Identifying EEG-based Correlations for Prediction of Psychiatric Disorders

Kashika Rathore, Sharai Barrera, Varsha Sampath, Sarita Raghunath, Jahnavi Handral

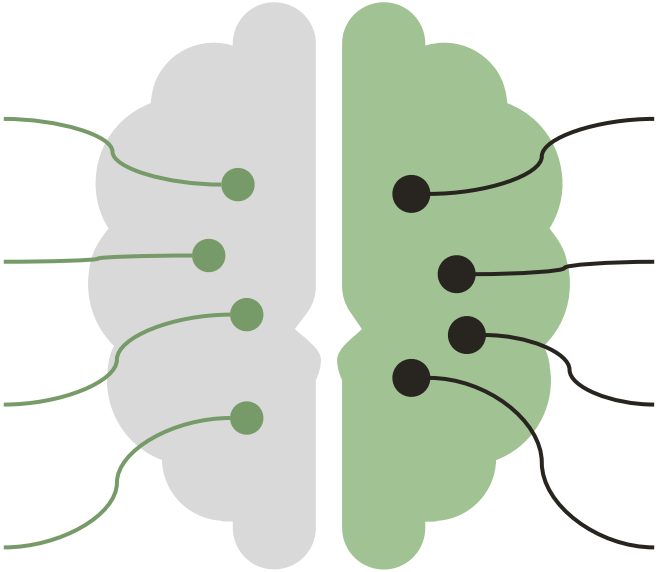
Table Of Contents

01

INTRODUCTION

WHAT IS EEG?

It most commonly studies the following waves: delta, theta, alpha, beta, gamma



It’s a test that measures electrical activity in the brain using small, metal discs (electrodes)

Research Question

Using pre-processed EEG datasets that reflect psychiatric disorders (specifically Schizophrenia & ADHD), how can we detect correlations/distinctions (biomarkers, electrodes, etc.) which signify the presence of a disorder, to aid in the development of diagnostic models and its prediction likelihood?

DATASET # 1

ADHD // Number of observations: 79

The study is separated into 4 different data sets: healthy males, healthy females, ADHD males, ADHD females. Each group was asked to complete a task while undergoing EEG specifically recording from the channels O1, F3, F4, Cz, and Fz. We will combine the female and male ADHD patients into a single file and omit parts of the task where they are not in a resting state.

DATASET # 2

SCHIZOPHRENIA // Number of observations: 28

This study is EEG data from 14 schizophrenic subjects and 14 healthy subjects measured from 19 channels: Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2. We are only using the schizophrenic patients data, transforming the edf file to a pandas dataframe and extracting certain channels.

Hypothesis

Utilizing resting state EEG recordings, we believe distinct EEG markers, such as increased delta power/ reduced alpha power and modified connectivity within attention-related networks, will result in higher accuracy for distinguishing ADHD from those with schizophrenia. We believe that the patterns of attentional deficits reflected in ADHD, resulting from impaired coordination amongst brain regions, are specific enough to develop a more precise model.

02

DATA WRANGLING AND CLEANING

```
'''Function that will take specified cell's data and loop through subjects to add to respective dataframe '''
def subj_loop(subj_list, cell_num,index_list,eyes,df_list,start_num):

    for i in range(len(subj_list[0])):

        subj = subj_list[cell_num-1][i]#
        subj_df = pd.DataFrame(subj.T,index = index_list)
        subj_df["subj"] = ["subj_"+ str(start_num+i+1)]  * 2
        subj_df["cell"] = [cell_num] * 2 #
        subj_df["eyes"] = [eyes] * 2#


        df_list[start_num + i] = pd.concat([df_list[start_num + i],subj_df])

'''Function that takes in a list of edf files,list of indexes, and df_name to convert edf files
into a dataframe,returns list of dataframes each corresponding to an edf subject file'''

def edf_to_df(file_list, index_list, df_name):

    #empty list to save dataframes
    df_list = []

    for i in range(len(file_list)): #loop through each file in file_list

    #start of code used from stack exchange


        file_name = pyedflib.data.get_generator_filename()
        f = pyedflib.EdfReader(sch_files[i])
        n = f.signals_in_file
        signal_labels = f.getSignalLabels()
        sigbufs = np.zeros((n, f.getNSamples()[0]))
        for j in np.arange(n):
            sigbufs[j, :] = f.readSignal(j)

        f.close()

    # end of code used from stack exchange
```

