ISSN 2563-7568

RESEARCH ARTICLE

Detecting Hidden Patterns in EEG Waveforms of Schizophrenia Patients using Convolutional Neural Network

Ephraim Nwoye[1*], Obinna Fidelis[2], Kenneth Umeh[1], Babatunde Fadipe[3], Charles Umeh[3], Theddeus Akano[4], Zaccheus Jesuloluwa[5], Wai Lok Woo[6]

[1]*Department of Biomedical Engineering, University of Lagos, Nigeria*
[2]*Department of Biomedical Technology, Federal University of Technology, Akure, Nigeria*
[3]*Department of Psychiatry and Clinical Psychology, Lagos University Teaching Hospital, Nigeria*
[4]*Department of System Engineering, University of Lagos, Nigeria*
[5]*Department of Biomedical Engineering, Afe Babalola University, Ado-Ekiti, Nigeria*
[6]*Department of Computer and Information Sciences, Northumbria University, United Kingdom*

Abstract

Schizophrenia is a severe mental disorder that affects 1% of the world's population and it is characterized by behavioral symptoms such as delusions, hallucinations and disorganized speech. The aim of this research was to develop an artificial intelligence model to detect hidden patterns in electroencephalogram (EEG) waveforms of schizophrenia patients. EEG waveforms of healthy subjects and schizophrenia patients were collected and processed. The data was used to develop a convolutional neural network (CNN) model which can automatically extract features and classify them. CNN does this by comparing the differences between the EEG waveforms of schizophrenia patients and healthy controls. These differences were used to train the classifier to differentiate the schizophrenia patients from the controls. The result of the CNN model showed a test accuracy of 60%, specificity of 55.55% and a precision of 55.55%. This early result shows that the model is promising. The next step will be to improve the accuracy of the model with a larger pool of data and many iterations, which is expected to lead to a better model that can be relied upon for schizophrenia diagnosis. In conclusion, CNN-based models like this one are relatively cheap and will improve the diagnosis of Schizophrenia, especially in low-income economies where the present study has been carried out.

Key Words: *Artificial intelligent; Confusion matrix; Electroencephalogram; Schizophrenia diagnosis*

***Corresponding Author**: Ephraim Nwoye, Professor, Department of Biomedical Engineering, University of Lagos, Nigeria, Tel: +234 8033735926; E-mail: enwoye@unilag.edu.ng*

***Received Date**: July 08, 2020, **Accepted Date**: September 09, 2021, **Published Date**: December 15, 2021*

***Citation**: Nwoye E, Fidelis O, Umeh K, et al. Detecting Hidden Patterns in EEG Waveforms of Schizophrenia Patients using Convolutional Neural Network. Int J Auto AI Mach Learn. 2021;2(2):01-19*

1. Introduction

Schizophrenia is a common disease with a prevalence of about 7 in 1000 people in which individuals interpret reality abnormally. To have a better understanding of the severity of prevalence, it is almost four times the global cases of Alzheimer's disease [1] and six times that of type 1 diabetes [2]. Schizophrenia is a complex mental disorder, with typical onset in late adolescence or early adulthood. In spite of the level of ongoing research, outcomes from best-practice treatment are often poor. A systematic review based on 50 outcome studies reported that the average proportion of people suffering from schizophrenia who met clinical and social recovery criteria was only 13.5% [3]. Accompanying poor recovery outcomes, schizophrenia patients have significantly reduced life expectancy [4]. High mortality appears across all age groups [5] and this differential mortality gap between those with and without schizophrenia may have increased in recent decades [6]. Schizophrenia has also been linked to higher rates of medical conditions coexisting in one individual and most deaths are due to underlying conditions, especially chronic diseases such as coronary heart disease, stroke, type II diabetes, respiratory diseases, and some cancers; while unnatural causes, including suicide, account for less than 15% of these deaths [5]. According to the Diagnostics and Statistical Manual for Mental Disorders, fifth edition (DSM-V), schizophrenia is diagnosed by an array of symptom clusters based on a careful patient population study. Symptoms may be positive or negative symptoms. Positive symptoms include delusions, hallucinations, and disorganized behavior or speech. They generally represent more apparent and outward symptoms of psychosis. Negative symptoms include apathy, social withdrawal, cognitive deficits and poverty of speech [7].

As many as 75% of the patients with schizophrenia pass through an early warning stage indicative of the presence of the medical condition [8]; with only a proportion eventually developing a psychotic disorder. At the moment, about 1 in 4 of the patients who can be categorized as high risk will convert to schizophrenia and therefore require intervention. According to different studies, about 20%–40% of those who meet UHR criteria convert to psychosis within 2 to 4 years. Young adults with "subthreshold" psychotic symptoms and some signs of functional impairment with a history of genetic risk for schizophrenia were identified and monitored longitudinally for the development of psychosis in studies undertaken in both Australia and the United States [8]. Prodromal status was linked to a 40%–50% conversion rate to psychosis in 1–2 years in a few previous investigations. A group of North American researchers explained why the rate of transformation was slower. Patients' 1-year transition rates in Australia have consistently fallen over time, dropping to 50% in 1995, 32% in 1997, 21% in 1992, and 12% in 2000 [8].

It is not known what causes schizophrenia, but researchers believe that a combination of genetics, brain chemistry, and the environment contributes to the development of the disorder. Problems with certain naturally occurring brain chemicals, including neurotransmitters called dopamine and glutamate, may contribute to schizophrenia. Neuroimaging studies show differences in the brain structure and central nervous system of people with schizophrenia. While researchers are not certain about the significance of these changes, they indicate that schizophrenia is a brain disease. Although the precise cause of schizophrenia is not known, certain factors seem to increase the risk of developing or triggering schizophrenia, including:

(i) Having a family history of schizophrenia
(ii) Increased immune system activation, such as form inflammation or autoimmune diseases
(iii) Older age of the father
(iv) Some pregnancy and birth complications, such as malnutrition or exposure of toxins or viruses that may impact brain development
(v) Taking mind-altering (psychoactive or psychotropic) drugs during teen years and young adulthood

There is no definite way to prevent schizophrenia, but following treatment plan may help prevent relapses or worsening of symptoms. Besides, researchers hope that learning more about risk factors for schizophrenia may lead to earlier diagnosis and treatment.

