# HIGH-PERFORMING MACHINE LEARNING MODEL FOR SCHIZOPHRENIA CLASSIFICATION

by

YUNUS Mujahid Olalekan

**A PROJECT SUBMITTED TO THE DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING IN THE FACULTY OF TECHNOLOGY, OBAFEMI AWOLOWO UNIVERSITY, ILE-IFE.**

**March 2024**

# Table of Contents

# CERTIFICATION

This is to certify that the project titled **HIGH-PERFORMING MACHINE LEARNING MODEL FOR SCHIZOPHRENIA CLASSIFICATION** was designed and constructed by YUNUS Mujahid Olalekan for the 2022/2023 academic session in the department of Electronic and Electrical Engineering, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria.

**Under the supervision of**

**Dr K. P. Ayodele**

**Department of Electronic/Electrical Engineering,**

**Obafemi Awolowo University,**

**Ile-Ife, Osun State.**

_____

# DEDICATION

This project is dedicated to my parents and sponsors.

# ACKNOWLEDGEMENT

I would like to acknowledge the mentorship of my Supervisor, Dr K.P. Ayodele throughout the period of the course.

# ABSTRACT

Schizophrenia (SZ) is a chronic, severe, debilitating mental disorder characterized by disorganized thoughts, unusual behaviors, and disruptive actions. It often leads individuals to lose touch with reality (Shalini et al., 2021). SZ poses significant harm to patients, underscoring the importance of timely and accurate detection (Jie et al., 2021). In recent years, there has been a growing interest in utilizing machine learning and deep learning models for detection. However, there is a lack of classification models trained on local data and tailored to specific age and gender groups.

This project aims to address these gaps by developing a high-performing classification model trained on local data that includes demographic information such as age and gender. Despite this focus, a simple neural network model trained on standardized EEG time series data achieved an impressive accuracy of 91.8%. This model holds promise for assisting healthcare practitioners in accurate and timely detection of SZ in patients.


# CHAPTER ONE: INTRODUCTION

Schizophrenia is a severe and prolonged brain disorder that disrupts normal thinking, speech, and behavioral characteristics of individuals. The National Institute of Mental Health considers schizophrenia a significant disease, affecting approximately 2.4 million people in the United States over the age of 18. According to the World Health Organization (WHO), approximately 21 million people worldwide are affected by schizophrenia (*Oh et al 2019).* The onset of this disorder typically occurs during youth, with men affected at around 18 years old and women around 25 years old, showing a higher prevalence among males (Sadeghi et al 2021). Men are typically affected by SZ 1.5 times more frequently than women. The annual expenditure on SZ treatment typically ranges between $32.5 and $65 billion, significantly impacting the economy. SZ is rare in young children. According to the National Institute of Mental Health (NIMH), only

1 in every 40,000 children experiences the onset of SZ symptoms before the age of 13. Currently, there are no clinical or physical tests available for diagnosing SZ. Diagnosis relies on identifying a constellation of symptoms that profoundly affect social or occupational functioning. According to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V), SZ has a lifelong prevalence rate of approximately 0.3-0.7%. Symptoms of SZ typically manifest during the mid-teenage years to mid-thirties. (Shalini et al 2021).

The causes of SZ encompass factors such as premature birth, low birth weight, perinatal hypoxia, exposure to intrauterine viruses early in life, and stressors associated with social isolation, migrant status, and urban life in adulthood, which subtly impede brain development. SZ is highly influenced by genetics. Individuals carrying genes such as neurogranin and zinc finger protein 804A have an elevated risk of developing SZ (Oh et al 2019).

Schizophrenia (SZ) is a brain anomaly marked by behavioral symptoms such as hallucinations and disorganized speech. SZ represents a combination of symptoms, which may involve hallucinations, auditory hallucinations, disorganized speech, and functional decline, among other manifestations. Approximately 20-40% of SZ patients attempt suicide at least once, with 5-10% ultimately committing suicide (Oh et al 2019).

In clinics, SZ diagnosis has traditionally relied on EEG readings. However, while effective to a certain extent, this method demands significant time and energy and isn't conducive to accurate diagnoses on a large scale. Subsequently, researchers have introduced a computer-based model that streamlines EEG analysis, reducing workload and accelerating diagnostic speed (Jie et al 2021). Afshin et al 2021 employed classification algorithms that include SVM, KNN, DT, Naïve Bayes, RF, ERT, and bagging. They also employed CNN, LSTM, and CNN-LSTM DL algorithms.

Research gaps in this field include developing deep learning (DL) classification models tailored for different age and gender groups, necessitating access to pertinent data. Additionally, future endeavors may involve leveraging a combination of machine learning (ML) and DL models for SZ diagnosis. This approach would entail extracting various nonlinear features from EEG signals using combinational ML techniques, followed by DL models to extract features directly from raw EEG signals. Subsequently, manual and DL features would be merged for classification

purposes. Utilizing graph models based on DL could also be promising for diagnosing SZ via EEG signals.

This project seeks to address several research gaps by introducing a model that incorporates data on the age and gender groups of subjects. Additionally, most classification models for Schizophrenia rely on foreign data. This project aims to develop a classification model using locally sourced data, making it applicable for use in local neurological centers.

The model proposed in this project leverages the innovation of Deep Learning, particularly suited for analyzing time series data. Among various feature extraction techniques explored, standardization emerged as the most effective. Employing a simple neural network on standardized data yielded a high accuracy of 91.8%. The confusion matrix of this best model indicates 25 true positives, 20 true negatives, 1 false positive, and 3 false negatives.