Schizophrenia

Data Wrangling code

```
'''Loop through Schizophrenia patient data files and create into dataframe using edf_to_df function '''

#14 schizophrenia data files
sch_files = ["s01.edf","s02.edf","s03.edf","s04.edf","s05.edf","s06.edf","s07.edf","s08.edf","s09.edf","s10.edf","s11.edf","s12.edf","s13.edf","s14.edf"]

# 19 EEG channels
channels = ["Fp1", "Fp2", "F7", "F3", "Fz", "F4", "F8", "T3", "C3", "Cz", "C4", "T4", "T5", "P3", "Pz", "P4", "T6", "O1", "O2"]

#apply functions
sch_dfs = edf_to_df(sch_files, channels,"Schiz_")
sch_bigdf = merge_df_list(sch_dfs)

sch_bigdf
```

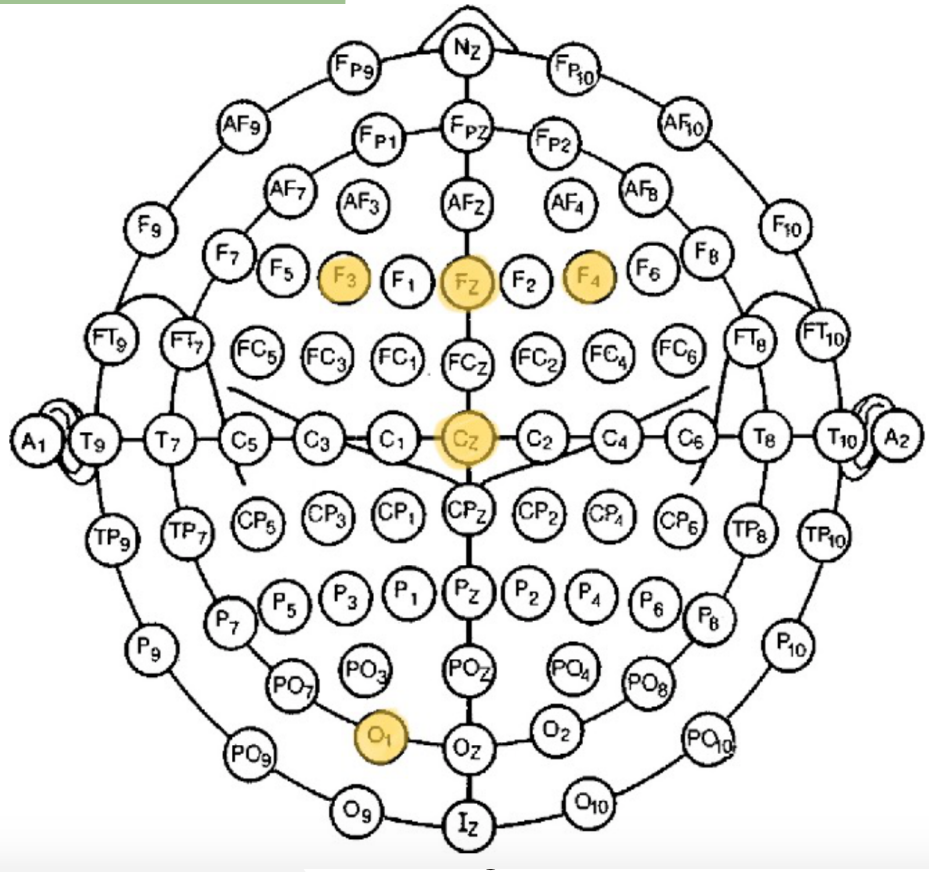
ADHD Data Cleaning code

```
# remove data corresponding to subject 34 (faulty)
mask = adhd_bigdf['subj'] != 'subj_34'
adhd_bigdf = adhd_bigdf[mask]

#remove columns with NaN values
adhd_bigdf= adhd_bigdf.dropna(axis=1)

adhd_bigdf
```