2. Previous Works

2.1. Electroencephalogram (EEG) Procedure

Electroencephalogram (EEG) is a test that records the electrical signals of the human brain in the form of waveforms. The work in [9] proposed that analyzing neurological conditions such as Parkinson's disease, epilepsy, depression and Alzheimer's disease by computerized means would be a breakthrough in the field of science. Moreover, a lot of research studies are ongoing on the use of computer in the diagnosis of schizophrenia. Medical procedures involving multimodal imaging techniques are currently being used to detect schizophrenia. Some of these modalities include computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and ultrasonography. Although, these modalities can be combined when one modality is not enough to diagnose the neuronal disease of the patient [10], a combination of these modalities is exorbitant, time-consuming and burdensome. Therefore, a more cost-effective method of diagnosing SZ is required. Kim et al. [11] extracted EEG signals with 21 gold cup electrodes placed according to the 10-20 international system while monitoring the horizontal and vertical eye movements of participants. Software solutions such as MATLAB and EEG Lab were used in the signal preprocessing and five frequency bands were selected for analysis. Fast Fourier transform was used to compute the spectral power of EEG data, after which the EEG power deviations were studied using the analysis of variance (ANOVA) for each of the five frequency bands examined. The receiver operating curve (ROC) was done to differentiate between normal and schizophrenia SZ patients. It was reported that the delta power had the highest classification accuracy of 62.2%. Kim et al. [11] also researched EEG recordings on 50 participants using 64 electrodes. The electrodes were placed above and beneath the right eye and laterally concerning the right and left eyes to monitor horizontal and vertical movements respectively. The EEG signals were preprocessed with the raw materials being partitioned into relevant intervals. The extracted features were used to represent normal and schizophrenia patients. This produced better classification accuracy with five different electrodes having a prediction accuracy ranging from 92.0% and 93.9%.

2.2. Convolutional Neural Network

Recently, several approaches for image classification were described and compared with other approaches. But in general, image classification refers to the process of extracting information from the image by labeling the pixels of the image to different classes. Image classification can be done in two ways, namely supervised and unsupervised classification. Silva et al. [12], has discussed the use of the unsupervised learning algorithm in the underwater fish recognition framework for classifying images. In this method, the pixels of the image are clustered into groups without the intervention of the analyst. Krizhevsky et al. [13], the author discussed supervised classification techniques that analyze and train the classifier on the labeled images and extracting features from them. By using the learned information of the training, the newly provided image will be classified based on the features observed in the image.

In current research, deep learning algorithms are providing successful results in the area of computer vision. The Convolutional Neural Network (CNN), a machine learning algorithm is being used for mage classification. Lunga et al. [14], uses deep learning algorithms for classifying the quality of the wood board by using extracted texture information from the wood images, also comparisons were made several other machine learning architectures. CNN is a type of feed-forward artificial neural network that has been successfully applied to analyze visual images. It is inspired by the biological processes and how the neurons are connected as in the animal visual cortex.

Lim et al. [15], the author discussed automatic recognition cattle images using CNN which helps to extract the necessary characteristic from the cattle images and Support Vector Machine (SVM) techniques was used for the classification of those images. Koh et al. [16], employed high-resolution images in ImageNet data set having 15 million labelled images with 1000 different classes for classification with the aid of deep Convolutional Neural Network. In general, images are constructed as a matrix of pixels and these pixel values are given as input to the input layer along with weights and biases (for non-linearity). The output layer will be a fully connected layer usually to classify the image to which class it belongs to. The hidden layer may be convolutional, pooling or fully connected.

The manifold and multi-view learning techniques for classifying remotely sensed hyperspectral data, medical chatbot and heart waveforms have also been presented in some studies [17,18]. The convolutional layer is a core building block and has learnable filters as parameters [19]. Each filter is spatially small (width and height) but extends across the depth of the input volume. The 2-dimensional activation map is produced by performing a dot product between input and entries of filter for every filter. As a result, the network learns the filter that activates when it detects specific features in some spatial position in the input. The pooling layer is used in down-sampling the image without losing any information from the image. Max pooling uses the maximum value from the cluster of neurons at the prior layer. The fully connected layer connects every neuron in one layer to every neuron in other layers [19]. CNN uses little pre-processing when compared to traditional classification algorithms that use filters that are hand-engineered. The independence of human intervention in learning filters is a good advantage of CNN. CNN is a supervised deep learning approach that requires large labeled

data for training on the network. After the training, the model will learn the weights and the accuracy of the classifier is improved.

There are other works that do not use CNN. For example, one study [2] discussed various approaches of extracting features for related biomedical problems in diabetes using machine learning algorithms. As an alternative to CNN, the authors in another study [20] developed a cost sensitive pruned decision tree model for improved schizophrenia diagnosis. The choice of using CNN is dependent on how the input features are extracted. In this present study, the authors have developed a pipeline that processes the EEG waveforms in parallel and turn them into an image for spatial features extraction. The contributions of the paper are as follows:

(i) Developing a new psychiatric dataset from the Sub-Saharan African community
(ii) Parallel preprocessing of EEG waveforms into a high-resolution image file
(iii) Capturing the spatial features of the time-domain EEG waveform-image file
(iv) Developing a CNN to extract the spatial features of EEG waveforms, and validating its performance

3. Methodology

3.1. Data Acquisition for the CNN Classifier

The experimental data were collected from the Psychiatric Department of Lagos University Teaching Hospital, Idi-Araba, Lagos State. The ethical approval letter with the reference number ADM/DCST/HREC/APP/3294 was obtained from the Ethical Committee of Lagos University Teaching Hospital, Lagos. The dataset consists of 81 healthy and 111 unhealthy EEG waveforms of mental health patients reported between the years of 2014 and 2019. The dataset used in this research includes patients diagnosed with schizophrenia and other mental health. Notwithstanding, only attributes based on DSM-V and ICD-11 specifications for diagnosis of Schizophrenia were used in training and validating the model.

The purpose of data or image preprocessing is to improve on the images by suppressing unwanted distortions to enhance some image features important for further processing. This is because the data acquired from hospitals are mostly in their raw form and would require some processing before, they can be fed into the CNN classifier. The various steps in data preprocessing are image enhancement, loading of the dataset into the CNN classifier and Image resizing and labelling.

The dataset was imported using Spyder IDE as shown in Figure 1 and the file path is given below:

path1='C:\\Users\KEN\Documents\PythonScripts\KERAS-CNN-PROJECT\Input_data2'.

The importation was made possible with the aid of some in-built libraries such as Keras, Theano and Numpy. The Spyder interface is shown in Figure 1.

