# AN HYBRID APPROACH FOR SCHIZOPHRENIA DISORDER CLASSIFICATION

## 15I720- PROJECT WORK I


**Akshaya B** **18I205**

**Barani Devi E** **18I208**

**Gayathri S** **18I218**


**November 2021**


**DEPARTMENT OF INFORMATION TECHNOLOGY**

**PSG COLLEGE OF TECHNOLOGY**

(Autonomous Institution), Affiliated to Anna University

**COIMBATORE – 641 004**

# PSG COLLEGE OF TECHNOLOGY

(Autonomous Institution), Affiliated to Anna University

## COIMBATORE – 641 004

**AN HYBRID APPROACH FOR SCHIZOPHRENIA DISORDER CLASSIFICATION**

Bonafide record of work done by


**Akshaya B 18I205**

**Barani Devi E 18I208**

**Gayathri S 18I218**


**November 2021**

Sindhu.
(G. SINDHU)

**Ms.Sindhu**
Faculty guide

**Dr.K.Umamaheswari**
Head of the Department

Certified that the candidate was examined in the viva-voce examination held on

**(Internal Examiner)**

**(External Examiner)**

ACKNOWLEDGEMENT

We express our sincere thanks to **Dr.K.Prakasan**, Principal-In-Charge, and PSG College of technology for providing an opportunity to take up this project work.

We are greatly indebted to **Dr.K.Umamaheshwari,** Professor and Head of Department of Information Technology for her kind support and encouragement was given to complete the project.

We are greatly indebted to **Dr.B.Vinoth Kumar,** Associate Professor and Programme Coordinator, Department of Information Technology for his kind support and encouragement given to complete the project.

We also express our humble and profound gratitude to our project guide **Ms.G.Sindhu**, Assistant Professor, Department of Information Technology for her valuable guidance and timely advice to complete the project successfully.

We express our sincere thanks to our tutor **Ms S.Shymala Gowri**, Assistant Professor (Senior Grade), Department of Information Technology, for her constant support and guidance that played a vital role in completing our project

We wish to record our gratitude to all the department faculty members for their support extended in the completion of our project.

# SYNOPSIS

Schizophrenia is a chronic brain disorder that affects a person's thinking ability. The symptoms of Schizophrenia include hallucination, delusions, disorganized speech, catatonic behavior and lack of motivation. There is no specific reason for the cause of Schizophrenia but the possible factors of this disease are physical, genetic, psychological, environmental impacts, drug addiction and inappropriate development of the Brain. Severe condition of this disorder leads to suicide. It is evaluated that one person out of two hundred individuals will be affected by Schizophrenia at some point during their lifetime. Currently, there are no specific laboratory tests or clinical equipment to diagnose Schizophrenia. Machine Learning plays a key role in the healthcare domain where there is a need for mental disorder prediction. The brain state is the main factor to detect Schizophrenia so Magnetic Resonance Imaging (MRI) is used for examining the brain. Datasets are taken COBRE (The Center for Biomedical Research Excellence). These datasets contain patient records with multiscale brain parcellations (fMRI). Data Preprocessing is used to extract Phenotypic Data (all kinds of clinical information regarding patients' disease symptoms, as well as relevant demographic data, such as age, ethnicity, and sex) from Raw fMRI. We have used several Deep learning models. Convolutional Neural Networks (CNNs) is one of the successful image classification and recognition models that results in high accuracy. CNN is made up of hidden layers that receive inputs as MRI images and are used to determine complex patterns. There are several diseases like bipolar disorder, Parkinson's disease which have symptoms similar to Schizophrenia. In these scenarios, highly efficient supervised algorithms are better in classifying Schizophrenia patients from other patients. Out of various Supervised Machine learning Algorithms, Support Vector Machine(SVM) results have high accuracy.Our project aim is to combine both CNN and SVM to classify and predict Schizophrenia patients from normal patients.

# LIST OF FIGURES

# LIST OF TABLES

# CONTENTS

**CHAPTER** **PAGE-NO**

# CHAPTER 1

# INTRODUCTION

## 1.1 PROBLEM DEFINITION

Schizophrenia is a serious psychological disorder that affects a person's thinking ability. The symptoms of Schizophrenia include hallucination, delusions, disorganized speech, disorganized thinking, change in body language and emotions, lack of motivation and change in sleep patterns. There is no specific reason for the cause of Schizophrenia but the possible factors of this disease are physical, genetics, stress, chemical or structural changes of brain environmental impacts, drug use, childhood trauma and inappropriate development of the Brain. Severe condition of this disorder leads to suicide.There is no cure for schizophrenia, so it has to be treated and managed with medication and behavioral therapy. At the same time, we don't have any specific laboratory tests or clinical equipment to diagnose Schizophrenia and another important cause is Overlapping Disorder. This is because many diseases have common Overlapping symptoms similar to Schizophrenia. For eg., Alzheimer's disease, Bipolar Disorder, Huntington's disease, etc.,

## 1.2 OBJECTIVE

Machine learning plays a vital role in the healthcare domain especially in diagnosing and predicting diseases. In our scenario, we make use of various Supervised Machine Learning Algorithms to analyze a Schizophrenia patient's MRI images to predict the disorder. The brain state is the main factor to detect Schizophrenia so Magnetic Resonance Imaging (MRI) is used for examining the brain. By analyzing various patterns, and which disease responds to what kind of treatment, we can arrive at a better solution that determines relevant subtypes of different diseases and which treatments are most efficient to deal with them. Our objective is to analyze the Schizophrenia patients MRI images using various Deep learning models and Supervised Machine Learning Algorithms.

## 1.3 LITERATURE SURVEY

### 1.3.1 SCHIZOPHRENIA CLASSIFICATION USING SVM

Machine learning helps to classify the problem of Schizophrenia data using the Northwestern University Schizophrenia Data. There are three simulations done in this study: SVM with linear, polynomial, and Gaussian kernels.

| **Classification of Schizophrenia Data Using Support Vector Machine (SVM)**<br><br>2018 IOP Conference Series Journal of Physics | |
|---|---|
| Problem Statement | Aims to distinguish people who are Schizophrenics and people who are not |
| Proposed Solution | Simulations with different data and training data's percentage are implemented and Model performance validation is done by averaging ten times Hold-Out Validations. |
| Results | SVM successfully classified Schizophrenics<br>Data with 90.1% accuracy. SVM with linear kernel and Gaussian kernel reached 95.0% accuracy in classifying Schizophrenia data. |
| Merits | For each simulation parameters of the models are optimized with the grid search method, so the evaluation of model performance is obtained with high accuracy |
| Demerits | Using SVM and Gaussian kernels with a 10% training dataset gives less accuracy. |
| Future Work | Recommended using other datasets, kernels, or validation methods. It is guessed that the results obtained from classifying Schizophrenia people or not might help to receive the right treatment for illness. |

**Table 1.1 Schizophrenia Classification using SVM**

### 1.3.2 BRAIN MRI-based 3D CONVOLUTIONAL NEURAL NETWORKS

In this paper, they focused on the potential of CNN for identifying patients with schizophrenia using 3D brain MRI images and found the way for image-based individual level diagnosis and prognosis in psychiatric disorders.

| **Brain MRI-based 3D Convolutional Neural Networks for Classification of Schizophrenia and Control** | |
|---|---|
| Problem Statement | To examine the feasibility of applying CNN to the classification of schizophrenia and controls based on structural Magnetic Resonance Imaging (MRI) |
| Proposed Solution | Built 3D CNN models with a handcrafted feature-based machine learning approach. Support vector machine (SVM) was used as a classifier and Voxel-based Morphometry(VBM) was used as a feature for handcrafted feature-based machine learning. |
| Results | 3D CNN models achieved higher accuracy than handcrafted feature-based machine learning approaches overall. The highest accuracy was obtained by Inception_resnet_1 at 79.27% |
| Merits | Comparing different 3D model architectures trained from scratch, the complex topologies such as inception module and residual module improved the accuracy. |
| Demerits | The sample size is relatively small especially for network training, which might affect the performance of the model and result in lower accuracy and generalizability. |
| Future Work | Future work with large sample size and multi-modal data should be conducted to further improve the accuracy. |

