99 between 0 and 500 epochs. The training and validation loss did not reduce in the same pattern thus indicating overestimation of the training data and misinterpretation of the test data.

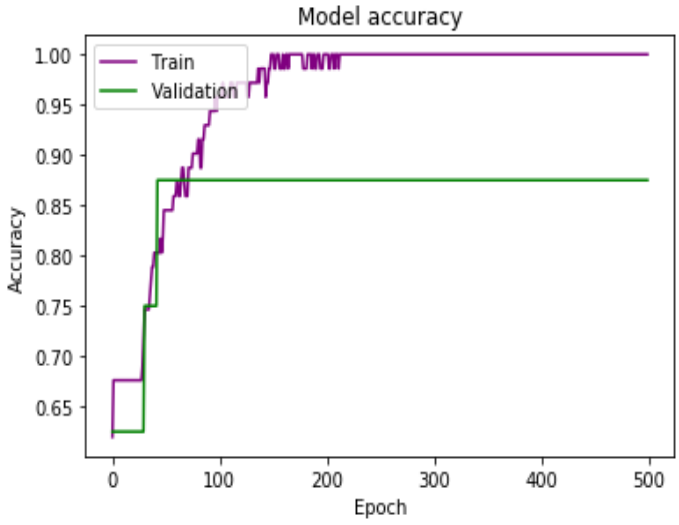

Figure 4: Model accuracy for 25 features at 40/60 Train-Test split.

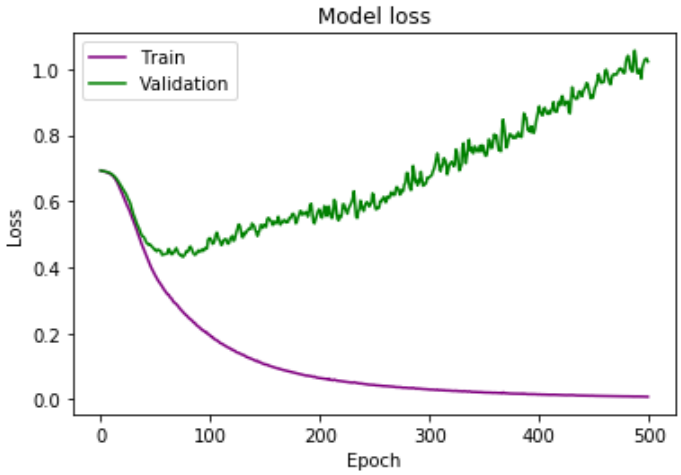

Figure 5: Model loss for 25 features at 40/60 Train-Test split.

Figures 4 and 5 indicate that both the validation loss and accuracy increased over 500 epochs; this suggests misdiagnosis of the model on the test dataset. The training accuracy recorded was at 0.84 and 1 between 100 and 200 epochs respectively while the training accuracy remained constant after 500 epochs.

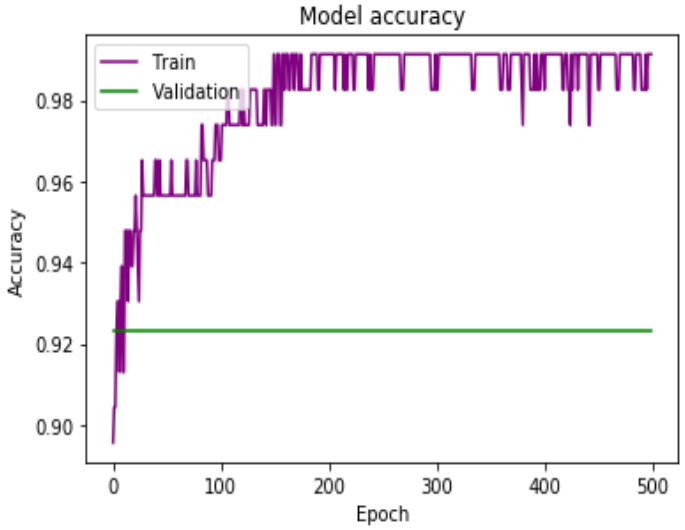

Figure 6: Model accuracy for 25 features at 65/35 Train-Test split.