# CHAPTER TWO: LITERATURE REVIEW

Shu Lih Oh et al 2019, in Deep Convolutional Neural Network (CNN) Model for Automated Diagnosis of Schizophrenia Using EEG Signals, developed two architectures of 11-layered CNN models to classify Schizophrenia based on the EEG data of 14 healthy subjects and 14 SZ patients. An architecture was utilized for subject-based testing while the other was employed for no-subject based testing. The CNN network used in this study was designed using Two Intel Xeon 2.40GHz (E5620) processors with 24GB RAM and the Intel (R) CPU E5-2650 v4 2.20GHz (2 Processors), 384GB RAM and NVIDIA Quadro K4200. Accuracy, sensitivity, specificity and positive predictive value were the metrics of performance and had obtained values of 98.07%, 97.32%, 98.17%, 98.45% and 81.26%, 75.42%, 87.59%, 87.59% respectively for the non-subject based testing and subject-based testing respectively. In spite of the high accuracy of the model, the cost of computation of the CNN model is relatively high. The authors suggested a web based diagnosis.

Carla et al 2021, in Advanced EEG-based learning approaches to predict schizophrenia: promises and pitfalls, conducted a review that provided a critical analysis of classical Machine

Learning and Deep Learning methods to detect Schizophrenia based on EEG signals, published in the last 5 years. *Goshvarpour 2020* utilized the Probabilistic Neural Network (PNN) model on 14 Healthy Control (HC) and 14 Schizophrenia (SZ) patients in subject based testing to get an exceptional accuracy of 100%. *Ye 2017* combined k-Nearest Neighbors (kNN) and Support Vector Machine (SVM) on 10 HC, 10 First Episode in Schizophrenia (FES) and 10 Clinical High Risk (CHR) Patients in subject independent testing to obtain an accuracy of 97.50% for HC vs FES and 77.3% for comparison of the three groups. *Sanots-Mayo 2017* aggregated SVM and Multi Linear Perceptron (MLP) on 16 HC, 31 SZ sample size and 14 subjects (not balanced) test set size to obtain an accuracy of 93.42% for subject independent testing. *Jahmunah 2019* employed SVM-RBF (SVM with Radial Basis Function kernel), SVM-Poly (SVM with a polynomial function kernel), DT (Decision Tree), LDA (Linear Discriminant Analysis), kNN, PNN on 14 HC, 14 SZ sample size and the same test size to get an accuracy of 92.19% in subject dependent testing. *Alimardani 2018* utilized LDA, QDA (Quadratic Discriminant Analysis), SVM, kNN, LRA (Logistic Regression Analysis) on 23 Bipolar Disorder (BD), 23 SZ patients sample size and the same test size. The subject independent testing had an accuracy of 91.30%. *Liu 2018* executed a subject independent testing on 40 HC, 40 FES and 40 CHR patients equal sample size and test size using SVM, RF (Random Forest), NB (Naïve Bayes), DT. The accuracy obtained for HC vs FES testing was 91.16% while 73.13% accuracy was obtained for comparison of the 3 groups. *Li 2019,* utilized LDA and SVM models on 25 HC, 23 SZ equal sample size and test set size and an accuracy of 90.48% for the subject independent test. These seven papers gave the highest accuracies in the range of 90-100%. The promising results obtained from the combination of CNN and LSTM by *Ahmedt-Aristizabal et al* encourage the applications of deep learning to EEG data, in particular to event-related signals, whose features may reflect cognitive, affective, or sensory changes in SZ. Deep Learning models enable the use of raw EEG data instead of EEG features and to learn subtle patterns from EEG data. These models are, however, based on diagnosis and cannot predict the existence of Schizophrenia in patients. They can also not replace standard diagnosis of SZ. It is also difficult to identify the most influential features of the data with Deep Leaning models. It was recommended that future studies should consider looking into comorbidity. Having a well-characterized public SZ EEG database is highly recommended direct comparison and objective evaluation of different algorithms' performances.

Afshin et al 2021 authored Automatic Diagnosis of Schizophrenia in EEG signals. They proposed LSTM, 1D-LSTM, 1D-CNN-LSTM model architectures for the classification of SZ using EEG signals. 1D-CNN models of 9, 3 and 2 Convolutional layers were introduced. LSTM models of 6 and 7 layers are employed. First version CNN-LSTM consists of 11 max, dropout, CNN, LSTM, flatten, pooling and dense layers – 2 convolutional layers, 3 dropout layers, 1 max pooling layer, 1 flatten layer, 1 LSTM layer, 2 dense layers with ReLU and Sigmoid activation functions. The second version of CNN-LSTM, the first 10 layers of this model are identical to those of the previous CNN-LSTM architecture. The dense layer with 50 neurons and the ReLU activation function is used in the 11th layer of this architecture. The 12th layer comprises a dropout with a rate of 0.25. Ultimately, in the 13th layer, the dense layer with a sigmoid activation function for classification is employed. The Dataset of Institute of Psychiatry and Neurology in Warsaw, Poland was used. This dataset includes recorded EEG signals from 14 females and males with ages between 27.9 and 28.3 years. Besides, 14 normal individuals matched with the patients in terms of age and gender were employed in the institution. EEG signals were divided into 25s time frames and then were normalized by z-score or norm-L2. All the experiments of the DL network were conducted using the keras library and using a GPU Nvidia TRX2080 Ti. ML experiments, however, were conducted in an Intel (R) Core (TM) i7-4810MQ CPU at 2.80GHz. The bagging conventional classification algorithms for EEG signals normalized using z-score normalization resulted in the maximum accuracy of 81.22±1.74. The second proposed CNN-LSTM model with the leaky ReLU activation function and combined normalization of z-score with LC resulted in maximum accuracy of 97.73±1.39. The second proposed CNN+LSTM model with ReLU activation function and combined normalization technique of z-score and L2 resulted in the overall maximum accuracy. Despite this high accuracy, limited number of cases in the available EEG datasets for SZ diagnosis has made access to the tools of SZ diagnosis via EEG signals and DL models challenging. These models, also, cannot determine severity neither are they suitable for prognosis or early diagnosis but to diagnose the SZ disorder. They are also not separately designed and compared for different age and gender groups. Recommended further researches include multiclass classifiers by adding classes of brain disorders with similar symptoms to SZ. Preparing datasets for prognosis or early diagnosis, determination of severity and classifying models for different age and gender groups is also a recommended potential future research work. The authors also suggested that CNN+AE