**Table 1.2 Brain MRI based 3D CNN**

### 1.3.3 MRI FINDINGS IN PATIENTS WITH SCHIZOPHRENIA

Brain atrophy was the most commonly seen brain change in the studied sample of patients with schizophrenia. MRI brain can be used to identify structural abnormalities in patients with schizophrenia.

| **Magnetic Resonance Imaging Findings in Patients with Schizophrenia** | |
|---|---|
| Problem Statement | To determine structural abnormalities in the brain of patients with schizophrenia by Magnetic Resonance Imaging (MRI). |
| Proposed Solution | Screening Magnetic Resonance Imaging (MRI) of the brain was done in order to see structural changes in brain matter and to measure the onset and course of psychosis ‘Interview for Retrospective Assessment of Age at Onset in Schizophrenia’ was applied (IROAS) |
| Results | A total of 66 MRI films studied for brain abnormalities, brain atrophy, presence of septum pallidum, and enlarged Virchow-Robins spaces were significantly associated with schizophrenia ($p < 0.001$). |
| Merits | Brain atrophy was the most commonly seen brain change in the studied sample of patients with schizophrenia. |
| Demerits | They had no primary psychiatric or neurological disorder and Patients with schizoaffective disorder or comorbid substance abuse, mental retardation, and organic mental disorders were excluded from the study. |
| Future Work | Using MRI various psychiatric disorders can be predicted. |

**Table 1.3 MRI Findings in Patients with Schizophrenia**

# CHAPTER 2
# DATASET

## 2.1 SOURCE

- COBRE (The Center for Biomedical Research Excellence)

COBRE concentrates on the brain processes, decision making and activities. The COBRE Center (Central Nervous System Function) is a unit of the Carney Institute for Brain Science. Human behavior needs attention, decisions, and all common functions are mediated by brain networks located in the neocortex but modulated by sub-cortical processing. Behavioral and brain functionalities of attention, orienting and action selection, are important gateways into high-level function. Generally, attention, decision making and the ensuing actions describe human mental activities. Deficiency in these functions is basic in both neurological and psychiatric disorders and can arrive at a wide range of higher-order behavioral deficits.

- Nilearn

It enables approachable and versatile analyses of brain volumes. It provides statistical and machine-learning tools, with instructive documentation & open community

## 2.2 ATTRIBUTES

- Phenotypic Data

Phenotypic data contains clinical information regarding a patient's disease symptoms. And also it has similar demographic data, such as age, gender, id and ethnicity. This information is collected and stored in patient registries and biobanks.

- Multiscale brain parcellation Data (Resting fMRI)

An MRI scan has magnetic fields and sound waves which are used to create two or three-dimensional images. This provides a good view of the structure of the brain and helps to predict schizophrenia by detecting abnormalities that may be causing symptoms similar to schizophrenia.

## 2.3 DETAILS

72 patients with Schizophrenia and 75 healthy controls (ages ranging from 18 to 65 in each group). All subjects were screened and excluded if they had; history of neurological disorder, history of mental retardation, history of severe head trauma with more than 5 minutes loss of consciousness, history of substance abuse or dependence within the last 12 months. Diagnostic information was collected using the Structured Clinical Interview used for DSM Disorders (SCID)

## 2.4 DATA FOR EVERY PARTICIPANT

- Resting fMRI
- Phenotypic data for every participant

| | id | current_age | gender | handedness | subject_type | diagnosis | frames_ok | fd | fd_scrubbed | path |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 40061 | 18 | Male | Right | Control | None | 133 | 0.25512 | 0.22657 | /root/nilearn_data/cobre/fmri_0040046.nii.gz |
| **1** | 40090 | 18 | Female | Right | Control | None | 150 | 0.16963 | 0.16963 | /root/nilearn_data/cobre/fmri_0040002.nii.gz |
| **2** | 40046 | 18 | Male | Left | Patient | 295.70 depressed type | 76 | 0.37504 | 0.30042 | /root/nilearn_data/cobre/fmri_0040117.nii.gz |
| **3** | 40002 | 19 | Male | Right | Patient | 295.3 | 67 | 0.40006 | 0.21575 | /root/nilearn_data/cobre/fmri_0040145.nii.gz |
| **4** | 40117 | 19 | Male | Right | Patient | 295.3 | 133 | 0.20975 | 0.18410 | /root/nilearn_data/cobre/fmri_0040000.nii.gz |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **141** | 40089 | 62 | Male | Right | Patient | 295.3 | 40 | 0.70368 | 0.72439 | /root/nilearn_data/cobre/fmri_0040144.nii.gz |
| **142** | 40040 | 63 | Male | Right | Patient | 295.3 | 42 | 0.58301 | 0.40646 | /root/nilearn_data/cobre/fmri_0040069.nii.gz |
| **143** | 40028 | 64 | Male | Right | Patient | 295.3 | 55 | 0.42364 | 0.26393 | /root/nilearn_data/cobre/fmri_0040111.nii.gz |
| **144** | 40086 | 65 | Male | Right | Control | None | 48 | 0.39595 | 0.32296 | /root/nilearn_data/cobre/fmri_0040066.nii.gz |
| **145** | 40007 | 65 | Female | Right | Patient | 295.3 | 40 | 0.70044 | 0.72077 | /root/nilearn_data/cobre/fmri_0040086.nii.gz |