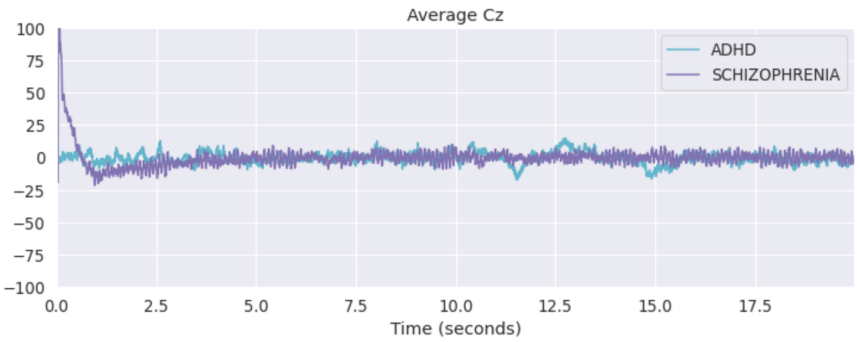
Schizophrenia Data

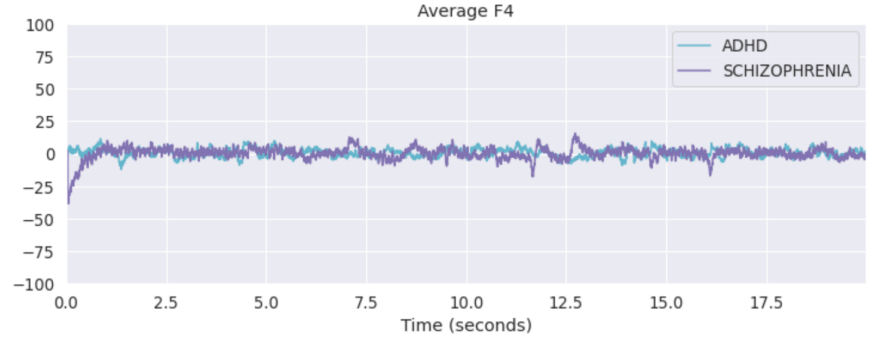
```
# remove the outliers from the dataset
from scipy import stats
z_scores = np.abs(stats.zscore(subset))
outlier_threshold = 3
outlier_mask = z_scores > outlier_threshold
no_outliers = subset[~outlier_mask.any(axis=1)]
subset
```

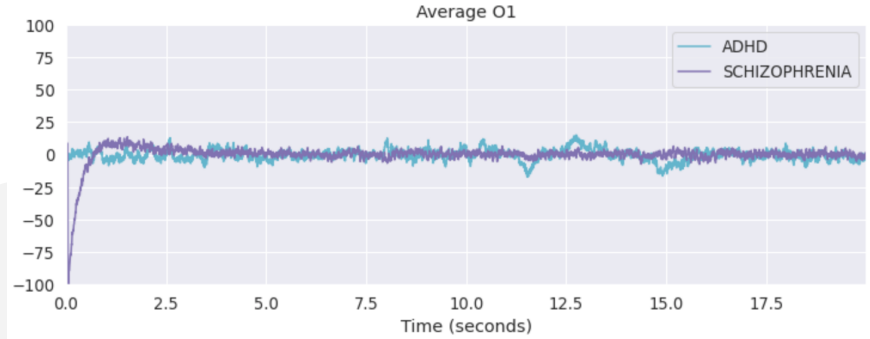
```
#standardize and normalize data
from sklearn.preprocessing import StandardScaler, MinMaxScaler
scale = StandardScaler()
standardized = pd.DataFrame(scale.fit_transform(subset), columns=subset.columns)
normal = MinMaxScaler()
df_normalized = pd.DataFrame(normal.fit_transform(subset), columns=subset.columns)
subset
```