should be considered for future diagnosis models. Combination of ML and DL models for SZ diagnosis such that different non-linear features are extracted from EEG signals first. Afterwards, the features are extracted from raw EEG signals by DL models. Graph models based on DL can be suitable for SZ diagnosis via EEG signals.

# CHAPTER THREE: METHODOLOGY

## DATA ACQUISITION

For this study, EEG data from both healthy individuals and those diagnosed with SZ at the Obafemi Awolowo University Teaching Hospital (OAUTHC) was collected. The EEG data for each participant was formatted in European Data Format (EDF). All EDF files for each participant were consolidated into a single directory, with each participant having their own dedicated subfolder. Additionally, each participant's folder included a GNR file containing individual-specific details.

## PYTHON PROCESSING OF EEG DATA FOR SCHIZOPHRENIA STUDY

Python libraries were utilized to process the EDF and GNR files, extracting their contents. Specifically, Pyedflib's "highlevel" function was employed to extract EEG data from the EDF files. The extracted EEG data and subjects' details from the EDF and GNR files, respectively, were then stored in separate data frames.

During the EEG data extraction process, the median of the 500 recordings per electrode in an EDF file was selected to alleviate high computational demands. This decision reduced the dataset length to 245, ensuring computational efficiency.

It's worth noting that "highlevel" encountered errors while processing some of the EDF files. As a result, the corresponding subjects were completely removed from the dataset, reducing the number of subjects to 27.

## EXPLORATORY DATA ANALYSIS

The data underwent exploration to better understand its characteristics. The EEG data frame extracted from EDF files, labeled as "eeg_data," consists of 25 columns. Among these, 24 columns represent electrodes, while the last column denotes participant ID. The info data frame obtained from the GNR files, named "participants_info," contains various details such as language, first name, date of birth, participant ID, surname, sex, last session ID, category, and age for each subject. The two dataframes – eeg_data and participants_info – were merged through the participantID column which is common to both of them.

To prepare the data for modeling, irrelevant columns including first name, date of birth, surname, and last session ID were dropped from the resulting data frame. Additionally, the participant ID column was removed to avoid uniqueness issues. Categorical columns were converted into binary columns using one-hot encoding. The resulting data set comprises 28 numeric features, with 24 representing continuous electrode data and the remaining 4 representing one-hot encoded categorical data.

```
model_df.head()
```

| | language | sex | category | age | Fp1[1] | Fp2[2] | F3[3] | F4[4] | C3[5] | C4[6] | ... | T5[15] | T6[16] | Fz[17] | Pz[18] | Cz[19] | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 1 | 1 | 44 | -0.714286 | 0.555556 | -1.190476 | 2.301587 | 0.396825 | 0.079365 | ... | 1.507937 | 0.555556 | -3.095238 | 0.873016 | -325.000000 | -3 |
| **1** | 1 | 1 | 1 | 44 | 0.317460 | 1.507937 | -0.238095 | 1.031746 | -0.158730 | 0.396825 | ... | 0.873016 | -1.666667 | -3.412698 | 0.714286 | -325.000000 | -3 |
| **2** | 1 | 1 | 1 | 44 | -1.507937 | 0.238095 | -0.714286 | 1.666667 | 0.476190 | 1.507937 | ... | 1.746032 | 0.714286 | -3.412698 | 0.873016 | -325.000000 | -3 |
| **3** | 1 | 1 | 1 | 44 | -0.714286 | 0.873016 | -0.396825 | 1.031746 | 0.079365 | 0.714286 | ... | 1.666667 | -0.714286 | -3.095238 | 0.873016 | -325.000000 | -3 |
| **4** | 1 | 1 | 1 | 44 | -15.317460 | -15.158730 | -21.825397 | -63.650794 | -23.333333 | -10.873016 | ... | -8.809524 | 2.698413 | -48.650794 | -0.158730 | -15.396825 | |

5 rows × 28 columns

*Figure 1: final model data*

Univariate analysis, including box plots and histograms in Figures 2 and 3, revealed significant outliers in the dataset, which influenced subsequent feature extraction decisions.