3.2 17 Features

The 17 features used in this case consist of 16 input features and 1 output feature. The features selected include (DUR_EPIS, MSE, ATTEN, EEG, P_PSY_HX, FAM_P_HX, FOR_HX, PERCEP, TH_CONTENT, SPEECH, JUDGM, MEM_ST, MEM_LT, MOOD, AFFECT, ORIENT and CLASS).

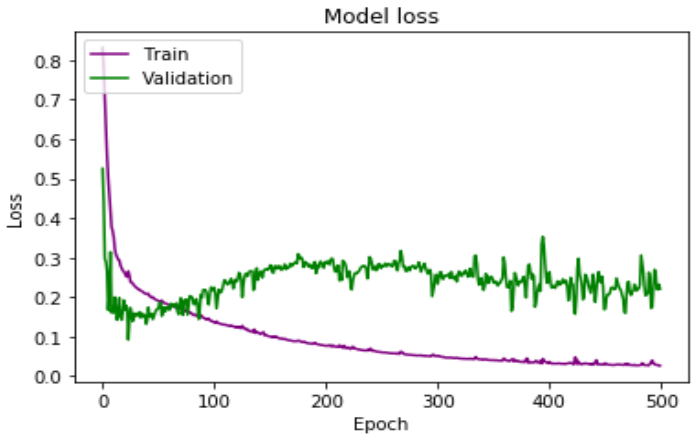

Figure 7: Model loss of 25 features at 65/35 Train-Test split

The results in Table 4 show the 65/35 train-test-split achieved the highest test accuracy of 82.86%. The 90/10 train-test-split had the lowest test accuracy of 75%. It indicates that the model was over-trained and could not generalize on unseen data, which led to over-fitting. The highest test accuracy was attained when the data set was split at a 65/35 train/test ratio. generally, when 17 features was used in training the model, overfitting reduced in comparison to when the model was trained on 32 and 25 features.

Table 4: Result of Train and Test Accuracy using 17 features from the dataset

Train/Test split	**65/35**	**90/10**	**50/50**	**60/40**	**40/60**
Training Accuracy	87.5%	86.52%	89.9%	89.91%	87.34%
Test Accuracy	82.86	75.00 %	80.81%	82.5%	82.35%

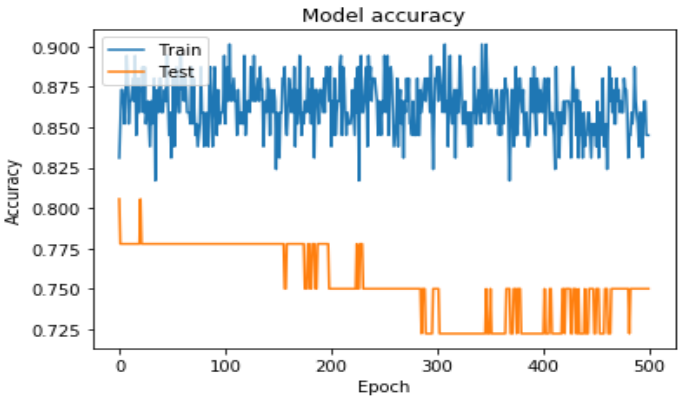

Figure 8: Model accuracy for 17 features at 90/10 Train-test split.

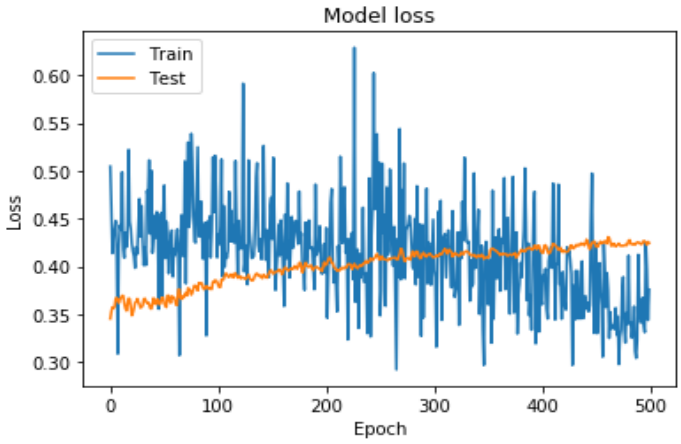

Figure 9: Model loss for 17 features at 90/10 Train-test split.