146 rows × 10 columns

**Fig 2.1 Schizophrenia Dataset**

# CHAPTER 3

# METHODOLOGY

## 3.1 FLOW CHART

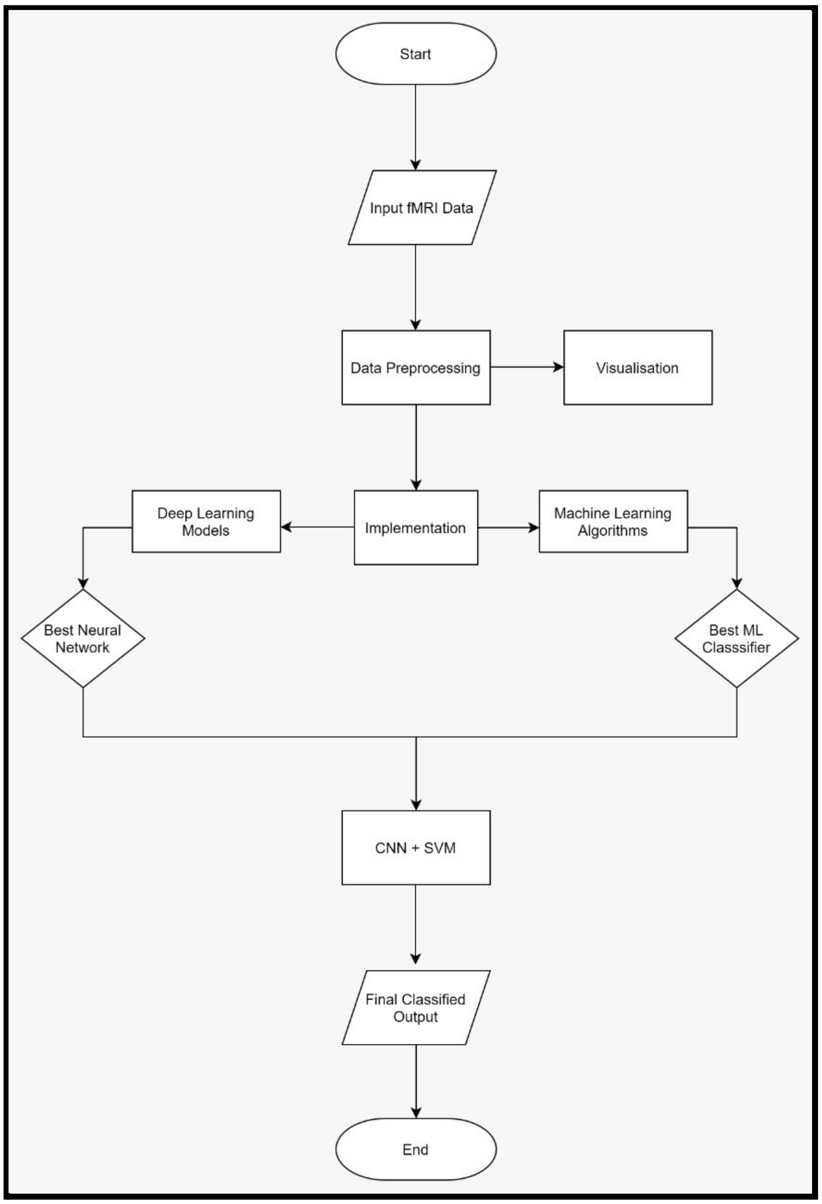



**Fig 3.1 Flow Chart**

First, we will load our dataset in google Colaboratory from COBRA and nilearn. Phenotypic data and MRI images are separately loaded. Then preprocessing is done on the dataset. In preprocessing we will merge the MRI image path with phenotypic data. Next, data visualization is done to understand the features available in the dataset which may improve accuracy. We implemented various Deep learning models and finalized the Convolutional Neural Network which gave high accuracy and figured that CNN fits for image classification as we are using MRI images. Then we tested our dataset on various Machine Learning algorithms like SVM, KNN, Random Forest, and Gradient Booster model. We choose the best model based on accuracy. The selected model with higher accuracy is used as a classifier in the last layer of CNN which gave good results.

## 3.2 MODULES OF THE PROJECT

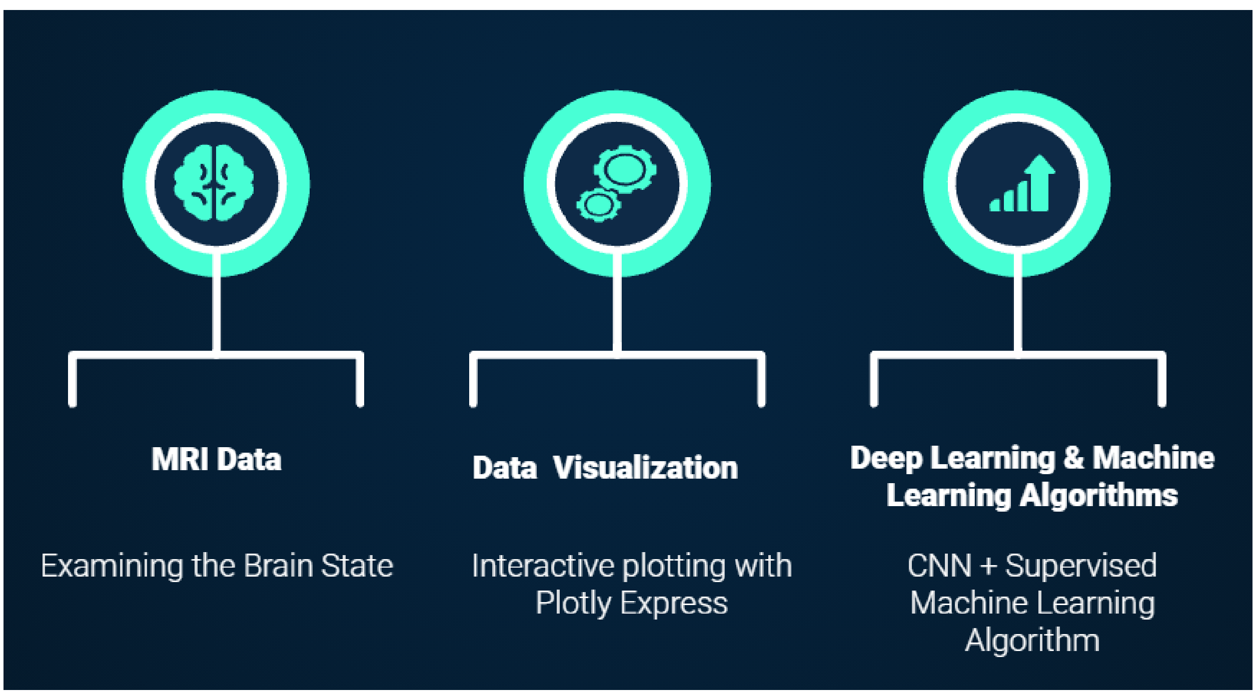



**Fig 3.2 Modules of the Project**

We used three different modules for the project. MRI Data is widely used in Schizophrenia prediction because it helps to examine the brain state. Data Visualization is used for better understanding feature selection and decides the importance of features. It is done using Interactive plotting with Plotly Express App. The next module is Machine Learning which plays an important part in Schizophrenia diagnosis. We used various Classifiers for selecting models with the best accuracy. The third important module is Deep learning which plays a major role in Image Classification

**Support Vector Machine Classifier**

SVC with parameter kernel='linear' implemented in terms of liblinear rather than libsvm, so it has more flexibility in the choice of penalties and loss functions and it should scale better to large numbers of samples. Also implemented using various kernels such as RBF, linear, poly, sigmoid kernels.

**Decision Tree Classifier**

For the Decision Tree Algorithm, we used Gradient Boost Classifier and Random Forest. GBT builds trees one at a time, where each new tree helps to correct errors made by previously trained trees whereas RF trains each tree independently, using a random sample of the data.

**K Nearest Neighbour Classifier**

The KNN which is a non-parametric classification method used for classification and regression is also implemented with a K value of 5.

**Multi-Layer Perceptron**

A multilayer perceptron (MLP) is a class of feedforward Artificial Neural Networks (ANN). MLP utilizes a supervised learning technique called backpropagation for training.

**Convolutional Neural Network**

A convolutional neural network (CNN) is a type of ANN used in image recognition and processing that is specifically designed to process pixel data.

# CHAPTER 4

# IMPLEMENTATION

## 4.1 DATA INTERPRETATION AND VISUALISATION

We Downloaded Raw fMRI Dataset from COBRE which includes multi-scale functional brain parcellations and Extracted Phenotypic data (medical-based data) from Raw fMRI Dataset by Pre-Processing. Then started creating a Data Frame for phenotypic data, extracted paths from data and stored it separately and Merged two Data frames and stored the result Next we created two Separate lists for patients and controls and generated a plotly express app to show the age distribution of subjects in the data. Data Visualization is done using plotly express app for patients and controls