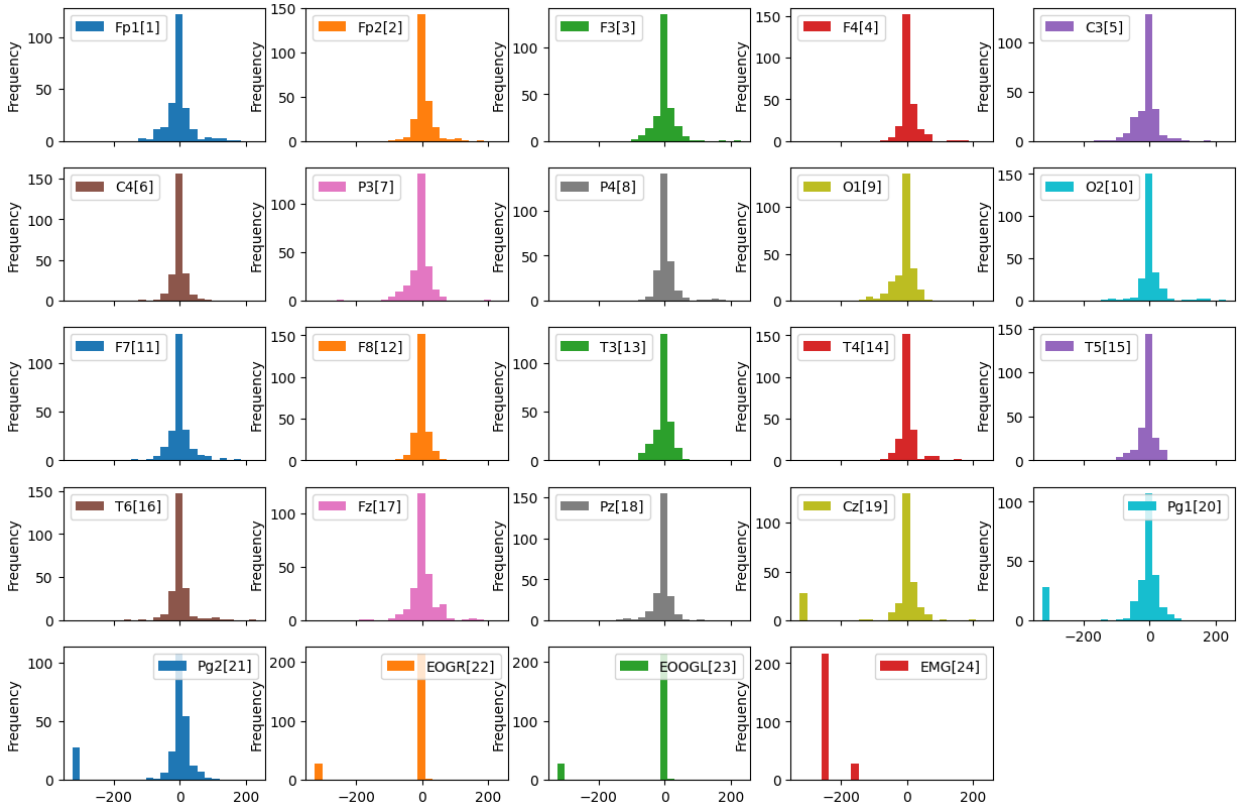


*Figure 2: Histogram of the final model data*

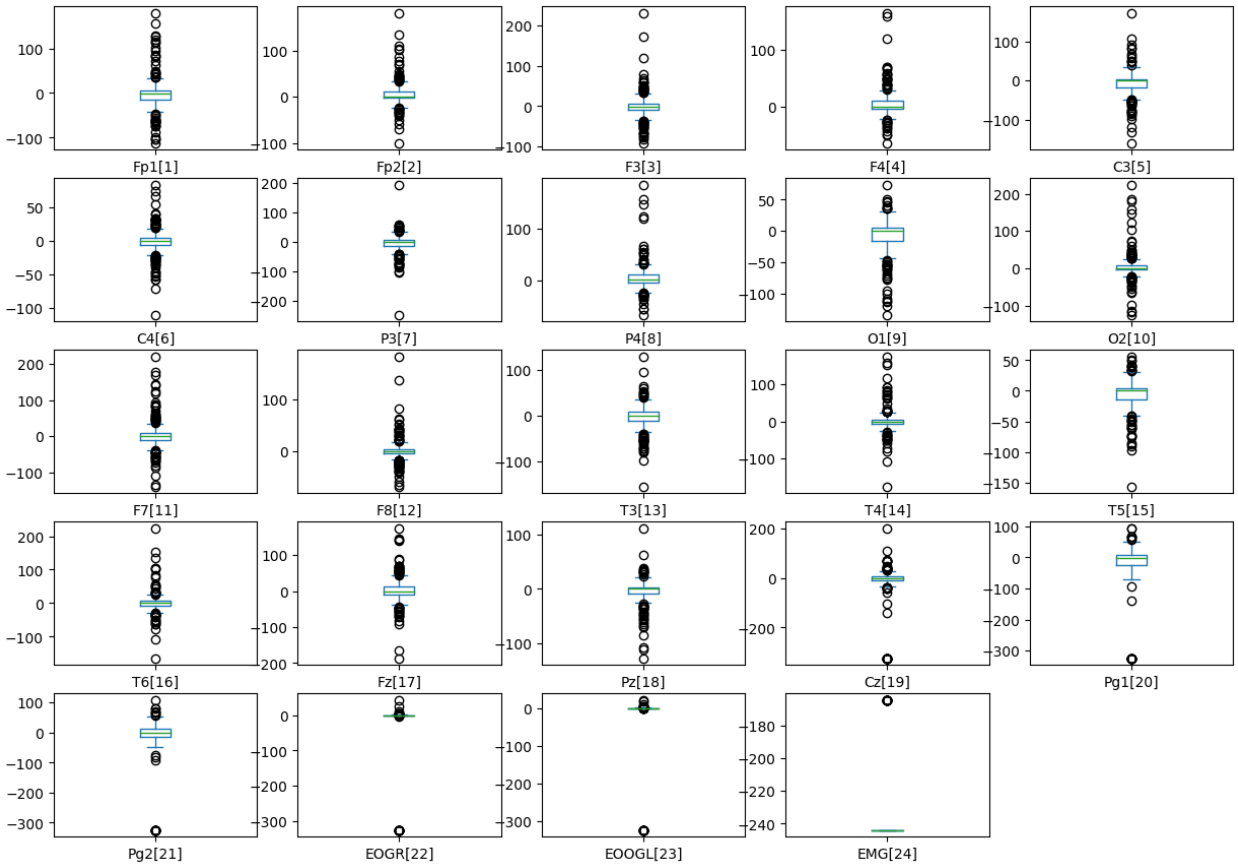


*Figure 3: Box Plot of the Final Model Data*

The pairplot in Figure 4 shows the correlations between each pair of all the features.