From Figures 8 and 9, the training accuracy of the model is high, but the validation accuracy is significantly lower which indicates overfitting. The loss was fluctuating and the validation loss was increasing, this indicates that the model was cramming values and not learning.

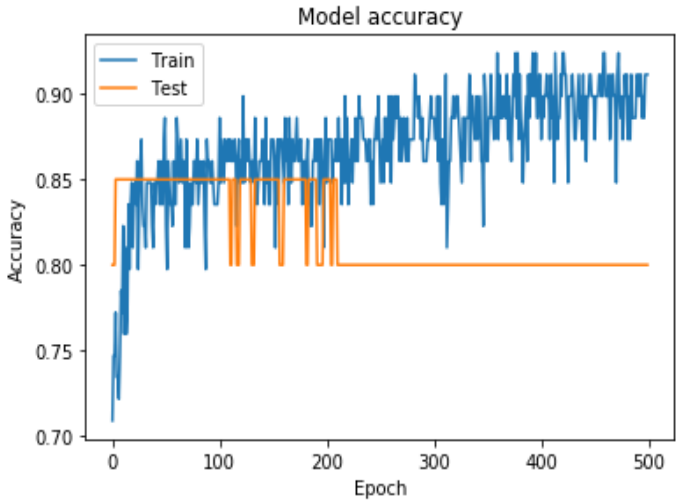

Figure 10: Model accuracy at 17 features and 50/50 Train-test split.

From Figure 10 and 11, the training accuracy increased from 0.7 and peaked at 0.92 over 500 epochs. There was a little overfitting in this model compared to other train/test split ratios. The model loss graph indicates that the model learnt at a normal rate. The training loss was lowest after 500 epochs at 0.2.

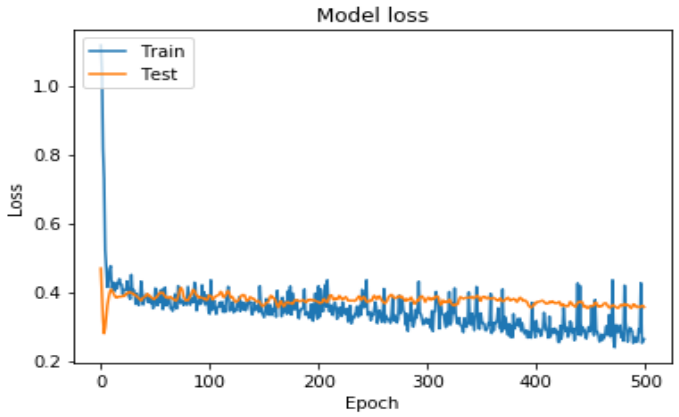

Figure 11: Model loss at 17 features and 50/50 Train-test split.

Figure 12 and 13 shows the training loss and validation loss were not declining at the same rate. The training loss was fluctuating between 0.25 and 0.6 which indicates overfitting. The validation accuracy was the highest after 500 epochs at 0.825 while the training accuracy recorded was greater than the validation accuracy and it peaked at 0.9 after 300 epochs.

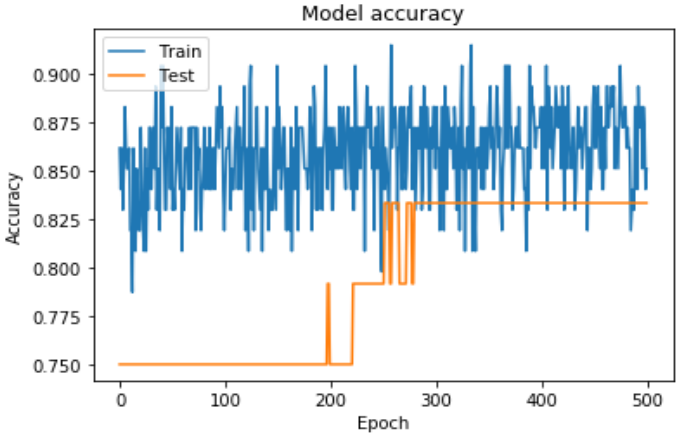

Figure 12: Model loss at 17 features and 60/40 Train-test split.