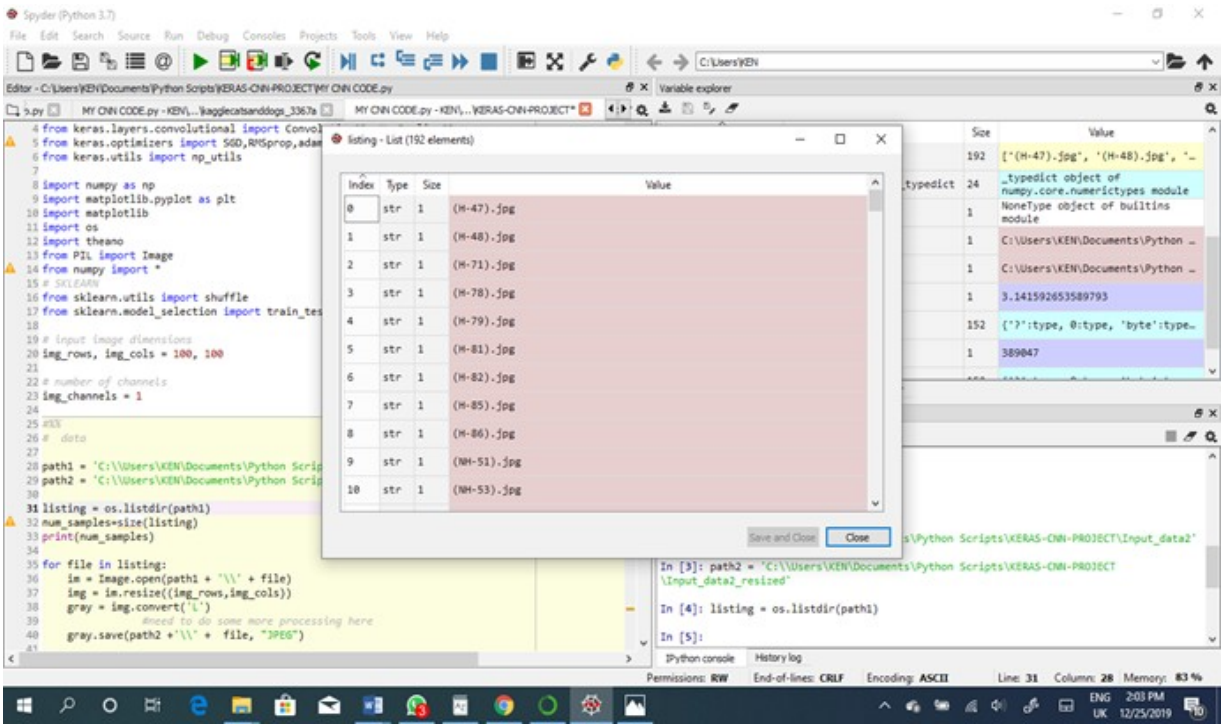

Figure 1: *Spyder IDE interface.*

3.1.1. Image Preprocessing

The EEG images acquired from the dataset were preprocessed in order to resize the images to a general pixel value of 100×100. After resizing the images, the images were converted to gray scale images in order to limit the amount of information fed into the network, only important features of the waveforms are required therefore noise was eliminated from said images, this technique helped in identifying key features in the waveforms thereby improving the accuracy of the model. The images were also labeled according to their various categories as shown in Figure 2 and Figure 3 i.e. Healthy and Unhealthy represented as H and U respectively.

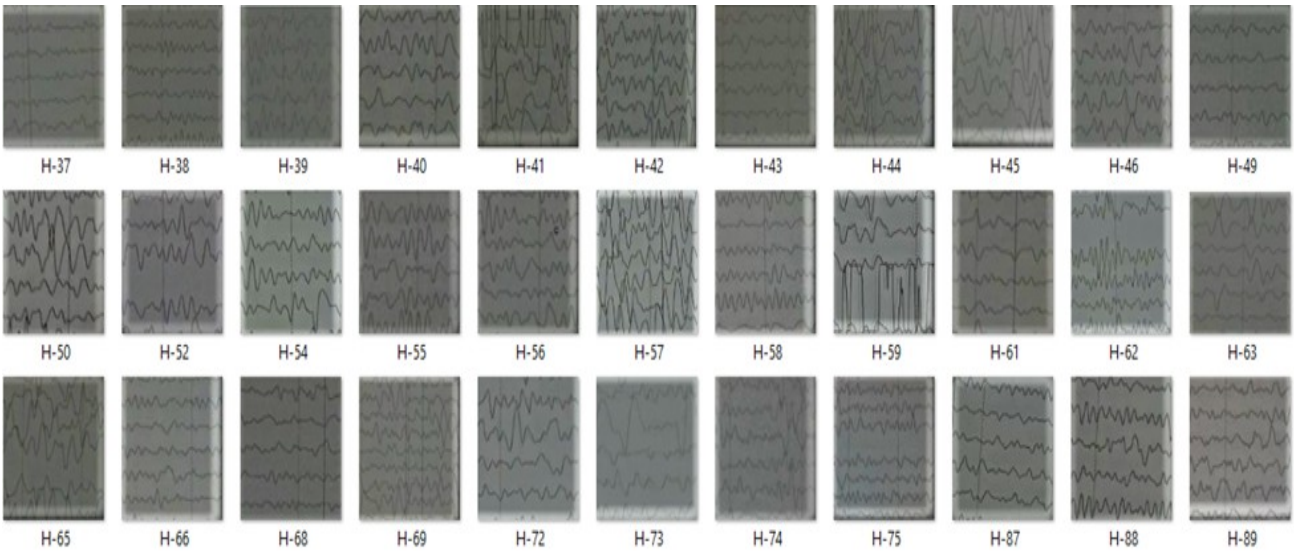

Figure 2: *Window showing resized images of healthy EEG waveforms.*

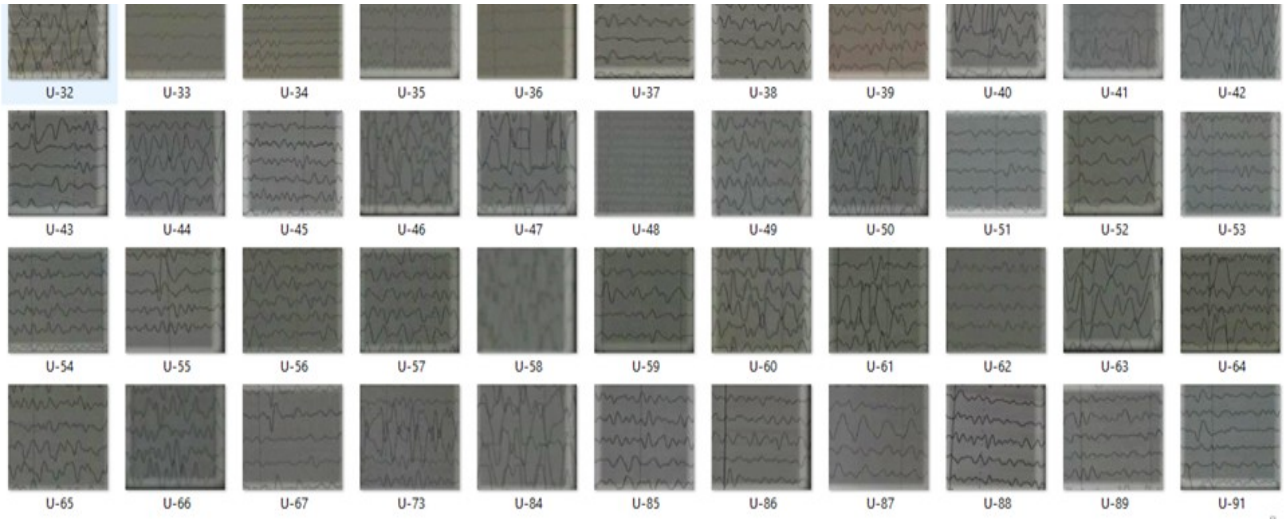

Figure 3: *Window showing resized images of unhealthy EEG waveforms.*

3.1.2. Dataset Splitting

The Sci-kit learn library was used in splitting the datasets into different train-test ratios during the research in order to obtain optimal result. The following train-test ratios were utilized (50:50, 60:40, 40:60, 30:70, and 70:30).