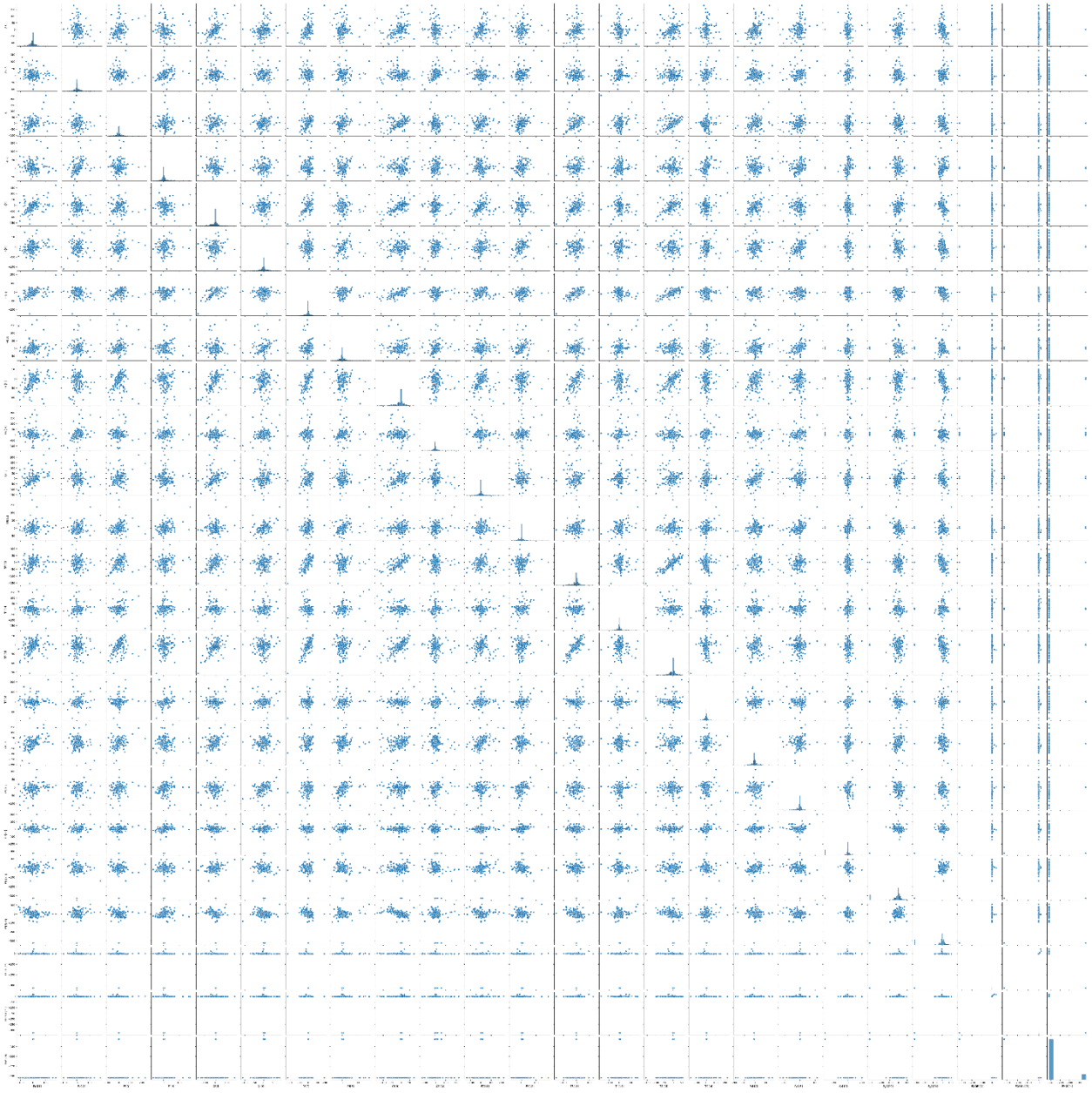

*Figure 4: Pairplot of All Features*

The pie chart in Figure 4 shows that even though, the data is almost distributed equally between the two genders, larger percentage of the subjects speak English and their ages are distributed between 20 and 74. Figure 5 shows that the control and SZ subjects are also almost equal in number