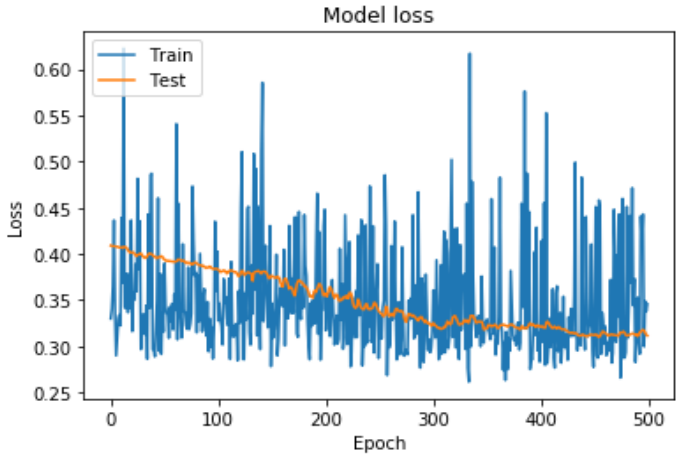

Figure 13: Model loss at 17 features and 60/40 Train-test split.

3.3 *Confusion Matrix of the Neural Network Classifier*

The model with the optimal performance was selected from the other models. The confusion matrix of said model is given below, and also the accuracy, specificity, recall, precision and F- measure are calculated from the confusion matrix.

True Positive (TP): The model predicted positive and it's true.

True Negative (TN): The model predicted negative and it's true.

False Positive (FP): The model predicted positive and it's false.

False Negative (FN): The model predicted negative and it's false.

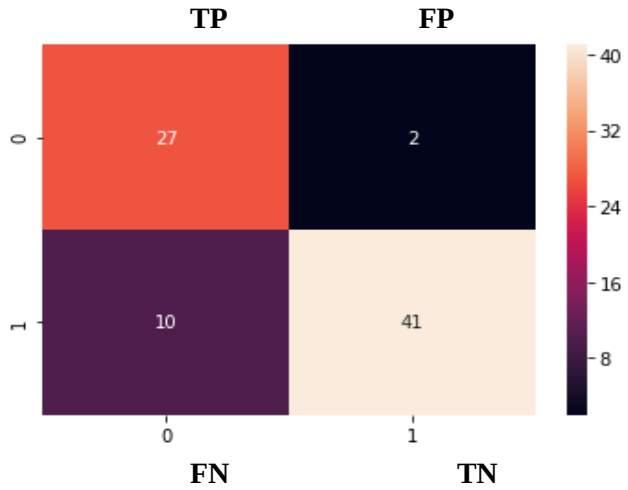

Figure 14: Confusion matrix of the test dataset from the model

From Figure 14, it is deduced that:

a. The model predicted 27 schizophrenic test samples and it is true.
b. The model predicted 2 schizophrenic test samples and it is false.
c. The model predicted 41 non-schizophrenic test samples and it is true.
d. The model predicted 10 non-schizophrenic test samples and it's false.

From the confusion matrix in Figure 14, the following values are calculated

$$Specificity = \frac{TN}{TN+FP} = \frac{41}{41+2} = \frac{41}{43} = 95.4\% \quad (23)$$

From equation (29) the model had a specificity of 95%

$$Recall = \frac{TP}{TP+FN} = \frac{27}{27+10} = \frac{27}{37} = 72.9\% \quad (24)$$

From equation (30) the model had a recall of 72.9%

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{27+41}{27+41+10+2} = 85\% \quad (25)$$

From equation (31) the model was 85% accurate

$$Precision = \frac{TP}{TP+FP} = \frac{27}{27+2} = 93.1\% \quad (26)$$

From equation (32) the model was 93% precise

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$= 2 \times \frac{93.1 \times 72.9}{93.1 + 72.9} = 81.77\% \quad (27)$$

The $F1\ score$ of the model was 82% approximately

From Table 5, the highest accuracy recorded was produced by the developed model with drop out nodes; although it had a recall value of 72.9% the model outperformed other models in areas of specificity and precision. The developed model outperforms the decision tree classifier with an F1 score of 81.77% to that of the decision tree at 73.9%. Generally the dropout technique greatly improved the precision and accuracy of the model when compared specifically to the developed model without dropout nodes.