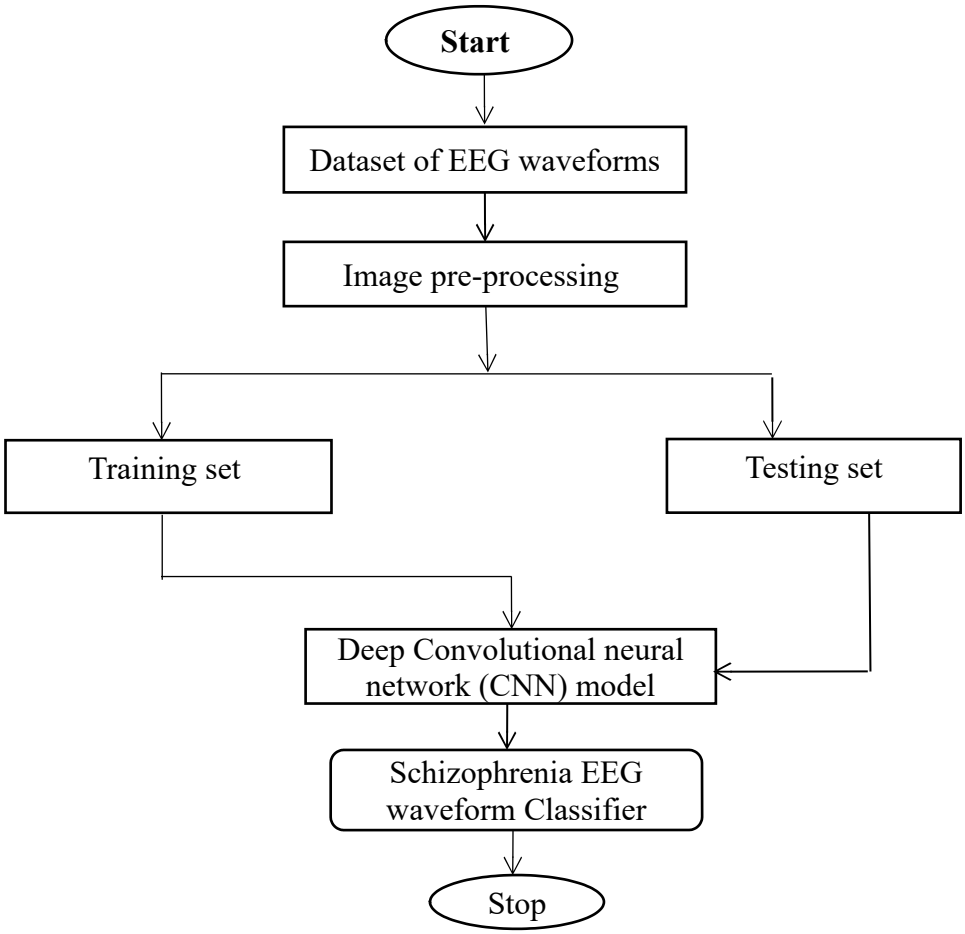

Figure 4: *Flowchart of Methodology.*

As shown in Figure 4, the EEG waveforms are first preprocessed using different techniques before being fed into the CNN for classification. The CNN model is developed using 2 convolutional layers in order to improve accuracy. The CNN model architecture is described in Table 1. The input data that goes into the CNN is an image which is constructed from the parallel time-domain EEG waveforms. Thus this image is a representation of time-spatial characteristics of the EEG events in a particular participant. As such, the CNN behaves like a spatio-temporal classifier although its structure does not contain any feedback path.

Table 1: *Proposed CNN Architecture.*

Layer	Description
Convolutional Layer (CL1)	Layer that uses a 3×3 feature detector that take a 1×1 step through the input data.
Max Pooling Layer (CL2)	In this layer the number of nodes in the output data from the previous layer is reduced by using a 2×2 feature detector that takes a 2×2 stride through the data, thereby reducing the size by half approximately.
Convolutional Layer (CL3)	Layer that uses a 3×3 feature detector that take a 1×1 step through the input data.
Max Pooling Layer (CL4)	Similar to the first max pooling layer, but a dropout of 0.5 is applied to the output of this layer.
Flattening (CL5)	The output data from the previous layer is converted to a 1-D vector.
Fully Connected Layer (CL6)	Layer with 128 neurons and relu activation function and a dropout of 0.25 applied.
Output Layer (CL7)	The output layer has a softmax activation function for classification.

We examine the mathematical definitions of convolution and cross-correlation:

3.2. Cross-correlation

Given an input image I and a filter (kernel) K of dimensions $k_1 \times k_2$, the correlation operation is given by:

$$(I \otimes K)_{ij} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} I(i+m, j+n) K(m,n) \qquad (1)$$

Given an input image I and a filter (kernel) K of dimensions $k_1 \times k_2$, the convolution operation is given by:

$$(I * K)_{ij} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} I(i-m, j-n) K(m,n) \qquad (2a)$$

$$\sum_{m=0}^{k_1-1}\sum_{n=0}^{k_2-1} I(i+m, j+n)K(-m,-n) \qquad (2b)$$

From equation (2), it is seen that convolution is the same as cross-correlation with a flipped kernel i.e. for a kernel K where K(-m,-n) = K(m,n)

Our CNN consists of convolutional layers that are characterized by an input map I, a bank of filters K and biases b.

In the case of images, we could have as input an image with height H, width W and C = 3 channels (red, blue and green) such that $I \in R^{H\times W\times C}$. Subsequently for a bank of D filters we have $K \in R^{k_1\times k_2\times C\times D}$ and biases b ϵ R^D, one for each filter. The output from this convolution procedure is as follows:

$$(I*K)_{ij} = \sum_{m=0}^{k_1-1}\sum_{n=0}^{k_2-1}\sum_{c=1}^{C} K_{m,n,c}.I_{i+m,j+n,c} + b \qquad (3)$$

The convolution operation carried out here is the same as cross-correlation, except that the kernel is "flipped" (horizontally and vertically). For simplicity, we shall use the case where the input image is gray scale i.e. single-channel C=1. Equation (3) will be transformed to:

$$(I*K)_{ij} = \sum_{m=0}^{k_1-1}\sum_{n=0}^{k_2-1} K_{m,n}.I_{i+m,j+n} + b \qquad (4)$$

To help us explore the forward and back propagation, the following notations are used:

1 l is the l^{th} the layer where $l=1$ is the first layer and $l=L$ is the last layer.
2 Input x is of dimension $H \times W$ and has i by j as the iterators.
3 Filter or kernel w is of dimension $k_1 \times k_2$has m by n as the iterators
4 $w^l_{m,n}$ is the weight matrix connecting neurons of layer l with neurons of layer $l-1$.
5 b^l is the bias unit at layer l.
6 $x^l_{i,j}$ is the convolved input vector at layer$l$ plus the bias represented by

$$x^l_{i,j} = \sum_m \sum_n w^l_{m,n} o^{l-1}_{i+m,j+n} + b^l \qquad (5)$$

7 $o^l_{i,j}$ is the output vector at layer l given by

$$o^l_{i,j} = f\left(x^l_{i,j}\right) \qquad (6)$$

8 $f(.)$is the activation function. Application of the activation layer to the convolved input vector at layer l is given by $f(x^l_{i,j})$

3.3. Forward Propagation

To perform a convolution operation, the kernel is flipped $180^0$and slid across the input feature map in equal and finite strides. At each location, the product between each element of the kernel and the input feature map element overlaps is computed and the results summed up to obtain the output at that current location.