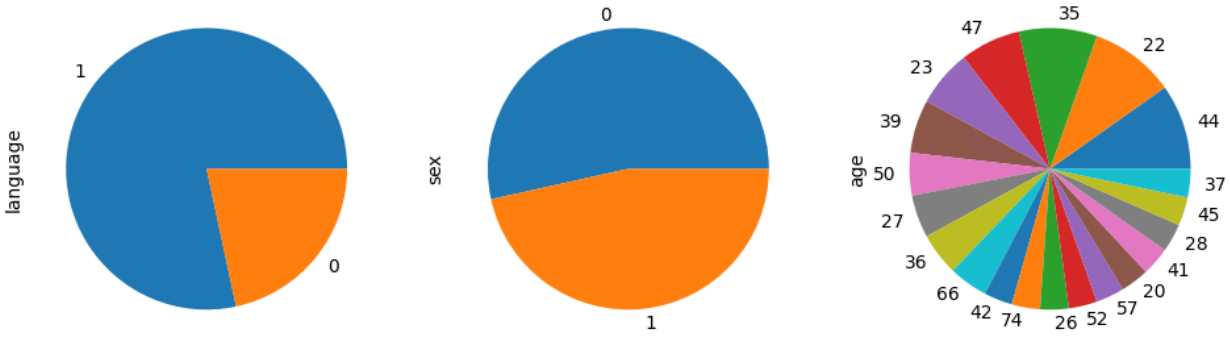


*Figure 5: Pie Chart of Encoded Categorical Features*

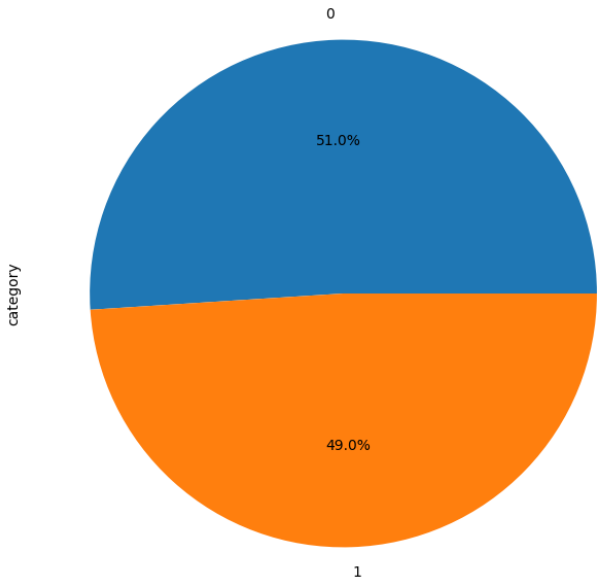


*Figure 6: Pie Chart of the Label Categories*

## FEATURE EXTRACTION

Various feature extraction techniques were attempted, but some, such as p3b, fuzzy entropy, and frequency features, yielded NaN values and were unsuccessful. However, the Tsfresh library provided a valuable solution with its extract_relevant_features function. This function was applied to the standardized time-series data, resulting in 54 relevant features extracted from the data. Subsequently, this set of features was utilized in the quest to find the most accurate model. The first five rows of the extracted relevant features are depicted in Figure 7 below.

```
df.head()
```

| | EOOGL[23]__benford_correlation | Pg1[20]__range_count__max_1__min_-1 | EOGR[22]__benford_correlation | Pg1[20]__root_mean_square | Pg1[20]__absolute_maximum | Pg1 |
|---|---|---|---|---|---|---|
| **0** | 0.295657 | 0.0 | 0.295657 | 2.702415 | 2.702415 | |
| **1** | 0.295657 | 0.0 | 0.295657 | 2.702415 | 2.702415 | |
| **2** | 0.295657 | 0.0 | 0.295657 | 2.702415 | 2.702415 | |
| **3** | 0.295657 | 0.0 | 0.295657 | 2.702415 | 2.702415 | |
| **4** | 0.062915 | 1.0 | 0.062915 | 0.184161 | 0.184161 | |

5 rows × 54 columns

*Figure 7: Extracted Relevant features*

Despite the successful extraction of features using the Tsfresh library, the resulting model accuracy was not optimal. Additional techniques were explored, including normalization, which, as anticipated due to the presence of large outliers, resulted in models with low accuracy. Ultimately, the best accuracy models were attained by training the model on standardized data. Standardization proved effective in scaling data with large outliers, leading to improved model performance.

## CROSS-VALIDATION AND VALIDATION

In attempt to find the best-performing model for the data at hand, some Machine Learning models were cross-validated and validated. Cross-validation refers to the evaluation of the performance of a model using its default hyperparameters. Validation on the other gives room for tuning the hyper parameters of a model. In both cross-validation and validation, the actual training of model will not be carried out. The cross-validated models are logistic regression, support vector classifier, K-Neighbors-Classfier, Random Forest Classifier, Gaussian Naïve Bayes, and Gradient Boosting Classifier. For the validation, logistic regression, support vector classifier, K-Neighbors-Classfier, and Random Forest Classifier were used. Support Vector

Classifier gave the best accuracy out of the validated models. The summary of the best hyperparameter combinations shall be given in the next chapter.

# CHAPTER FOUR: RESULTS AND DISCUSSION

As mentioned in the previous chapter, different models were cross-validated and validated on the EEG time-series data. Figure 8 shows the cross-validation accuracies of the first set of machine-learning models, which are: logistic regression, support vector classifier, KNeighborsClassifier, and RandomForestClassifier. The highest validation accuracy for this round of cross-validation is 62.86% produced by the logistic regression model.

```
Cross Validation accuracies for the LogisticRegression(max_iter=1000) = [0.6122449  0.73469388 0.32653061 0.81632653 0.65306122]
Acccuracy score of the  LogisticRegression(max_iter=1000) = 62.86 %
----------------------------------------------------------------
Cross Validation accuracies for the SVC(kernel='linear') = [0.6122449  0.59183673 0.40816327 0.7755102  0.63265306]
Acccuracy score of the  SVC(kernel='linear') = 60.41 %
----------------------------------------------------------------
Cross Validation accuracies for the KNeighborsClassifier() = [0.48979592 0.75510204 0.3877551  0.44897959 0.28571429]
Acccuracy score of the  KNeighborsClassifier() = 47.35 %
----------------------------------------------------------------
Cross Validation accuracies for the RandomForestClassifier(random_state=0) = [0.51020408 0.7755102  0.46938776 0.3877551  0.42857143]
Acccuracy score of the  RandomForestClassifier(random_state=0) = 51.43 %
----------------------------------------------------------------
```