Table 5: Results showing different classifiers and their performance

Classifier	Accuracy	Precision	Recall	Specificity	F1 score
CLASSISCHIZ	82.5%	75.9%	**75.9%**	86.3%	75.9%
CLASSISCHIZ*	**85%**	**93.1%**	72.9%	**95.4%**	**81.77%**
Decision Tree	80.8%	72.9%	75%	84%	73.9%

*Represents that dropout technique was used in the model

Figure 15 represents the accuracy of the model developed. It is a plot of the epoch versus the accuracy. After 100 epochs, the training accuracy was 65%, the training accuracy increased to 77% after 250 epochs. The training accuracy peaked at 90% at 350 epochs, and then at 500 epochs the training accuracy recorded was 86%. The average training accuracy recorded was 88.5%.

Figure 16 represents the training and validation loss of the model. The plot indicates that has the loss is decreasing the model is actually learning the patterns between the features of the dataset. The loss function was at the lowest at 500 epochs as shown in the graph.

The performance of the model for different configurations shows the accuracy of the model and the ability of the model to be used as reliable software for schizophrenia diagnosis. During the study, it was observed that the model performed best with 17 features and has the least degree of overfitting. The 10/90 and 90/10 train/test split produced the greatest overfitting and is therefore not recommended as a reliable configuration. However the 60/40 and 50/50 train/test split produced the best train accuracies with the least overfitting and was considered as a reliable configuration to be selected, the 60/40 paradigm performed better than the 50/50 configuration.

Figure 15: Training and Validation accuracy of the Model using Dropout nodes

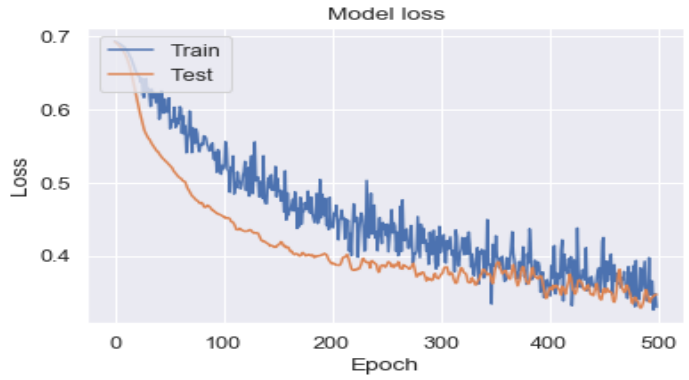

Figure 16: Training and Validation Loss of the Model using Dropout nodes

Overfitting was a major issue in most of the model configurations tested, this was due to irregular train/test split ratio and small datasets. This was solved by tuning hyper-parameters and using the dropout regularization technique. The normal 60/40 paradigm was improved and optimized by introducing the dropout technique. The dropout technique improved the accuracy of the model by 2.5%, the precision by 17.2% and the F1 score by 5.9%. Both models were compared with the decision tree classifier as shown in Table 6; the developed models emerged the best. Finally, from performed experiments it was observed that some features in the data set are redundant and does not improve the accuracy of the model.

4.0 CONCLUSION

In this study, a neural network model was developed for schizophrenia diagnosis and was tested and validated. The model outperformed existing neural network models as shown in the research. In relation to medical application, the model can aid psychiatrists in quick and accurate detection of schizophrenia and also in designing treatment plans and the management of the condition as duration of psychotic or manic episode is a factor in the effectiveness of antipsychotics. The model can be tuned to work on smartphones, to improve accessibility while the accuracy of the model can be improved greatly by acquiring more data from other psychiatric hospitals.