Units in the convolutional layer illustrated above have receptive fields of size 4 in the input feature map and are thus only connected to 4 adjacent neurons in the input layer. This is the idea of sparse connectivity in CNNs where there are local connectivity patterns between neurons in adjacent layers. The color codes of the weights joining the input layer to the convolutional layer show how the kernel weights are distributed (shared) amongst neurons in the adjacent layers. Weights of the same color are constrained to be identical. The convolution process here is usually expressed as a cross-correlation but with a flipped kernel. In the diagram below we illustrated a kernel that has been flipped both horizontally and vertically:

The convolution equation of the input at layer l is given by:

$$x_{i,j}^{l}=rot_{180^0}\left\{w_{m,n}^{l}\right\}*o_{i,j}^{l-1}+b_{i,j}^{l} \tag{7}$$

$$x_{i,j}^{l}=\sum_{m}\sum_{n}w_{m,n}^{l}o_{i+m,j+n}^{l-1}+b_{i,j}^{l} \tag{8}$$

$$o_{i,j}^{l}=f\left(x_{i,j}^{l}\right) \tag{9}$$

For a total of P predictions, the predicted network outputs y_p and their corresponding targeted values t_p the mean squared error is given by:

$$E=\frac{1}{2}\sum_{p}\left(t_p-y_p\right)^2 \tag{10}$$

For backpropagation, there are two updates performed, for the weights and the deltas. Let's begin with the weight update. We are looking to compute $\frac{\partial E}{\partial w_{m',n'}^{l}}$which can be interpreted as the measurement of how the change in a single pixelm',n' in the weight, kernel affects the loss function E.

During forward propagation, the convolution operation ensures that the yellow pixel m',n' in the weight, the kernel contributes all the products (between each element of the weight kernel and the input feature map element it overlaps). This means that pixel m',n'will eventually affect all the elements in the output feature map.

Convolution between the input feature map of dimension $H \times W$ and the weight kernel of dimension $k_1 \times k_2$ produces an output feature map of size $H-k_1+1$ by $W-k_2+1$. The gradient component for the individual weights can be obtained by applying the chain rule in the following way:

$$\frac{\partial E}{\partial w^l_{m',n'}} = \sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} \frac{\partial E}{\partial x^l_{i,j}} \frac{\partial x^l_{i,j}}{\partial w^l_{m',n'}} \tag{11}$$

$$= \sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} \delta^l_{i,j} \frac{\partial x^l_{i,j}}{\partial w^l_{m',n'}} \tag{12}$$

In equation 11, $x^l_{i,j}$ is equivalent to $\sum_m \sum_n w^l_{m,n} o^{l-1}_{i+m,j+n} + b^l$ and expanding this part of the equation gives us:

$$\frac{\partial x^l_{i,j}}{\partial w^l_{m',n'}} = \frac{\partial}{\partial w^l_{m',n'}} \left(\sum_m \sum_n w^l_{m,n} o^{l-1}_{i+m,j+n} + b^l \right) \tag{13}$$

Further expanding the summations in equation (12) and taking the partial derivatives for all the components results in zero values for all except the components where $m=m'$ and $n=n'$ in $w^l_{m,n} o^{l-1}_{i+m,j+n}$ as follows:

$$\frac{\partial x^l_{i,j}}{\partial w^l_{m',n'}} = \frac{\partial}{\partial w^l_{m',n'}} \left(w^l_{0,0} o^{l-1}_{i+0,j+0} + \ldots + w^l_{m',n'} o^{l-1}_{i+m',j+n'} + \ldots + b^l \right) \tag{14}$$

$$= \frac{\partial}{\partial w^l_{m',n'}} \left(w^l_{m',n'} o^{l-1}_{i+m',j+n'} \right) \tag{15}$$

$$= o^{l-1}_{i+m',j+n'} \tag{16}$$

Substituting equation (13) in equation (11) gives us the following results:

$$\frac{\partial E}{\partial w^l_{m',n'}} = \sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} \delta^l_{i,j} o^{l-1}_{i+m',j+n'} \tag{17}$$

$$= rot_{180^0}\{\delta^l_{i,j}\} * o^{l-1}_{m',n'} \tag{18}$$

$$\delta^l_{i,j} = \frac{\partial E}{\partial x^l_{i,j}} \tag{19}$$

$$\delta_{i,j}^{l} = \frac{\partial E}{\partial x_{i,j}^{l}} \tag{20}$$

Using chain rule and introducing sum gives us the following equation:

$$\frac{\partial E}{\partial x_{i',j'}^{l}} = \sum_{i,j \epsilon Q} \frac{\partial E}{\partial x_{Q}^{l+1}} \frac{\partial x_{Q}^{l+1}}{\partial x_{i',j'}^{l}} \tag{21}$$

$$= \sum_{i,j \epsilon Q} \delta_{Q}^{l+1} \frac{\partial x_{Q}^{l+1}}{\partial x_{i',j'}^{l}} \tag{22}$$

$$\frac{\partial E}{\partial x_{i',j'}^{l}} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \frac{\partial E}{\partial x_{i'-m,j'-n}^{l+1}} \frac{\partial x_{i'-m,j'-n}^{l+1}}{\partial x_{i',j'}^{l}} \tag{23}$$

$$= \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \delta_{i'-m,j'-n}^{l+1} \frac{\partial x_{i'-m,j'-n}^{l+1}}{\partial x_{i',j'}^{l}} \tag{24}$$

The method is summarized by the flow chart in Figure 5:

The steps involved in the convolutional neural network algorithm are as follows:

Step 1: Initialize the number of inputs, outputs, and hidden layers.

Step 2: Select the learning rate of 0.001

Step 3: Initialize weight to random.

Step 4: Enter the output of each layer for the training.

Step 5: Compute error derivative in the output layer

Step 6: If there is an error, the model updates the input, output, and hidden layers.

Step 7: But if there is no error, the model generates the final output and stops operation.