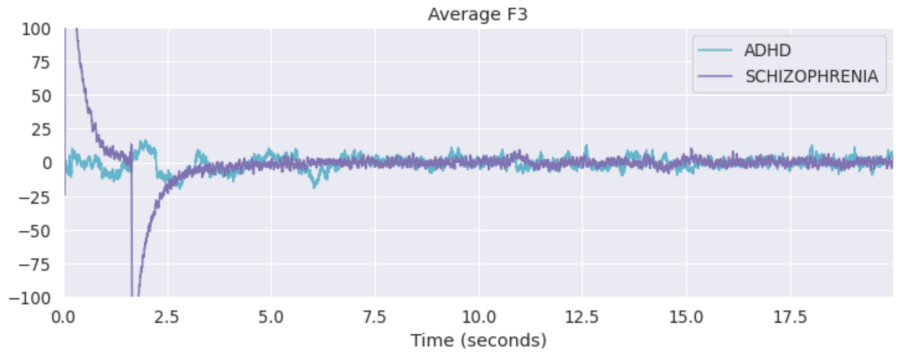


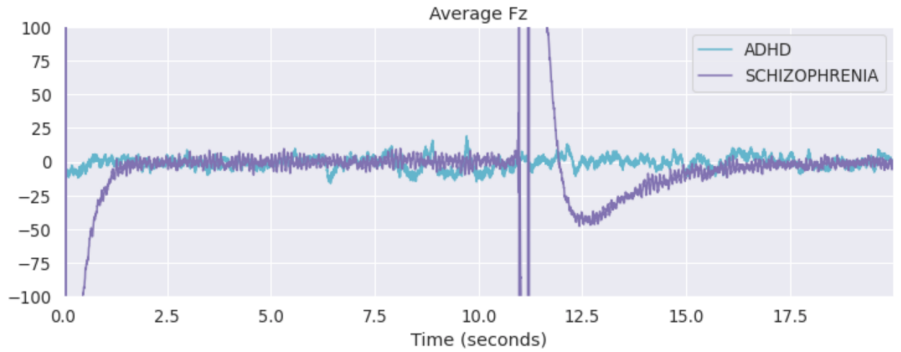
Map of Electrodes we are analyzing



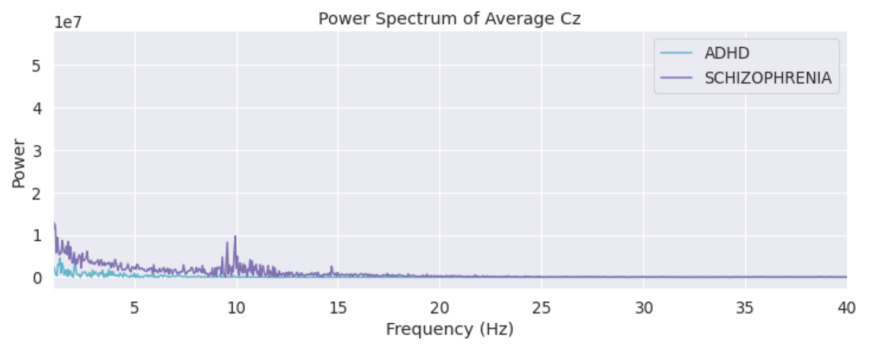


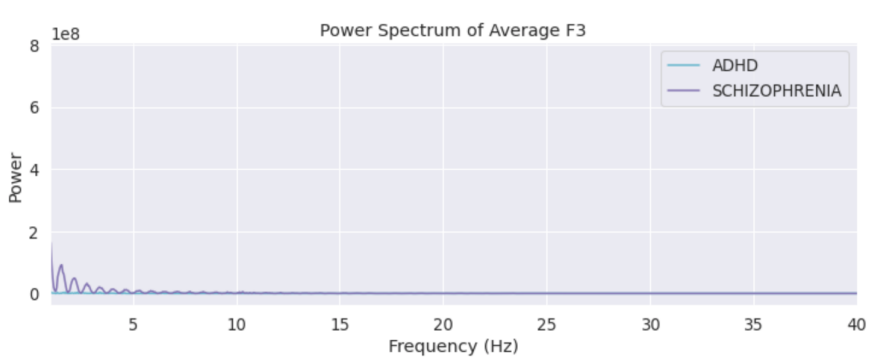


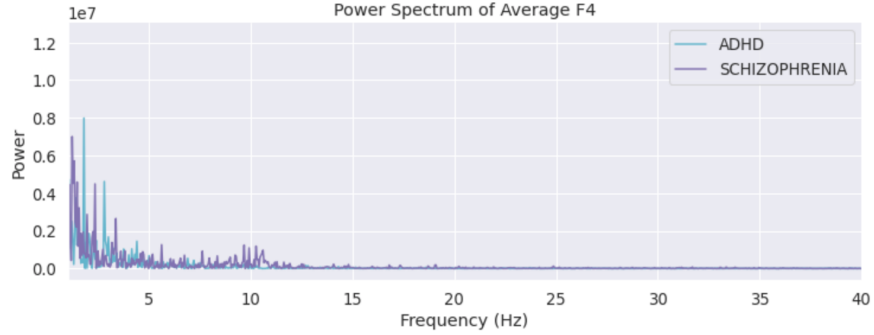


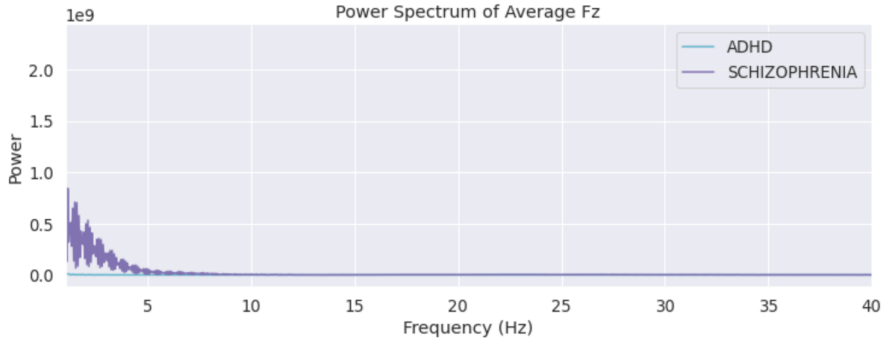