| | id | current_age | gender | handedness | subject_type | diagnosis | frames_ok | fd | fd_scrubbed | path | features | file |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 40061 | 18 | Male | Right | Control | None | 133 | 0.25512 | 0.22657 | /root/nilearn_data/cobre/fmri_0040046.nii.gz | [0.08678037911431263, 0.06380639929294542, 0.2... | /root/nilearn_data/cobre/fmri_0040046.nii.gz |
| 1 | 40090 | 18 | Female | Right | Control | None | 150 | 0.16963 | 0.16963 | /root/nilearn_data/cobre/fmri_0040002.nii.gz | [0.14562583490409123, -0.063139770457594, 0.05... | /root/nilearn_data/cobre/fmri_0040002.nii.gz |
| 2 | 40046 | 18 | Male | Left | Patient | 295.70 depressed type | 76 | 0.37504 | 0.30042 | /root/nilearn_data/cobre/fmri_0040117.nii.gz | [0.17462308889198713, -0.11825290188442354, -0... | /root/nilearn_data/cobre/fmri_0040117.nii.gz |
| 3 | 40002 | 19 | Male | Right | Patient | 295.3 | 67 | 0.40006 | 0.21575 | /root/nilearn_data/cobre/fmri_0040145.nii.gz | [-0.030583214196109036, -0.034635104115588664,... | /root/nilearn_data/cobre/fmri_0040145.nii.gz |
| 4 | 40117 | 19 | Male | Right | Patient | 295.3 | 133 | 0.20975 | 0.18410 | /root/nilearn_data/cobre/fmri_0040000.nii.gz | [-0.005440252066354106, -0.0004542389777897926... | /root/nilearn_data/cobre/fmri_0040000.nii.gz |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 141 | 40089 | 62 | Male | Right | Patient | 295.3 | 40 | 0.70368 | 0.72439 | /root/nilearn_data/cobre/fmri_0040144.nii.gz | [0.23066576780245926, 0.12127764968872, 0.246... | /root/nilearn_data/cobre/fmri_0040144.nii.gz |
| 142 | 40040 | 63 | Male | Right | Patient | 295.3 | 42 | 0.58301 | 0.40646 | /root/nilearn_data/cobre/fmri_0040069.nii.gz | [0.049459034002268934, -0.13945100378772768, 0... | /root/nilearn_data/cobre/fmri_0040069.nii.gz |
| 143 | 40028 | 64 | Male | Right | Patient | 295.3 | 55 | 0.42364 | 0.26393 | /root/nilearn_data/cobre/fmri_0040111.nii.gz | [-0.10218465716657227, -0.1638331558659602, 0.... | /root/nilearn_data/cobre/fmri_0040111.nii.gz |
| 144 | 40086 | 65 | Male | Right | Control | None | 48 | 0.39595 | 0.32296 | /root/nilearn_data/cobre/fmri_0040066.nii.gz | [0.06591001691077762, 0.1350120773424914, 0.08... | /root/nilearn_data/cobre/fmri_0040066.nii.gz |
| 145 | 40007 | 65 | Female | Right | Patient | 295.3 | 40 | 0.70044 | 0.72077 | /root/nilearn_data/cobre/fmri_0040086.nii.gz | [0.44868677474051166, -0.19578477957852558, -0... | /root/nilearn_data/cobre/fmri_0040086.nii.gz |

146 rows × 12 columns

**Fig 4.1 Schizophrenia Dataset with Features**

fig 4.1 shows a dataset for schizophrenia. It contains 147 rows and 10 columns which includes phenotypic data and path id. Phenotypic data contains id, age, gender, handedness which are independent variables and subject type which is a dependent variable that tells the subject is control or patient. Also, path id for which contains the path for the fmri image of the subject and features are extracted from fMRI images.

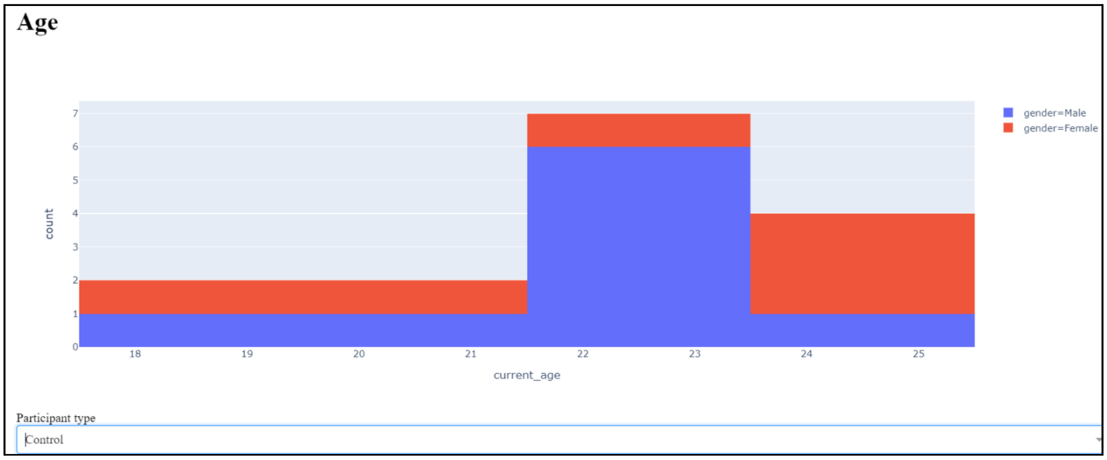



**Fig 4.2 Age Distribution for healthy Controls**

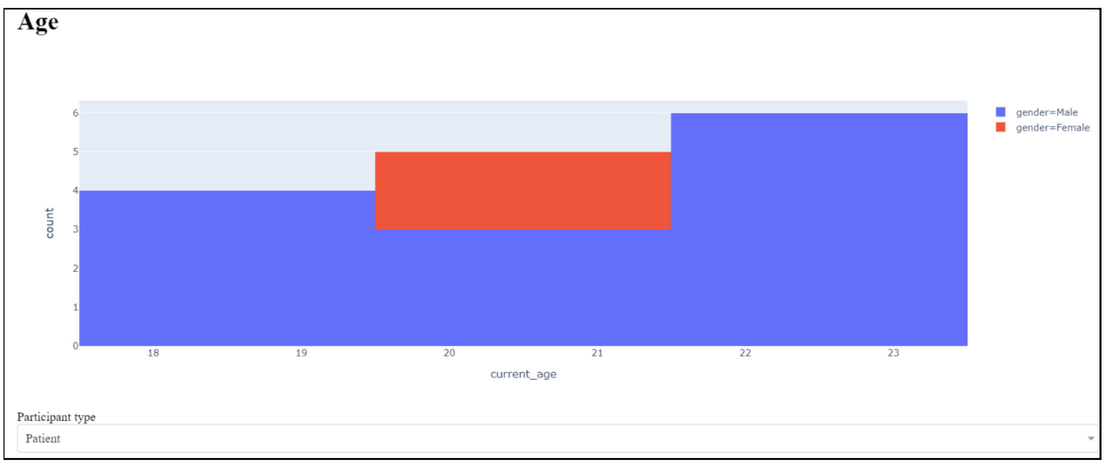



**Fig 4.3 Age Distribution for Schizophrenia patients**

In fig 4.2 and 4.3, the current age is on the X-axis and the count of the subjects is on the Y-axis. The plot shows the total count of females which indicates red color and males in blue color according to age in the X-axis. In the above plot, we can conclude for healthy control and patients for males the age is between 23 to 25 and 21 to 23. For females 21 to 23 are controls and 23 to 25 are patients.