*Figure 8: Cross-Validation 1 on standardized data*

In attempt to get a higher cross-validation accuracy, the same process was executed on Gaussian Naïve Bayes and Gradient Boosting Classifier. The resulting cross-validation accuracies, as shown in Figure 9, are not better than the previous accuracies. In fact, the highest cross-validation accuracy produced by these two models is 59.18% which is significantly lower than the previous 62.86% produced by the logistic regression model.

```
Cross Validation accuracies for the GaussianNB() = [0.32653061 0.75510204 0.51020408 0.67346939 0.69387755]
Acccuracy score of the  GaussianNB() = 59.18 %
----------------------------------------------------------------
Cross Validation accuracies for the GradientBoostingClassifier() = [0.51020408 0.83673469 0.57142857 0.42857143 0.44897959]
Acccuracy score of the  GradientBoostingClassifier() = 55.92 %
----------------------------------------------------------------
```

*Figure 9: Cross Validation 2 on Standardized Data*

Additionally, efforts were geared toward validation since it allows tuning of hyperparameters unlike cross-validation that uses only the default hyperparameters. Three different models – Logistic Regression (LR), Support Vector Classifier (SVC), KNeighborsClassifier (KNC), and RandomForestClassifier (RFC) – were validated with different sets of hyperparameters on the standardized data. This process of validation gave a higher accuracy as SVC model gave an accuracy of 68.16% setting hyperparameters C and kernel to 1 and sigmoid respectively. The summary of the best hyperparameter combination for each of the models is given in Figure 10.

| | model used | highest score | best hyperparameters |
|---|---|---|---|
| 0 | LogisticRegression(max_iter=10000) | 0.628571 | {'C': 1} |
| 1 | SVC() | 0.681633 | {'C': 1, 'kernel': 'sigmoid'} |
| 2 | KNeighborsClassifier() | 0.546939 | {'n_neighbors': 10} |
| 3 | RandomForestClassifier(random_state=0) | 0.514286 | {'n_estimators': 100} |

*Figure 10: Validation of Standardized Data*

The same cross-validation and validation processes were executed on the extracted relevant features. Figures 11 through 12 show the cross-validation and validation accuracy results. It can be observed that the highest of these accuracies is 64.9%, which was produced by quite a number of model architectures.

.

```
Cross Validation accuracies for the LogisticRegression(max_iter=1000) = [0.59183673 0.69387755 0.53061224 0.73469388 0.69387755]
Acccuracy score of the  LogisticRegression(max_iter=1000) = 64.9 %
----------------------------------------------------------------
Cross Validation accuracies for the SVC(kernel='linear') = [0.59183673 0.69387755 0.53061224 0.73469388 0.69387755]
Acccuracy score of the  SVC(kernel='linear') = 64.9 %
----------------------------------------------------------------
Cross Validation accuracies for the KNeighborsClassifier() = [0.51020408 0.69387755 0.53061224 0.73469388 0.69387755]
Acccuracy score of the  KNeighborsClassifier() = 63.27 %
----------------------------------------------------------------
Cross Validation accuracies for the RandomForestClassifier(random_state=0) = [0.59183673 0.69387755 0.53061224 0.73469388 0.69387755]
Acccuracy score of the  RandomForestClassifier(random_state=0) = 64.9 %
----------------------------------------------------------------
```

*Figure 11: Cross Validation 1 on Relevant Features*

```
Cross Validation accuracies for the GaussianNB() = [0.51020408 0.55102041 0.57142857 0.59183673 0.63265306]
Acccuracy score of the  GaussianNB() = 57.14 %
----------------------------------------------------------------
Cross Validation accuracies for the GradientBoostingClassifier() = [0.59183673 0.69387755 0.53061224 0.73469388 0.69387755]
Acccuracy score of the  GradientBoostingClassifier() = 64.9 %
----------------------------------------------------------------
```

*Figure 12: Cross Validation 2 on relevant features*

| | model used | highest score | best hyperparameters |
|---|---|---|---|
| 0 | LogisticRegression(max_iter=10000) | 0.648980 | {'C': 1} |
| 1 | SVC() | 0.648980 | {'C': 1, 'kernel': 'linear'} |
| 2 | KNeighborsClassifier() | 0.632653 | {'n_neighbors': 5} |
| 3 | RandomForestClassifier(random_state=0) | 0.648980 | {'n_estimators': 10} |

*Figure 13: Validation on Relevant features*

These consistently low accuracies provoked the need for the robust neural network models. A first attempt of a simple neural network gave an astonishing 79.59%. This drastic increase in accuracy and good confusion matrix parameters as shown in figures 14 and 15 bought a strong buy-in for the simple neural network model.

Further permutations were made on the model as it was fed with normalized, standardized, and unscaled data, as well as the extracted relevant features. Table 1 shows the summary of the accuracies of these trials and their confusion matrix parameters.