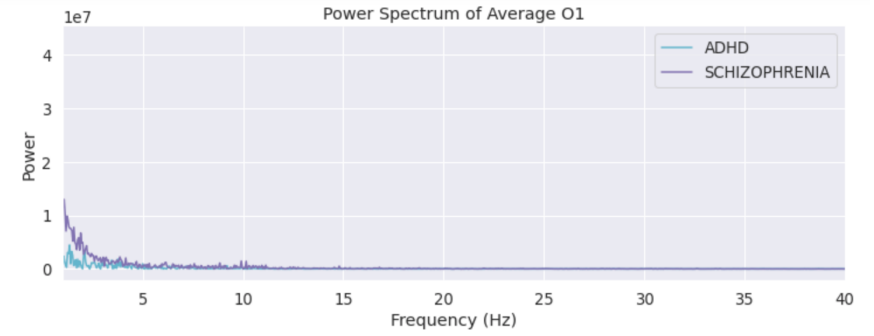
average RAW eeg signal per electrode



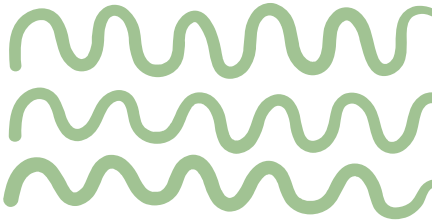


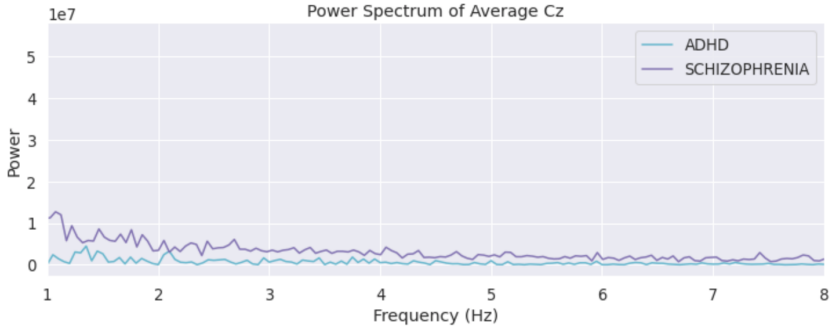


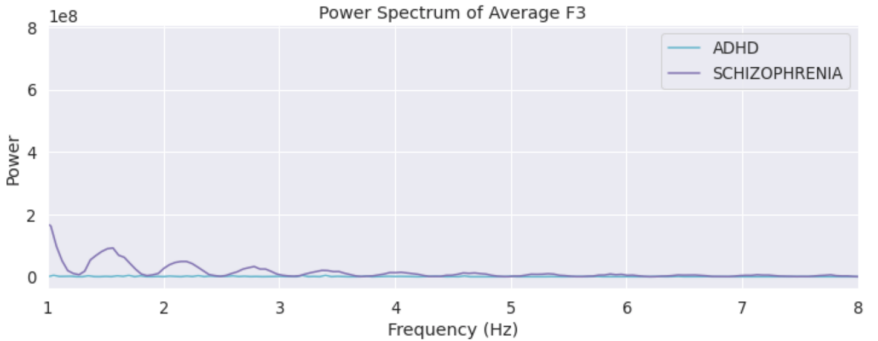


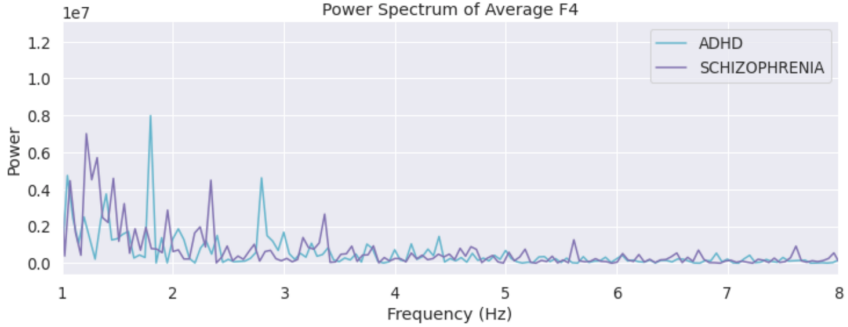


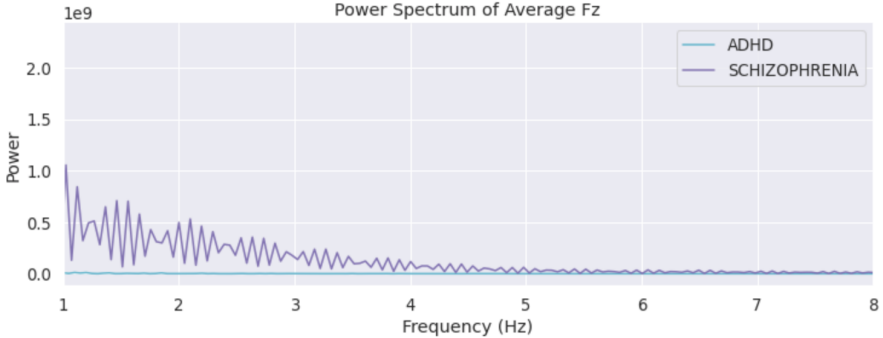