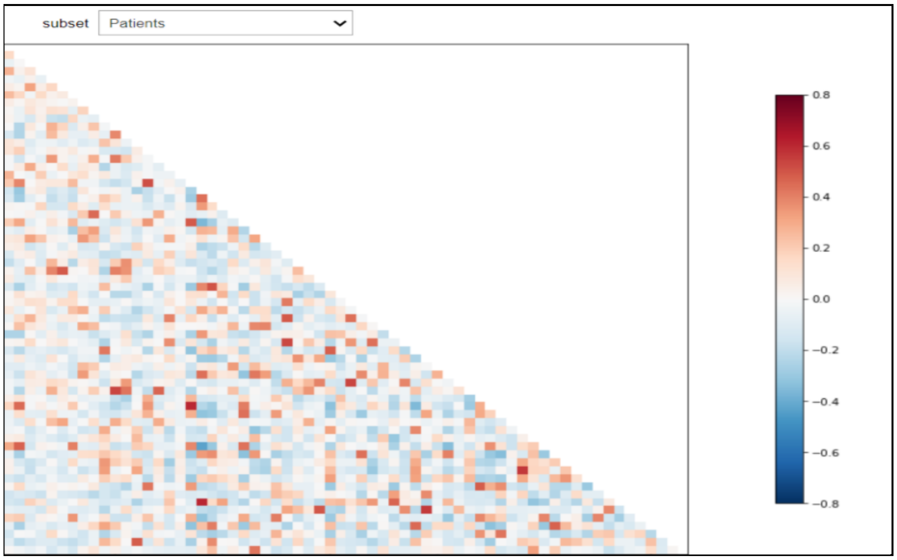



**Fig 4.4 Average correlation matrix for Schizophrenia patients**

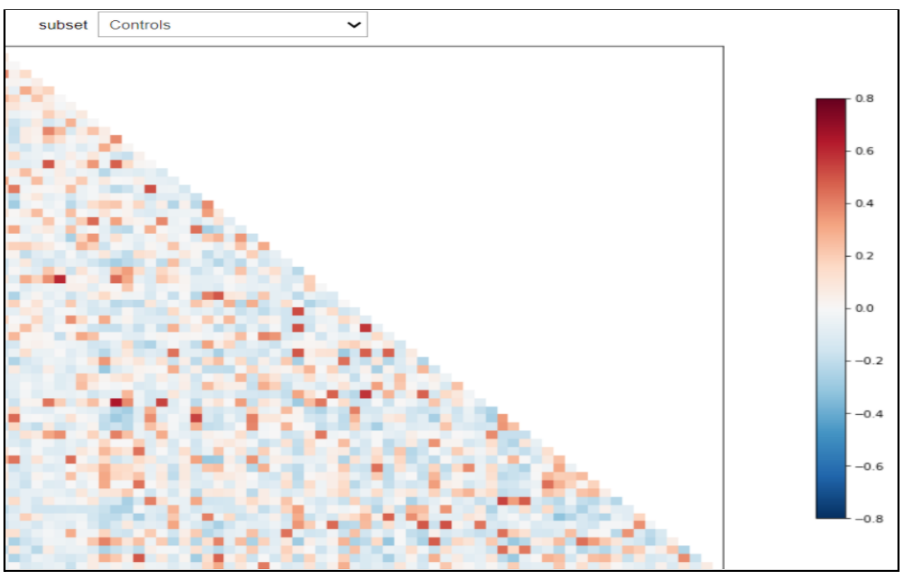



**Fig 4.5 Average correlation matrix for Healthy Control**

A correlation matrix is used to summarize data. A correlation matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables. Fig 4.4 shows the correlation matrix for the control type which is patient and Fig 4.5 shows the correlation matrix for the control type which is healthy.

## 4.2 MACHINE LEARNING CLASSIFIERS

The schizophrenia dataset contains 147 subjects in total. The dataset split in such a way that 80% for training and 20% for testing. Approximately, 117 subjects for training and 30 subjects for testing. We used the below mentioned Supervised Machine Learning Algorithms to predict and classify schizophrenia patients from others.

1. Support Vector Machine Classifier with various kernels
2. K Nearest Neighbors Classifier
3. Gradient Boost Classifier
4. Random Forest Classifier

### 4.2.1 SUPPORT VECTOR MACHINE CLASSIFIER

**SVM with Radial Basis Function(RBF) Kernel**

The fig represents the Confusion Matrix built using SVM with RBF Kernel. Here, the X-axis represents the Predicted label and the Y-axis represents the True label. Out of 30 subjects, 11 controls and 6 patients were predicted correctly. The remaining 4 controls and 9 patients were predicted wrongly. So this classifier gives an accuracy of 56.66%.

```
predictions = svc_rbf.predict(x_test)
print(predictions)
print("Accuracy = ",metrics.accuracy_score(y_test, predictions)*100)

['Control' 'Patient' 'Patient' 'Control' 'Patient' 'Control' 'Control'
 'Patient' 'Control' 'Control' 'Patient' 'Patient' 'Control' 'Patient'
 'Control' 'Control' 'Control' 'Control' 'Control' 'Patient' 'Control'
 'Control' 'Patient' 'Control' 'Control' 'Control' 'Control' 'Patient'
 'Control' 'Control']
Accuracy =  56.666666666666664
```

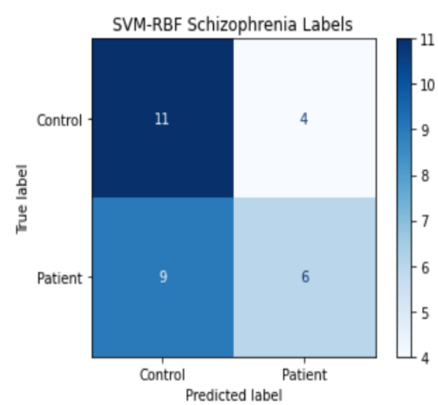



**Fig 4.6 Predicted Output and Confusion Matrix for SVM-RBF Kernel**

## SVM with Linear Kernel

The fig represents the Confusion Matrix built using SVM with Linear Kernel. Here, the X-axis represents the Predicted label and the Y-axis represents the True label. Out of 30 subjects, 11 controls and 8 patients were predicted correctly. The remaining 4 controls and 7 patients were predicted wrongly. So this classifier gives an accuracy of 63.33%.

```
predictions = svc_linear.predict(x_test)
print(predictions)
print("Accuracy = ",metrics.accuracy_score(y_test, predictions)*100)

['Patient' 'Patient' 'Patient' 'Control' 'Patient' 'Control' 'Control'
 'Control' 'Control' 'Control' 'Patient' 'Patient' 'Control' 'Patient'
 'Control' 'Patient' 'Control' 'Control' 'Control' 'Patient' 'Control'
 'Control' 'Patient' 'Control' 'Control' 'Control' 'Control' 'Patient'
 'Patient' 'Control']
Accuracy =  63.33333333333333
```

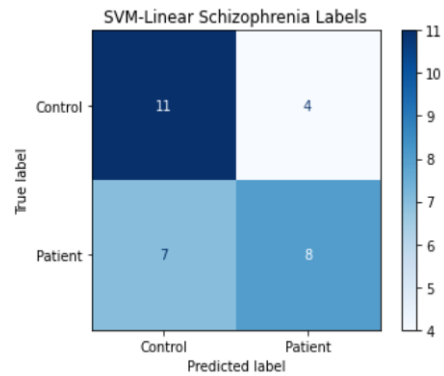



**Fig 4.7 Predicted Output and Confusion Matrix for SVM-Linear Kernel**

## SVM with Sigmoid Kernel

The fig represents the Confusion Matrix built using SVM with Sigmoid Kernel. Here, the X-axis represents the Predicted label and the Y-axis represents the True label. Out of 30 subjects, 15 controls and 4 patients were predicted correctly. The remaining 0 controls and 11 patients were predicted wrongly. So this classifier gives an accuracy of 63.33%.

```
predictions = svc_sig.predict(x_test)
print(predictions)
print("Accuracy = ",metrics.accuracy_score(y_test, predictions)*100)

['Control' 'Patient' 'Patient' 'Control' 'Control' 'Control' 'Control'
 'Control' 'Control' 'Control' 'Control' 'Patient' 'Control' 'Control'
 'Control' 'Control' 'Control' 'Control' 'Control' 'Control' 'Control'
 'Control' 'Control' 'Control' 'Control' 'Control' 'Control' 'Patient'
 'Control' 'Control']
Accuracy =  63.33333333333333
```