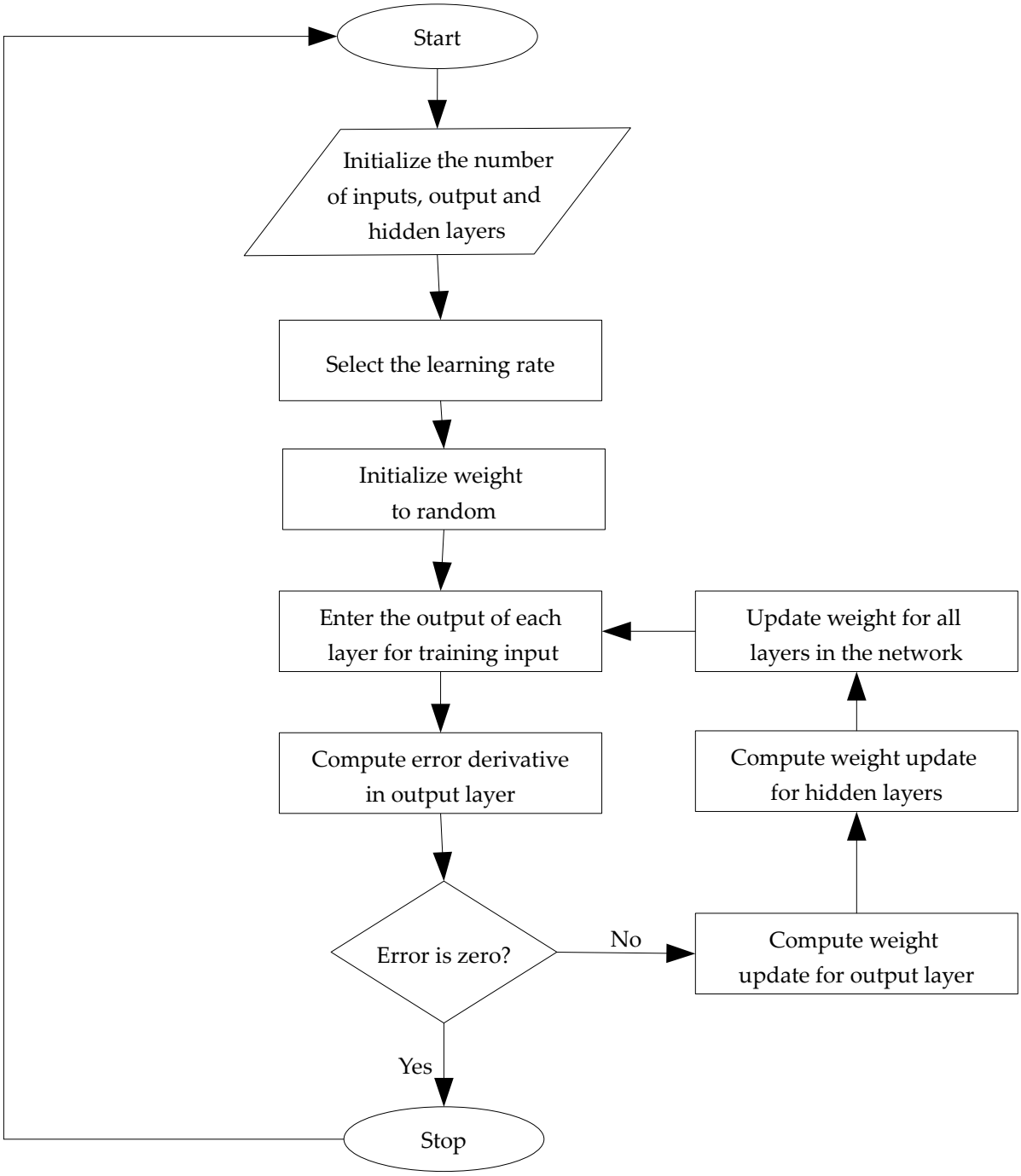

Figure 5: *Convolutional neural network algorithm.*

4. Results and Discussions

The results of the study is presented in Figure 6-11 for the various model accuracy and loss, which represent the best results, observations, and analysis as related to the study. In Figure 6, the training accuracy at 10 epochs was 0.97 and from 11 epochs to 50 epochs it was constant at 0.97, while the validation accuracy was around the same point. In Figure 7, the model loss was between 0.1 and 0.15 which is minimal. From Figure 8, between 0 and 50 epochs, the model

recorded a train and validation accuracies of 0.87 and 0.91 respectively. In Figure 9, the model loss increased a little bit between 0.27 and 0.34 against 0.1 and 0.15 as recorded in Figure 8.

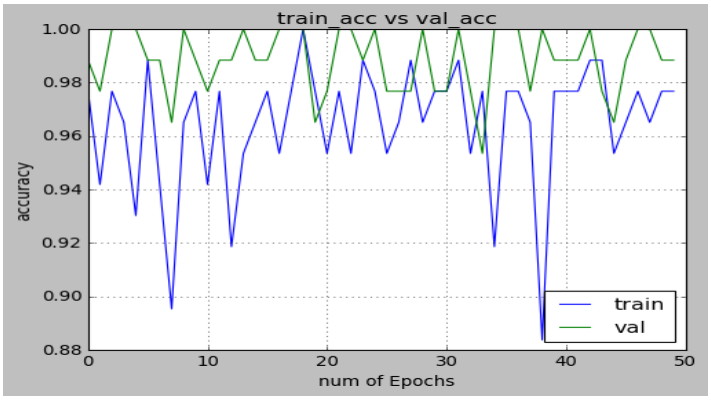

Figure 6: *Model accuracy for 50/50 Train-test-split.*

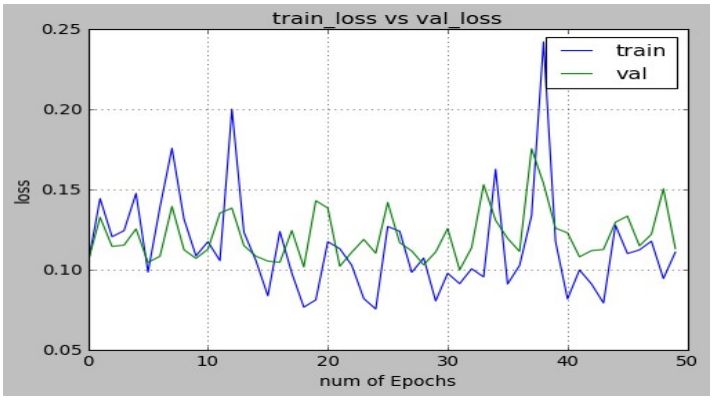

Figure 7: *Model loss for 50/50 Train-test-split.*

From Figures 6 and 7, the training accuracy at 10 epochs was 0.97 and from 11 epochs to 50 epochs it was constant at 0.97, while the validation accuracy was around the same point. Also the model loss was between 0.1 and 0.15 which is minimal.