REFERENCES

[1] James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abdelalim, A. "Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017", *The Lancet*, 392(10159), (2018), 1789–1858. Doi: 10.1016/s0140-6736(18)32279-7

[2] Schizophrenia (2015, March 22) https://www.medicalnewstoday.com/articles/36942.php Accessed on 9/12/19

[3] Koch S.P., Hägele C., Haynes J-D, Heinz A., Schlagenhauf F., Sterzer P. "Diagnostic Classification of Schizophrenia Patients on the Basis of Regional Reward-Related fMRI Signal Patterns", *PLoS ONE* 10(3), (2015), e0119089. doi: 10.1371/journal. pone.0119089

[4] Diagnostic and Statistical Manual of Mental Disorders: DSM-5, 5th ed., American Psychiatric Association, 2013. DSM-V. doi.org.db29.linccweb.org/10.1176/appi.books

[5] World Health Organization. (2018). International classification of diseases for mortality and morbidity statistics (11th Revision). Retrieved from https://icd.who.int/browse11/l-m/en

[6] Mihalopoulos, C., Harris M., Henry, L., Harrigan, S. and McGorry, P. (2009). "Is early intervention in psychosis cost-effective over the long term?" *Schizophrenia Bulletin*, 35(5), (2009), 909–918.

[7] Owen, M. J., Sawa, A., and Mortensen, P. B. "Schizophrenia", *The Lancet*, 388(10039), (2016,) 86–97.

[8] Han, S., Huang, W., Zhang, Y., Zhao, J., & Chen, H. (2017). "Recognition of early-onset schizophrenia using deep-learning method", *Applied Informatics*, 4(1), (2017), 16.

[9] Hsieh, T.-H., Sun, M.-J., & Liang, S.-F. "Diagnosis of schizophrenia patients based on brain network complexity analysis of resting-state fMRI" In the 15th International conference on Biomedical Engineering, New York, 2014, 203–206.

[10] Bois, C., Whalley, H., McIntosh, A., & Lawrie, S. "Structural magnetic resonance imaging markers of susceptibility and transition to schizophrenia: A review of familial and clinical high risk population studies", *Journal of Psychopharmacology*, 29(2), (2015), 144–154.

[11] Amin, J., Sharif, M., Yasmin, M., & Fernandes, S. L. "Big data analysis for brain tumor detection: Deep convolutional neural networks", *Future Generation Computer Systems*, 87, (2018), 290-297.

[12] Sumathi, M.R. & Poorna, B. "Design and development of ensemble of naïve bayes classifiers to predict social and communication deficiency among children", *International Journal of Applied Engineering Research*, 12, (2017), 14190-14198.

[13] Dabek, Filip, and Jesus J. Caban. "A Neural Network Based Model for Predicting Psychological Conditions", *Brain Informatics and Health*, Springer International Publishing, 2015, 252-261.

[14] Masri, R.Y.; Jani, H.M., "Employing artificial intelligence techniques in Mental Health Diagnostic Expert System," in *Computer & Information Science (ICCIS), 2012 International Conference on*, 1, 12-14 June 2012,495-499. doi:10.1109/ICCISci.2012.6297296

[15] Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. "Do we need hundreds of classifiers to solve real world classification problems?", *Journal of Machine Learning Research*, 15, (2014), 3133–3181.

[16] Mitchell M, Hollingshead K, and Coppersmith G. "Quantifying the language of schizophrenia in social media", in *Proc. 2nd Work- shop Comput. Linguistics Clin. Psychol., Linguistic Signal Clin. Reality*, Denver, CO, USA, Jun. 2015, pp. 11_20. [Online]. Available: http://www.aclweb.org/anthology/W15-1202

[17] Strous R. D., Koppel M., J. Fine, S. Nachliel, G. Shaked, and A. Z. Zivotofsky, "Automated characterization and identification of schizophrenia in writing", *J. Nervous Mental Disease*, 197(8), (2009), 585-598.

[18] Nair, Vinod; Hinton, Geoffrey E. "Rectified Linear Units Improve Restricted Boltzmann Machines", *27th International Conference on International Conference on Machine Learning*, USA, Omnipress, 2010, 807–814

[19] Von Seggern, D. *CRC Standard Curves and Surfaces with Mathematics, 2nd ed.* Boca Raton, FL: CRC Press, 2007.

[20] Srivastava, Nitish & Hinton, Geoffrey & Krizhevsky, Alex & Sutskever, Ilya & Salakhutdinov, Ruslan. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". *Journal of Machine Learning Research*, 15, (2014), 1929-1958.