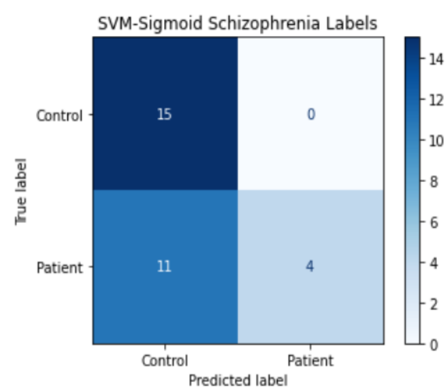



**Fig 4.8 Predicted Output and Confusion Matrix for SVM-Sigmoid Kernel**

### SVM with Poly Kernel

The fig represents the Confusion Matrix built using SVM with Poly Kernel. Here, the X-axis represents the Predicted label and the Y-axis represents the True label. Out of 30 subjects, 9 controls and 7 patients were predicted correctly. The remaining 6 controls and 8 patients were predicted wrongly. So this classifier gives an accuracy of 53.33%.

```
predictions = svc_poly.predict(x_test)
print(predictions)
print("Accuracy = ",metrics.accuracy_score(y_test, predictions)*100)

['Control' 'Patient' 'Patient' 'Control' 'Patient' 'Patient' 'Control'
 'Patient' 'Control' 'Control' 'Patient' 'Patient' 'Control' 'Control'
 'Control' 'Control' 'Patient' 'Control' 'Control' 'Patient' 'Control'
 'Control' 'Patient' 'Patient' 'Control' 'Control' 'Control' 'Patient'
 'Patient' 'Control']
Accuracy =  53.333333333333336
```

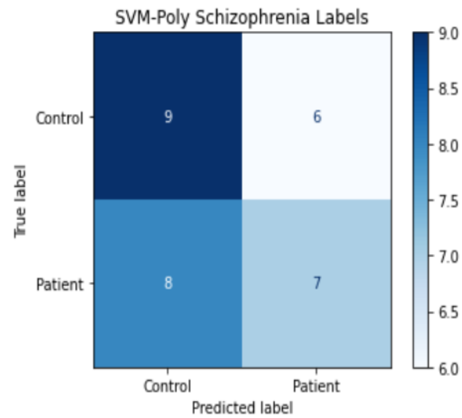



**Fig 4.9 Predicted Output and Confusion Matrix for SVM-Poly Kernel**

### 4.2.2 K NEAREST NEIGHBOR (KNN) CLASSIFIER

The fig represents the Confusion Matrix built using KNN with K value as 5. Here, the X-axis represents the Predicted label and the Y-axis represents the True label. Out of 30 subjects, 6 controls and 7 patients were predicted correctly. The remaining 9 controls and 8 patients were predicted wrongly. So this classifier gives an accuracy of 43.33%.

```
predictions = knn.predict(x_test)
print(predictions)
print("Accuracy = ",metrics.accuracy_score(y_test, predictions)*100)

['Patient' 'Control' 'Control' 'Patient' 'Patient' 'Patient' 'Control'
 'Patient' 'Patient' 'Control' 'Patient' 'Patient' 'Control' 'Control'
 'Control' 'Control' 'Patient' 'Patient' 'Control' 'Patient' 'Control'
 'Control' 'Patient' 'Patient' 'Control' 'Patient' 'Patient' 'Patient'
 'Control' 'Control']
Accuracy =  43.333333333333336
```

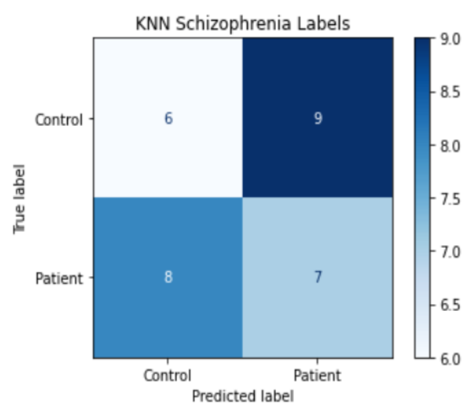



**Fig 4.10 Predicted Output and Confusion Matrix for KNN**

### 4.2.3 GRADIENT BOOST CLASSIFIER

The fig represents the Confusion Matrix built using Gradient Boost Classifier. Here, the X-axis represents the Predicted label and the Y-axis represents the True label. Out of 30 subjects, 7 controls and 8 patients were predicted correctly. The remaining 8 controls and 7 patients were predicted wrongly. So this classifier gives an accuracy of 50%.

```
predictions = boost.predict(x_test)
print(predictions)
print("Accuracy = ",metrics.accuracy_score(y_test, predictions)*100)

['Patient' 'Patient' 'Patient' 'Patient' 'Patient' 'Control' 'Patient'
 'Patient' 'Control' 'Control' 'Patient' 'Patient' 'Control' 'Patient'
 'Control' 'Control' 'Control' 'Patient' 'Control' 'Patient' 'Patient'
 'Control' 'Control' 'Control' 'Control' 'Control' 'Control' 'Patient'
 'Patient' 'Patient']
Accuracy =  50.0
```

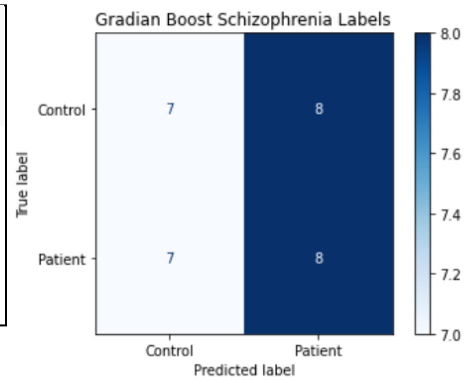
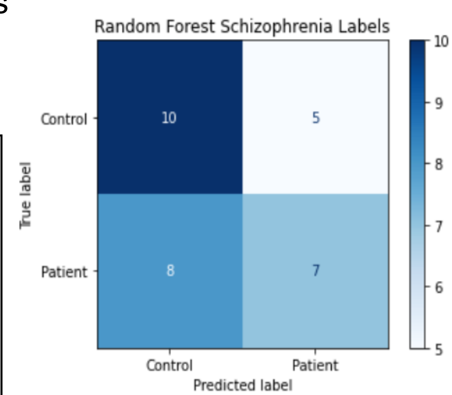



**Fig 4.11 Predicted Output and Confusion Matrix for Gradient Boost Classifier**

### 4.2.4 RANDOM FOREST CLASSIFIER

The fig represents the Confusion Matrix built using Gradient Boost Classifier. Here, the X-axis represents the Predicted label and the Y-axis represents the True label. Out of 30 subjects, 10 controls and 7 patients were predicted correctly. The remaining 5 controls and 8 patients were predicted wrongly. So this classifier gives an accuracy of 56.66%.




```
predictions = rfc.predict(x_test)
print(predictions)
print("Accuracy = ",metrics.accuracy_score(y_test, predictions)*100)

['Control' 'Patient' 'Patient' 'Control' 'Patient' 'Patient' 'Control'
 'Patient' 'Control' 'Control' 'Patient' 'Patient' 'Control' 'Control'
 'Control' 'Patient' 'Control' 'Control' 'Control' 'Patient' 'Control'
 'Patient' 'Control' 'Control' 'Control' 'Control' 'Control' 'Patient'
 'Patient' 'Control']
Accuracy =  56.666666666666664
```

**Fig 4.12 Predicted Output and Confusion Matrix for Random Forest Classifier**

## 4.3 PERFORMANCE EVALUATION FOR ML CLASSIFIERS

The evaluation metrics are used for evaluating the performance of a machine learning model. It aims to estimate the generalization accuracy of a model on the future (out-of-sample) data.