Power Spectrum ANalysis

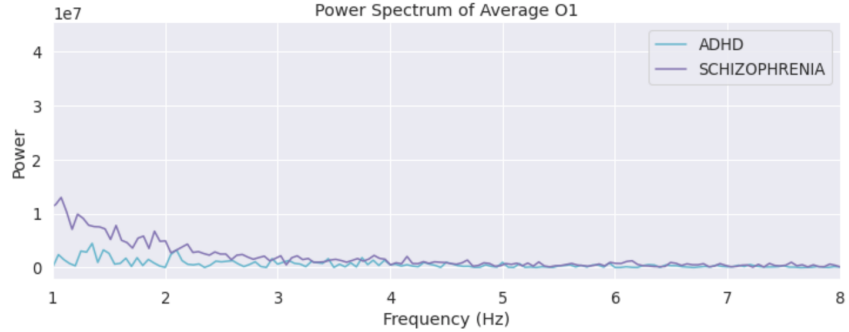




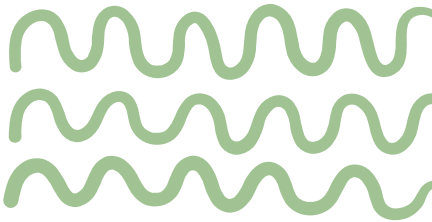








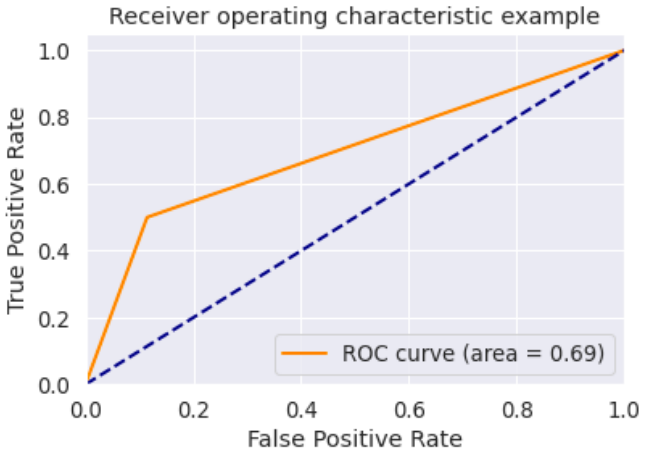
Looking closer at slow frequencies

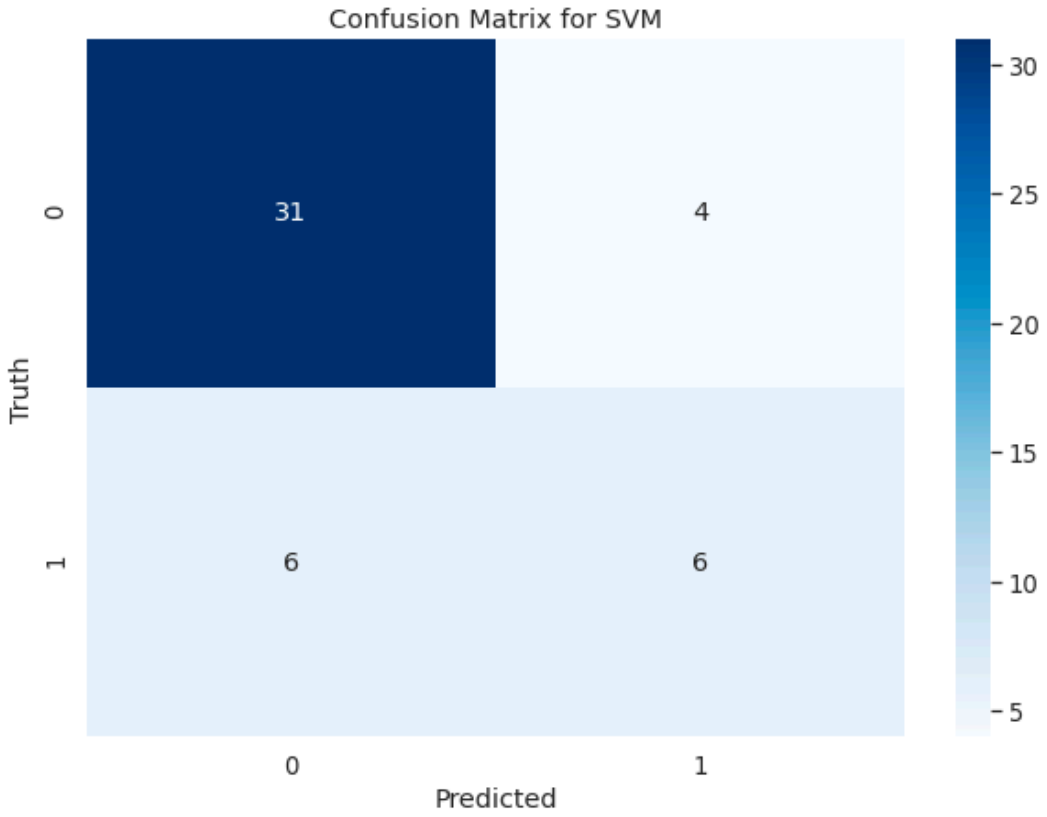


03

MODELS AND RESULTS

Support VECTOR MACHINE(SVM)