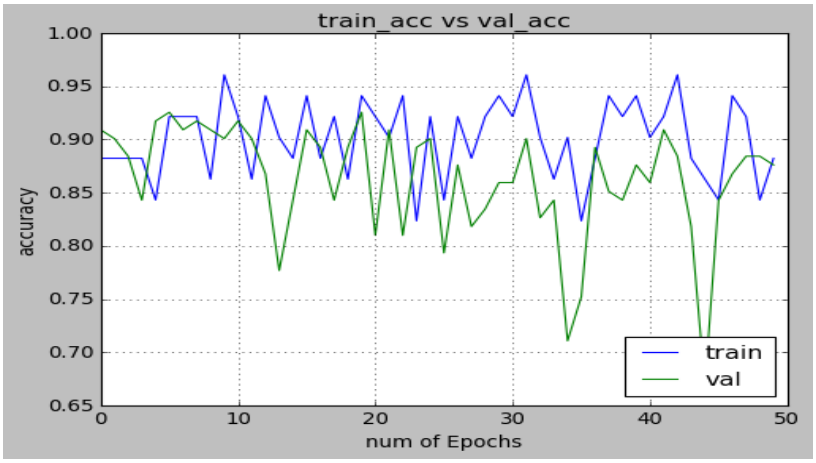

Figure 8: *Model accuracy for 70/30 Train-test-split.*

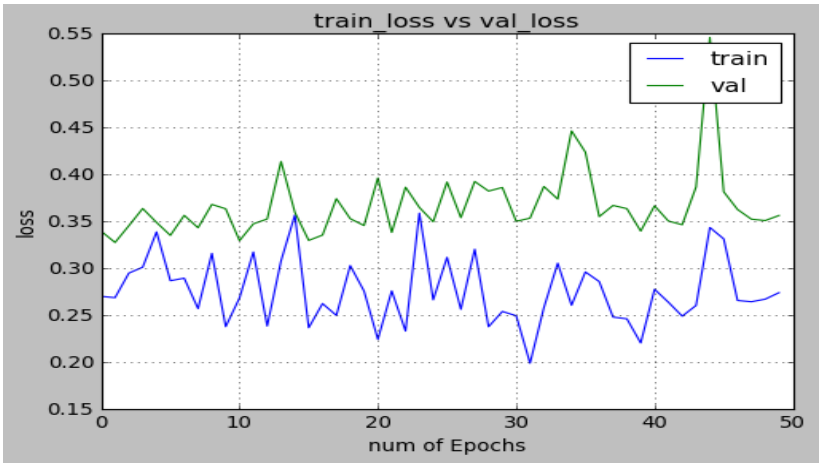

Figure 9: *Model loss for 70/30 Train-test-split.*

From Figures 10 and 11, from 0 to 50 epochs, there was a perfect tie between the training accuracy and validation accuracy of 0.95 thereby producing the best result. Also the model loss indicated perfect tie of 0.16 between the train and validation loss.

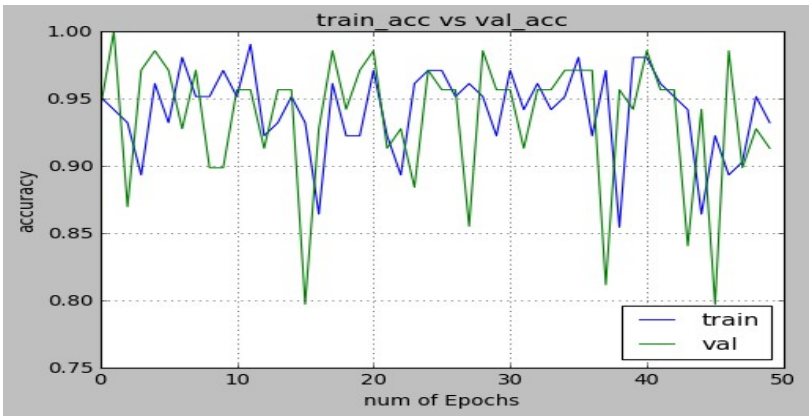

Figure 10: *Model accuracy for 60/40 Train-test-split.*

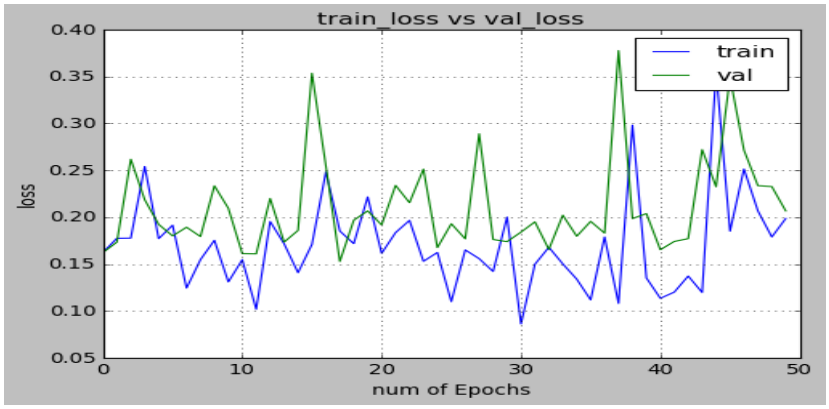

Figure 11: *Model loss for 60/40 Train-test-split.*

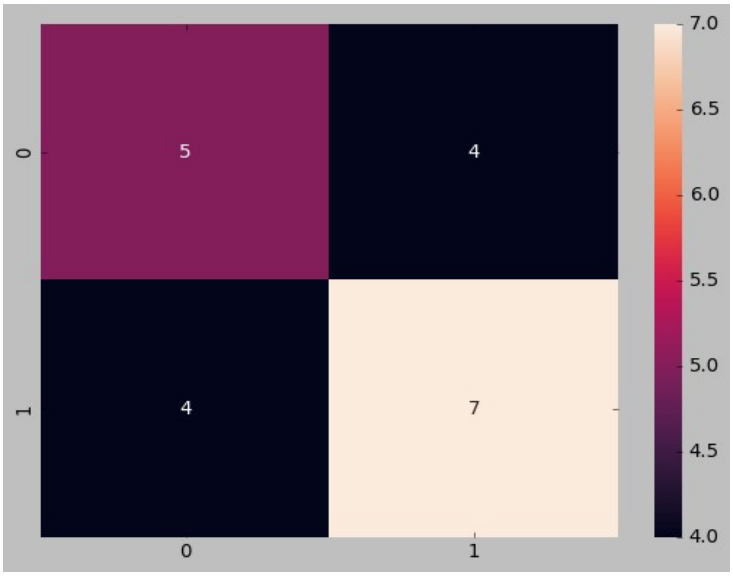

Figure 12: *Confusion matrix plot the 50/50 Train-test-split.*

Figures 12 shows the confusion matrix shows that 60% of the dataset in the model was accurately predicted. 55.6% of schizophrenic cases were precisely predicted with a specificity of 63.6%.

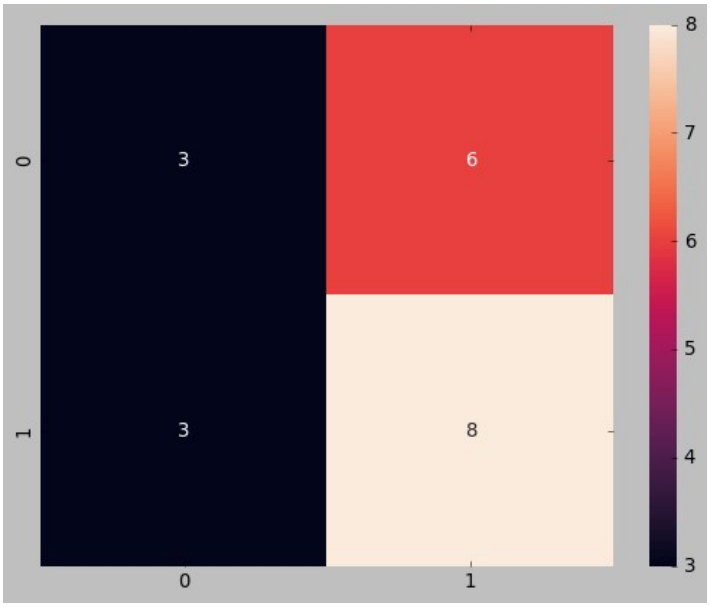

Figure 13: *Confusion matrix plot the 70/30 Train-test-split.*

Figures 13 shows the confusion matrix shows that 57.8% of the dataset in the model was accurately predicted. 33.3% of schizophrenic cases were precisely predicted with a specificity of 57.1%.