**For Patients:**

| Classifiers | Precision | Recall | F1 Score | Support |
|---|---|---|---|---|
| SVM - RBF | 0.60 | 0.40 | 0.48 | 15 |
| SVM - Linear | 0.67 | 0.53 | 0.59 | 15 |
| SVM - Poly | 0.54 | 0.47 | 0.50 | 15 |
| SVM - Sigmoid | 1.00 | 0.27 | 0.42 | 15 |
| KNN | 0.44 | 0.47 | 0.45 | 15 |
| Gradient Boost | 0.50 | 0.53 | 0.52 | 15 |
| Random Forest | 0.58 | 0.47 | 0.52 | 15 |

**Table 4.1 Performance Measures for Patients**

**For Controls:**

| Classifiers | Precision | recall | F1 Score | Support |
|---|---|---|---|---|
| SVM - RBF | 0.55 | 0.73 | 0.63 | 15 |
| SVM - Linear | 0.61 | 0.73 | 0.67 | 15 |
| SVM - Poly | 0.53 | 0.60 | 0.56 | 15 |
| SVM - Sigmoid | 0.58 | 1.00 | 0.73 | 15 |
| KNN | 0.43 | 0.40 | 0.41 | 15 |
| Gradient Boost | 0.50 | 0.47 | 0.48 | 15 |
| Random Forest | 0.56 | 0.67 | 0.61 | 15 |

**Table 4.2 Performance Measures for Controls**

## 4.4 TRAIN AND TEST SCORE FOR VARIOUS ML CLASSIFIERS

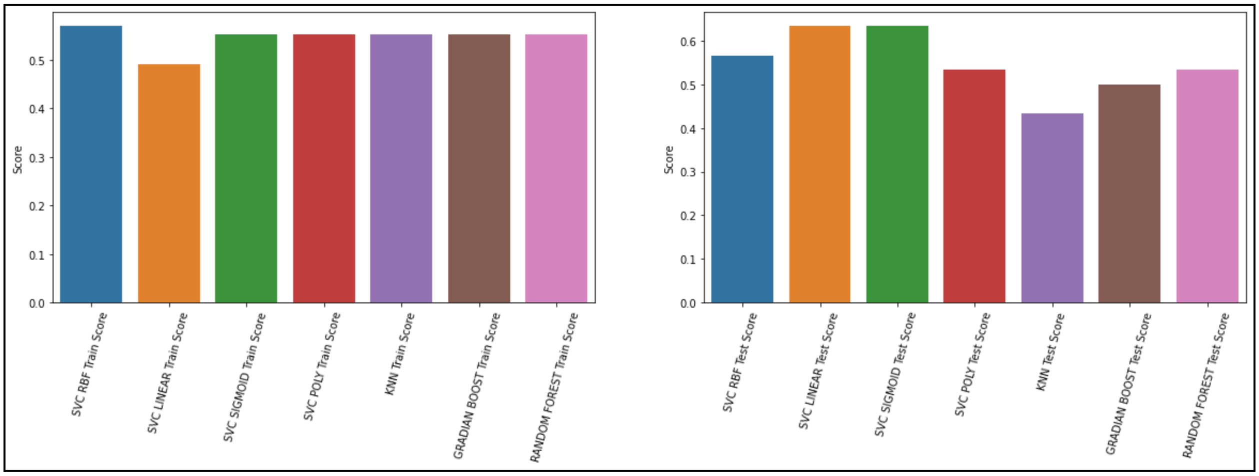



**Fig 4.13 Train and Test Score Results**

In Fig 4.13 Training and Testing Score for Various ML Classifiers is shown. The X-axis represents each classifier's Training and Testing Results respectively. Y-axis represents the accuracy score. Out of all classifiers, we got SVC with high accuracy. So we applied SVC at the last layer of the Deep learning model.

## 4.5 DEEP LEARNING MODELS

### 4.5.1 MULTI-LAYER PERCEPTRONS (MLP) WITH DIFFERENT HYPERPARAMETER

Multilayer Perceptrons, or MLPs for short, are the classical type of neural network. They are comprised of one or more layers of neurons. Data is fed to the input layer, there may be one or more hidden layers providing levels of abstraction, and predictions are made on the output layer, also called the visible layer. In MLP model one we have used a hidden layer size is (9,2) and got an accuracy of around 47%. In MLP model two tuning the hidden layer size is (10,2) and in MLP Model three tuning the hidden layer size is (22,3) the accuracy was around 59% and 64% respectively.

## MLP Model 1

```
mlp = MLPClassifier(hidden_layer_sizes = (9, 2), max_iter = 270)
mlp.fit(x_train, y_train)
pred_train = mlp.predict(x_train)
pred_test = mlp.predict(x_test)
print("Train Data Score =", (accuracy_score(y_train, pred_train)))
print("Test Data Score =", (accuracy_score(y_test, pred_test)))
print("Accuracy :", (accuracy_score(y_test, pred_test)) * 100, "%")
print('Confusion Matrix:')
print(confusion_matrix(y_test, pred_test))
print(classification_report(y_test, pred_test))
```

```
Train Data Score = 0.7586206896551724
Test Data Score = 0.47863247863247865
Accuracy : 47.863247863247864 %
Confusion Matrix:
[[37 22]
 [39 19]]
              precision    recall  f1-score   support

           0       0.49      0.63      0.55        59
           1       0.46      0.33      0.38        58

    accuracy                           0.48       117
   macro avg       0.48      0.48      0.47       117
weighted avg       0.48      0.48      0.47       117
```

**Fig 4.14 Accuracy and Performance Measures for MLP Model 1**

## MLP Model 2

```
mlp = MLPClassifier(hidden_layer_sizes = (10, 2), max_iter = 300)
mlp.fit(x_train, y_train)
pred_train = mlp.predict(x_train)
pred_test = mlp.predict(x_test)
print("Train Data Score =", (accuracy_score(y_train, pred_train)))
print("Test Data Score =", (accuracy_score(y_test, pred_test)))
print("Accuracy :", (accuracy_score(y_test, pred_test)) * 100, "%")
print('Confusion Matrix:')
print(confusion_matrix(y_test, pred_test))
print(classification_report(y_test, pred_test))
```

```
Train Data Score = 0.7931034482758621
Test Data Score = 0.5982905982905983
Accuracy : 59.82905982905983 %
Confusion Matrix:
[[29 30]
 [17 41]]
              precision    recall  f1-score   support

           0       0.63      0.49      0.55        59
           1       0.58      0.71      0.64        58

    accuracy                           0.60       117
   macro avg       0.60      0.60      0.59       117
weighted avg       0.60      0.60      0.59       117
```

**Fig 4.15 Accuracy and Performance Measures for MLP Model 2**

### MLP Model 3

```
mlp = MLPClassifier(hidden_layer_sizes = (22, 3), max_iter = 800)
mlp.fit(x_train, y_train)
pred_train = mlp.predict(x_train)
pred_test = mlp.predict(x_test)
print("Train Data Score =", (accuracy_score(y_train, pred_train)))
print("Test Data Score =", (accuracy_score(y_test, pred_test)))
print("Accuracy :", (accuracy_score(y_test, pred_test)) * 100, "%")
print('Confusion Matrix:')
print(confusion_matrix(y_test, pred_test))
print(classification_report(y_test, pred_test))
```

```
Train Data Score = 0.7241379310344828
Test Data Score = 0.6410256410256411
Accuracy : 64.1025641025641 %
Confusion Matrix:
[[34 25]
 [17 41]]
              precision    recall  f1-score   support

           0       0.67      0.58      0.62        59
           1       0.62      0.71      0.66        58

    accuracy                           0.64       117
   macro avg       0.64      0.64      0.64       117
weighted avg       0.64      0.64      0.64       117
```

**Fig 4.16 Accuracy and Performance Measures for MLP Model 3**

### 4.5.2 CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional Neural Networks, or CNNs, were designed to map image data to an output variable. They have proven so effective that they are the go-to method for any type of prediction problem involving image data as an input. In our model we have implemented one input layer and one output layer and 3 hidden layers which used to classify both image and phenotypic data. In the output layer we have used the svc as activation function and adam as an optimizer and hinge as a loss function which acts as a support vector classifier. The model gives 97% accuracy for the 80% of the dataset which is higher than the all classifier and other models.