Table 1: Summary of Simple Neural Network Model Training Trials

*No of epochs = 50*

| Model | Accuracy | True Positive | True Negative | False Positive | False Negative | Standardized | Normalized | No scaling |
|---|---|---|---|---|---|---|---|---|
| Simple Neural Network | 77.55 | 19 | 6 | 19 | 5 | 0 | 0 | 1 |
| | 81.63 | 19 | 6 | 21 | 5 | 0 | 0 | 1 |
| | 77.55 | 17 | 21 | 8 | 3 | 0 | 0 | 1 |
| | 69.39 | 15 | 19 | 4 | 11 | 0 | 0 | 1 |
| | 71.43 | 12 | 23 | 6 | 8 | 0 | 0 | 1 |
| | 85.71 | 20 | 22 | 5 | 2 | 0 | 1 | 0 |
| | 63.75 | 18 | 15 | 5 | 11 | 0 | 1 | 0 |
| | 73.47 | 23 | 13 | 8 | 5 | 0 | 1 | 0 |
| | 71.43 | 17 | 18 | 7 | 7 | 0 | 1 | 0 |
| | 85.71 | 19 | 23 | 6 | 1 | 1 | 0 | 0 |
| | 85.71 | 19 | 23 | 6 | 1 | 1 | 0 | 0 |
| | 91.84 | 25 | 20 | 1 | 3 | 1 | 0 | 0 |
| | 83.67 | 24 | 17 | 3 | 5 | 1 | 0 | 0 |
| | 79.59 | 18 | 21 | 4 | 6 | 1 | 0 | 0 |
| | 89.80 | 19 | 25 | 1 | 4 | 1 | 0 | 0 |
| **Average Values** | | | | | | | | |
| Unscaled | 75.51 | 16.4 | 15 | 11.6 | 6.4 | 0 | 0 | 1 |
| Normalized | 73.59 | 19.5 | 17 | 6.25 | 6.25 | 0 | 1 | 0 |
| Standardized | 86.01 | 20.7 | 21.5 | 3.5 | 3.3 | 1 | 0 | 0 |

From the table, the Simple Neural Network model trained with standardized data produces the highest average accuracy of 86.01 and peak accuracy of 91.84%. By comparison, it also has the highest average true positive and true negative as well as the least false positive and false negative. It can thus be concluded that the best model architecture is a simple neural network trained with standardized time-series data.

EVALUATION OF THE BEST PERFORMING MODEL

The best-performing model – the simple neural network's 91.84% model – was loaded in a new script. The summary of the evaluation is given in Figure 14. The time taken for the prediction is 0.37s. The evaluation accuracy is 89.4%. Since the evaluation was done on the same, it implies that the model works fine on the current acquired dataset.

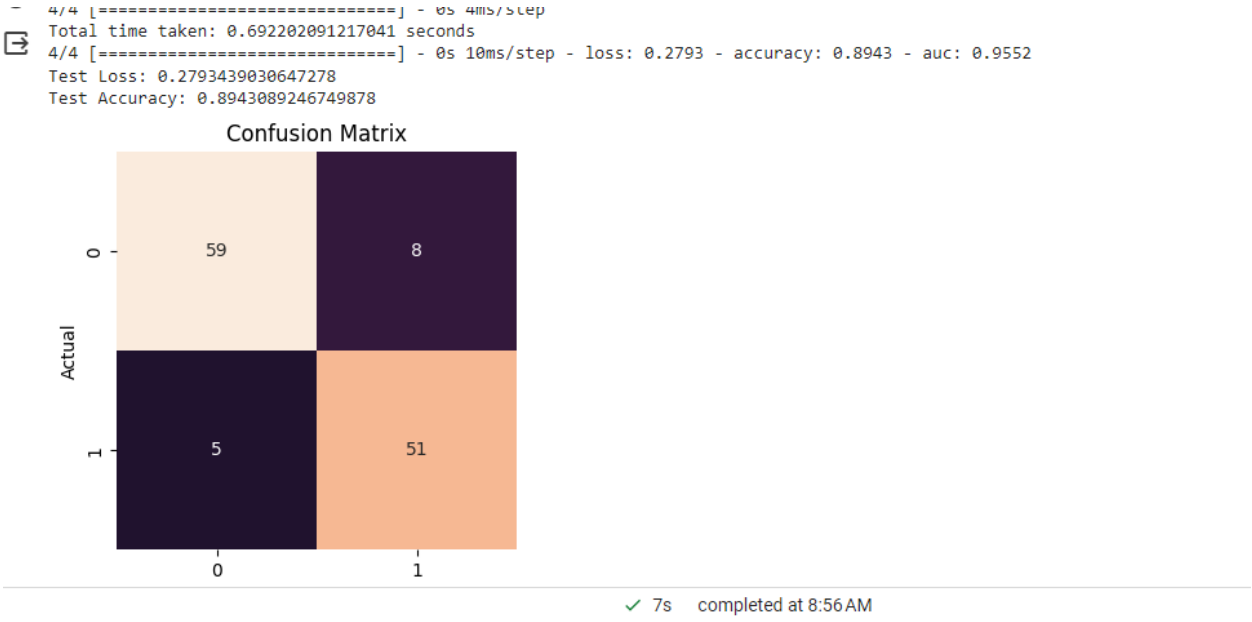


*Figure 14: Summary of Best Model Evaluation*

In an attempt to evaluate the model based on a different data such that there is no probability that the model had seen the model earlier, a new data ingest process was executed with a slight change to the initial data ingest. Instead of choosing the median of the 500 signal readings from each electrode of an edf file, the new data was created by computing the mean of the standardized signal readings. All other columns used for modelling were directly since the number of rows and order is the same with the initial model data. Evaluating the model using this new data gave low accuracies compared to orders of the 80's and 90's. Although, it would been better to get new edf files, pass them through the same ingestion process as the training data, and then use the new data to evaluate the mode, this evaluation also implies that the model is not very robust and might require further optimization in further research.

# CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

This project proposes a high-performing simple neural network model for classification of Schizophrenia subjects. The model's input data is a 27-feature data which consists of 24 timeseries data and 3 encoded categorical data. The timeseries data is the EEG recorded values while the categorical data are language, sex and age. Therefore, this model can be used to make classifications tailored to age and gender groups which is one of the research gaps in this field of research. The major setback of this model is that it was not tested with a new dataset. The robustness of the model cannot be affirmed until it is evaluated with a new dataset.

Further research could be carried out to try more complex neural networks like LSTM, CNN, Bi-LSTM, and other deep learning models for higher accuracies. A combination of machine learning and deep learning models can also be considered in new research.

# REFERENCES