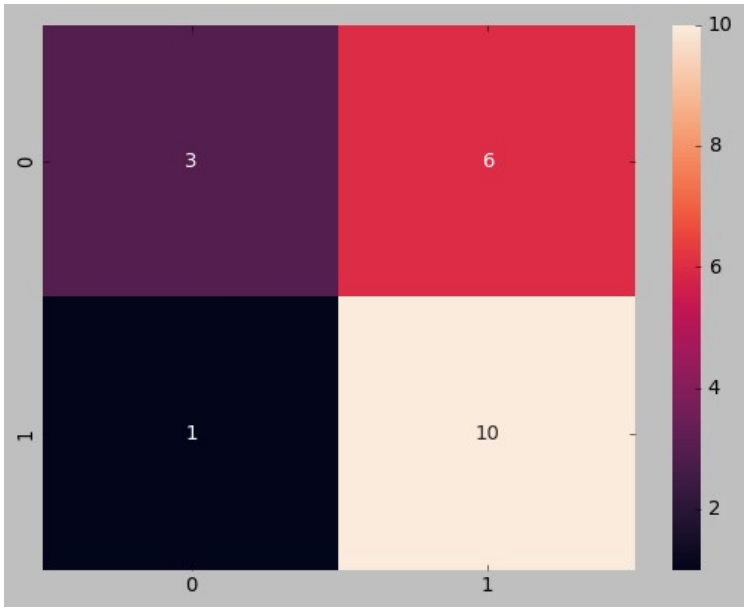

Figure 14: *Confusion matrix plot the 40/60 Train-test-split.*

From Figure 14, the confusion matrix shows that 65% of the dataset in the model was accurately predicted. 33.3% of schizophrenic cases were precisely predicted with a specificity of 62.5%.

The summary of the results is shown in Table 2.

Table 2: *Result summary of the train-test split.*

Train	Test	F-Measure	Test Accuracy
30	70	40%	55%
40	60	46.2%	65%
50	50	55.6%	60%
60	40	60%	60%
70	30	40%	50%

At the end of the experiment, the test results from the various configurations such as 50/50, 60/40, 70/30, 40/60 and 30/70 train-test splits were observed to rate the performance of the model. During the iterations, it was observed that the model performed best with a 60/40 train-test split with an accuracy of 60% and an f-measure value of 60%, followed by a 50/50 train-test split with an accuracy of 60% and f-measure value of 55.55% thereby producing the least over-fitting. The performance of the model was enhanced and optimized by the introduction of the dropout technique thereby preventing the model from overfitting and increasing the accuracy of the model as well. This early result shows that the model is promising. The next step will be to improve the accuracy of the model with a larger pool of data and many iterations, which is expected to lead to a better model that can be relied upon for schizophrenia diagnosis. The main limitation of our study is the small sample size. Another major challenge in the experiment was the issue of overfitting in some of the configurations, which was due to irregular train/test split ratio and small datasets.

5. Conclusions

In conclusion, the study shows that the development of an artificial intelligence-based software in low-income economies for detecting schizophrenia using EEG waveforms is possible. With the software, exhaustive screening of schizophrenia patients to alert for behavioral markers of the disease will no longer be required, as the model will be satisfactorily efficient, automatically assisting with the diagnosis. This system is foreseen to be a windfall to clinicians as a diagnostic tool, aiding them in schizophrenia assessment. It is also relatively cheap especially for those in the rural areas who cannot afford regular medical check-ups. The results of this initial study will support future work as the performance of this model can be improved with a larger dataset. To the best of authors' knowledge, not many CNN-based studies exist for Schizophrenia diagnosis in low-income countries. In fact, this seem to be the first of such studies in Nigeria and is targeted towards providing an accurate and standardized means of schizophrenia diagnosis and early detection, other than the conventional method of interviewing patients, based on symptoms presented and the presence or absence of behavioral signs.

Conflict of interests: The authors declare that there is no conflict of interest.

References

1. Hebert LE, Weuve J, Scherr PA, et al. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. Neurology. 2013;80:1778-83.
2. Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract. 2010;87:4-14.
3. Charlson FJ, Ferrari AJ, Santomauro DF, et al. Global epidemiology and burden of schizophrenia: findings from the global burden of disease study 2016. Schizophr Bull. 2018;44:1195-203.
4. Laursen TM, Nordentoft M, Mortensen PB. Excess early mortality in schizophrenia. Annu Rev Clin Psychol. 2014;10:425-48.
5. Charlson FJ, Baxter AJ, Dua T, et al. Excess mortality from mental, neurological and substance use disorders in the Global Burden of Disease Study 2010. Epidemiol Psychiatr Sci. 2015;24:121-40.
6. Deutsch-Link S. Language in schizophrenia: what we can learn. Yale University. 2016.
7. APA. Diagnostic and Statistical Manual of Mental disorders:DSM-5. Washington, D.C. American Psychiatric Association. 2013.
8. George M, Maheshwari S, Chandran S, et al. Understanding the schizophrenia prodrome. Indian J Psychiatry. 2017;59:505-9.
9. Jack Jr CR, Lowe VJ, Weigand SD, et al. Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: Implications for sequence of pathological events in Alzheimer's disease. Brain. 2009;132:1355-65.

10. Chuang MC, Hwan JN, Williams K. Supervised and unsupervised feature extraction methods for underwater fish species recognition. 2014 ICPR Workshop on Computer Vision for Analysis of Underwater Imagery. Stockholm, Sweden. 2014.
11. Kim H, Koo J, kim D, et al. Image-Based monitoring of jellyfish using deep learning architecture. IEEE S J. 2016;16:2215-16.
12. Silva C, Welfer D, Gioda FP, et al. Cattle brand recognition using convolutional neural network and support vector machines. IEEE Lat Am Trans. 2017;15:310-6.
13. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA. 2012.
14. Lunga D, Prasad S, Crawford MM, et al. Manifold-learning-based feature extraction for classification of hyperspectral data: review of advances in manifold learning. IEEE Signal Process Mag. 2014;31:55-66.
15. Lim CLP, Woo WL, Dlay S, et al. Deep Multi- Signal View Heartwave Authentication. IEEE Trans Industr Inform. 2019;15:777-86.
16. Koh BHD, Woo WL. Multi-view Temporal Ensemble for Classification of Non-Stationary Signals. IEEE Access. 2019;7:32482-91.
17. Woo WL, Gao B, Al-Nima RRO, et al. Development of conversational artificial intelligence for pandemic healthcare query support. Int J Auto AI Mach learn. 2020;1:54-79.
18. Koh BHD, Lim CLP, Rahimi H, et al. Deep Temporal Convolution Network for Time Series Classification. Sensors. 2021;21:603.
19. Rodríguez-Rodríguez I, Rodríguez JV, Woo WL, et al. A comparison of feature selection and forecasting machine learning algorithms for predicting glycaemia in type 1 diabetes mellitus. Appl Sci. 2021;11:1742.
20. Nwoye E, Woo WL, Fidelis O, et al. Development and investigation of cost-sensitive pruned decision tree model for improved schizophrenia diagnosis. Int J Auto AI Mach learn. 2020;1:17-41.