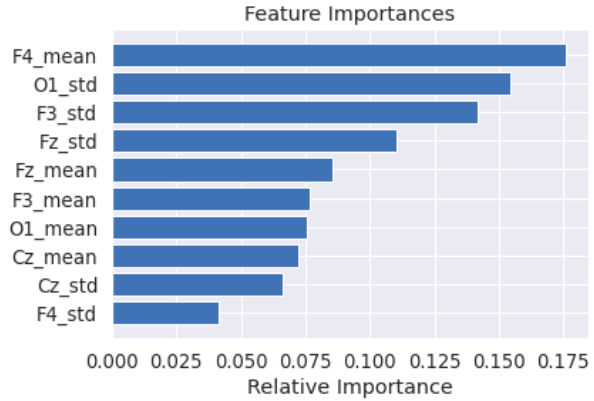
Accuracy: 0.7872340425531915

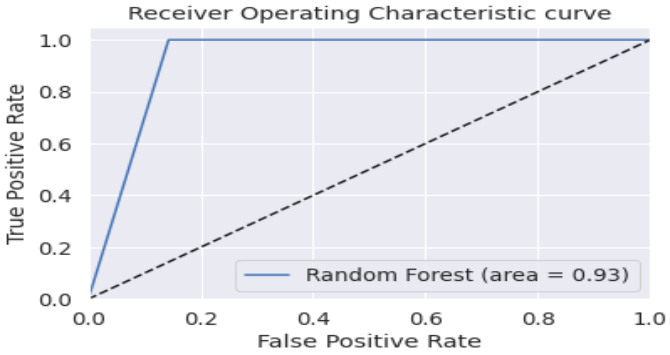
	precision	recall	f1-score	support
ADHD	0.84	0.89	0.86	35
Schizophrenia	0.60	0.50	0.55	12
accuracy			0.79	47
macro avg	0.72	0.69	0.70	47
weighted avg	0.78	0.79	0.78	47