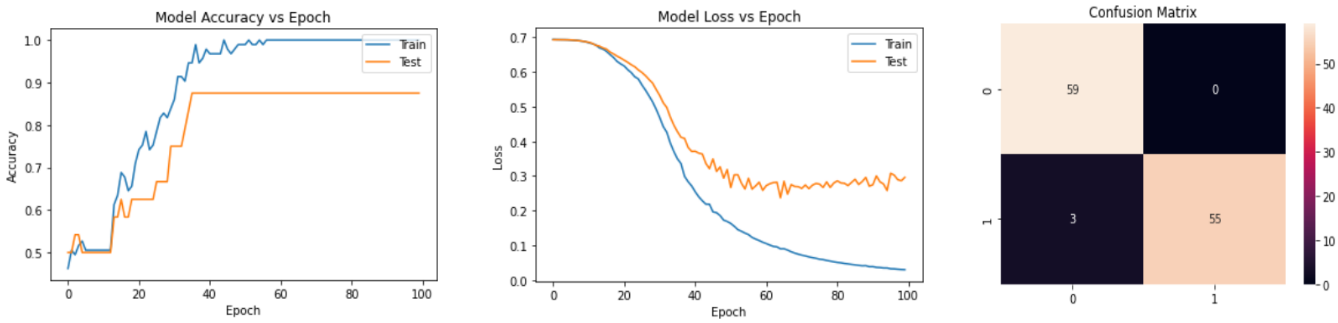



**Fig 4.17 Accuracy and Loss for Train and Test using CNN model with Confusion Matrix**

Fig 4.17 shows the CNN model accuracy and loss graph in which no of epochs are in the x-axis and accuracy in y axis. The plot shows that increasing the epochs also increases the test accuracy and at a certain point it becomes stable and also decreases the loss. In the Confusion matrix 59 controls are correctly classified. For patients 55 subjects are correctly classified and only 3 wrong.

## 4.6 FINAL RESULT USING CNN WITH SVC

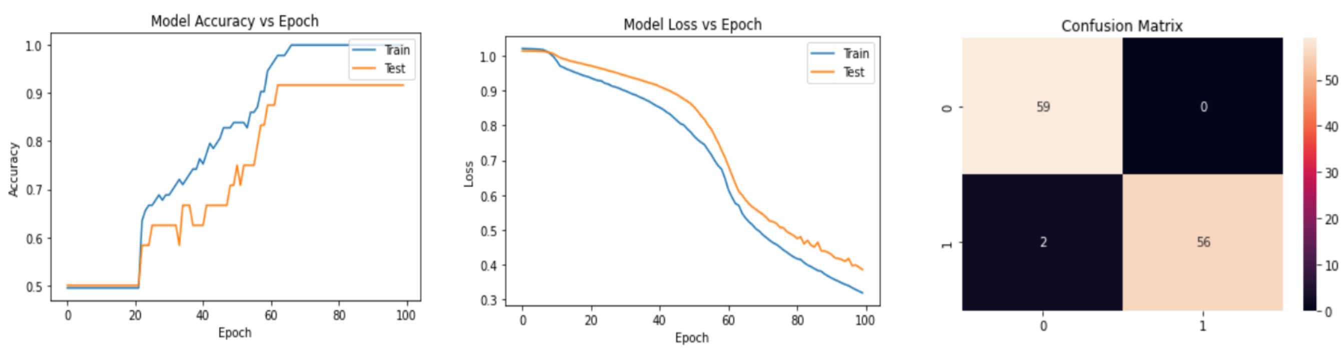



**Fig 4.18 Accuracy and Loss for Train and Test using CNN+SVC model with Confusion Matrix**

Fig 4.18 shows the CNN+SVC model accuracy and loss graph in which no. of epochs are in the x-axis and accuracy in the y-axis. The plot shows that increasing the epochs also increases the test accuracy and at a certain point it becomes stable and also decreases the loss. In the Confusion matrix, 59 controls are correctly classified. For patients, 56 subjects are correctly classified and only 2 wrong. So it ensures 98% accuracy.

# CHAPTER 5

# CONCLUSION

Schizophrenia is a chronic and severe mental disease, which largely influences the daily life and work of patients. The diagnosis is important because schizophrenia's severe conditions lead to death. In this work, we proposed a hybrid approach to detect schizophrenia based on FMRI images and phenotypic data. FMRI images are used to state the functionality of the brain which plays a key role in detecting schizophrenia. visualization of phenotypic data gives an understanding of the importance of the features in detecting schizophrenia. Various Deep Learning Models and Supervised Machine Learning algorithms are trained and then tested on a dataset which contains both FMRI images and phenotypic data. This project finds that a combination of deep learning models and Supervised machine learning algorithms are more effective in terms of detecting schizophrenia.

# CHAPTER 6

# FUTURE WORKS

- For future activity, we plan to use other datasets, characters, or test and verification methods. Therefore, it is expected that the results may assist the medical field in predicting whether or not a person has schizophrenia, so they can get proper treatment for their health.

- Future work with a large sample size and multi-modal data can be performed to improve model performance and accuracy.

- Data sets can be added to produce better performance in differentiating schizophrenia.

- Larger data sets also offer the opportunity to include additional classes to differentiate the next stage of depression. As well as the Prodromal stage, the critical stage and the remaining stage will expand the scope of use.

# BIBLIOGRAPHY

[1] R. de Filippis, E. A. Carbone, R. Gaetano, A. Bruni, V. Pugliese, et al., “Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: A systematic review,” Neuropsychiatric Disease and Treatment, vol. 15, pp. 1605–1627, 2019.

[2] Anant V Nimkar & Divesh R Kubal,” Optimization of Schizophrenia Diagnosis Prediction using Machine Learning Techniques” DOI: 10.1109/ICCOINS.2018.8510599.

[3] T V Rampisela & Z Rustam. “Classification of Schizophrenia Data Using Support Vector Machine (SVM), ”Journal of Physics Conference Series 1108(1):012044 DOI:10.1088/1742-6596/1108/1/012044, 2018.

[4] Kanghan Oh, Woo Sung Kim, Guang fan Shen, Yanhong Piao, Nam-In Kang, Il-Seok Oh, Young Chul Chung, “Classification of schizophrenia and normal controls using 3D convolutional neural network and outcome visualization”, DOI:10.1016/j.schres.2019.07.034.

[5] Mengjiao Hu, Kang Sim, Juan Helen Zhou, Xudong Jiang, Cuntai Guan, “Brain MRI-based 3D Convolutional Neural Networks for Classification of Schizophrenia and Controls”, DOI: 10.1109/EMBC44109.2020.9176610

[6] Jihoon Oh, Baek-Lok Oh, Kyong-Uk Lee, Jeong-Ho Chae and Kyongsik Yun, “Identifying Schizophrenia Using Structural MRI With a Deep Learning Algorithm”, Front. Psychiatry, 03 February 2020, DOI: 10.3389/fpsyt.2020.00016