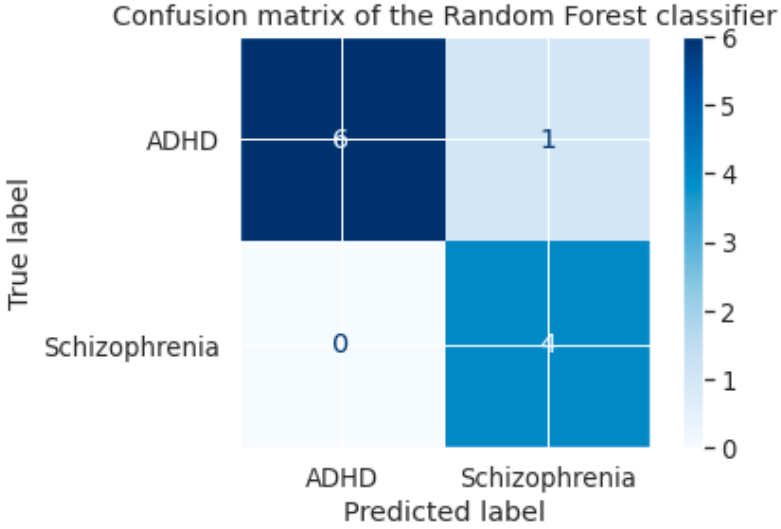
Random forests(RF)

Scores: [1. 1. 0.8 0.8 0.7]
Mean: 0.86
Accuracy: 0.9090909090909091

	precision	recall	f1-score	support
ADHD	1.00	0.86	0.92	7
Schizophrenia	0.80	1.00	0.89	4
accuracy			0.91	11
macro avg	0.90	0.93	0.91	11
weighted avg	0.93	0.91	0.91	11







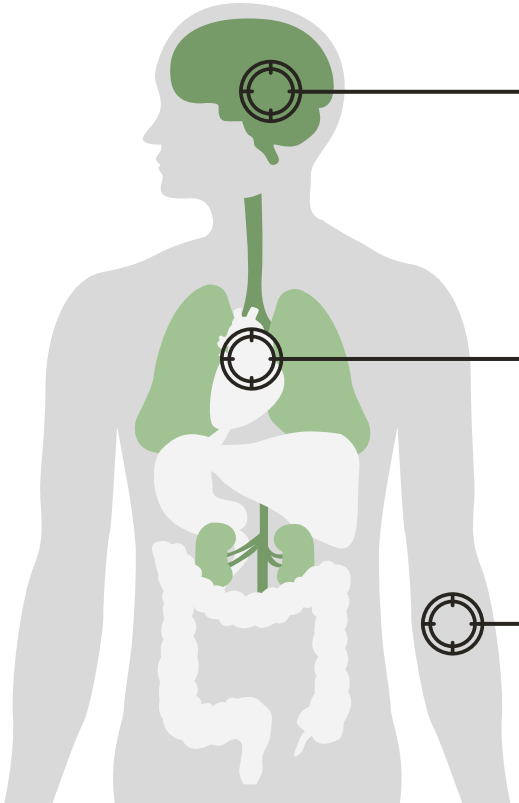
Model Comparison

RF was more suitable overall.

SVM was weaker in Schizophrenia cases.

RF had better accuracy and AUC_ROC score

pHYSICAL EXAMINATION



BRAIN

Mercury is the smallest planet

HEART

Saturn is a gas giant with several rings

MUSCLES

Jupiter is the biggest planet of them all

Discussion & Conclusion

Our research demonstrates that it is possible to use EEG data to distinguish between different psychiatric disorders, especially schizophrenia, although it should not be the only diagnostic tool. Further research and data collection is needed to make even more accurate diagnoses.