**Diagnosis of Parkison’s Disease and Schizophrenia using Machine Learning Techniques on EEG Data.**

Mustafa Mehmood,

Student of Computer Science

at University of Nottingham

Under Supervision of

Mr Haffez Ullah Amin

Asst. Professor

School of Computer Science

**Introduction**

According to WHO, 1 in every 8 people in the world live with a mental disorder. Mental disorders are characterised by clinically significant disturbances in an individual’s cognition, emotional regulation, or behaviour. It is usually associated with distress or impairment in important areas of functioning. Effective prevention and treatment options exist. But most people do not have access to effective care, therefore it is necessary to develop cheap and easy to access diagnosis and care.

There are many different types of mental disorders, Schizophrenia and Parkinson’s Disease are one of the most important mental disorders and the focus of this project.

An individual with schizophrenia interprets reality abnormally. They may experience hallucinations, delusions, and extremely disordered thinking and behaviour, which can have a devastating effect on their daily functioning and may require lifelong treatment. It disrupts brain development [1].Worldwide, there are nearly 21 million individuals suffering from this type of brain disorder, according to the World Health Organization.

Parkinson's disease is a brain disorder that causes uncontrolled movements, such as trembling, stiffness, and difficulty coordinating and balancing. There is usually a gradual onset of symptoms that worsen over time. Walking and speaking may become difficult as the disease progresses. It is unclear what causes Parkinson's disease. Both environmental and inherited factors may contribute [2] . The substantia nigra is damaged in this neurodegenerative condition. An important part of these neurons is the production of dopamine. The chemical dopamine functions as a messenger in the brain between neurons. For body movements and speech to function properly, it assists the brain in sending signals to various parts of the body. If there is a high loss of dopaminergic neurons in the brain or a low level of dopamine in the brain, PD symptoms will appear [3]. According to WHO, global estimates in 2019 showed over 8.5 million individuals with PD.

The final diagnosis will always be determined by the neurologist's opinion, however, any tool that can help them contrast their diagnosis is always welcome. Therefore, automated procedures that can improve diagnosis accuracy and speed are in demand

In order to automate the diagnosis of these disorders, several methods have been proposed.

There have been several approaches presented in this area, most of which use voice signals, gait signals, handwriting signals, and MRI [4][5]. Neuroimaging-based techniques are among the most promising of these techniques. Among the most practical and affordable functional neuroimaging techniques is EEG (electroencephalogram). Brain electrical activities are recorded from the head surface at high temporal and spatial resolutions. Brain neurons produce voltage fluctuations through ionic currents that are measured by EEG (Niedermeyer E.; da Silva F.L. (2004)). With millisecond-range temporal resolution, it offers a unique advantage over CT, PET, and MRI.

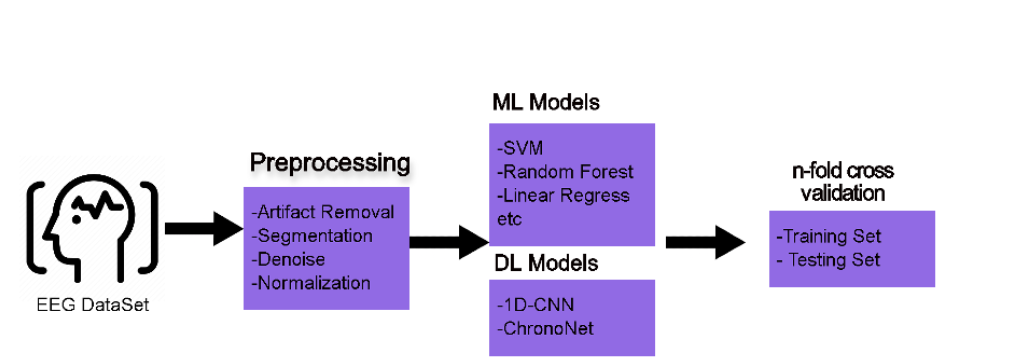
It has a reasonable cost and portability to it. The EEG also records brain activity faster than other technologies and for a longer period of time. In conjunction with machine learning techniques, EEG analysis has already been used to detect autism spectrum disorder, schizophrenia, Alzheimer's disease, epilepsy, and major depression. However EEGs have not been fully investigated as a method of studying PD.

According to Ly et al, six patients with Parkinson's disease were classified based on their EEG data during Timed Up and Go Task studies[6]. Stockwell transforms were used to extract features. In order to separate the EEG sources, independent component analysis was used with entropy bound minimization. The frequency-based features of selected independent components of EEG were extracted and classified using Bayesian Neural Networks, resulting in an 86.2% accuracy in TF detection. A different study discussed various methods of identifying gait initiation failures (GIFs), which are a type of freezing of gait[7]. An SVM was used to classify GIF events in five PD patients with FOG using wavelet transforms. These approaches were accurate in identifying GIF episodes, with an accuracy rate of 86.3%. Data collected from resting EEGs of patients with Rapid Eye Movement Behavior Disorder was used for developing deep learning models to identify GIF episodes. 25. Eighty-one percent and seventy-nine percent classification accuracy was obtained using RNN-LSTMs and five-layer CNNs, respectively. There are two main types of studies discussed so far: those that aim to detect Parkinson's disease by asking participants to perform mental or muscular tasks, or those that aim to determine the most effective EEG parameters/characteristics for the detection of Parkinson's disease.

The following studies are dedicated to the detection of PD patients from HC using resting-state EEG recordings.[8] To develop an automated diagnosis of Parkinson's disease, Yuvaraj et al used a higher-order spectrum (HOS) feature extractor. An assessment of the relevance of the bispectrum features was conducted. In addition to decision trees, KNNs, fuzzy KNNs, NBs, probabilistic neural networks, and SVMs, the use of resting-state EEG signals to detect Parkinson’s disease has also been investigated .[9] According to the study, it was possible to achieve 88.25% classification accuracy using a thirteen-layer convolutional neural network (CNN). [10] To decompose EEG signals into several subbands, Smith K.K. et al. proposed the wavelet transform. Statistical measurements were used to extract five features from these subbands, which were then categorised using several machine learning techniques.

This project uses various conventional Machine Learning methods, e.g., SVM, K-NN, Random forest, Logistic Regression, Naive Bayes and some deep learning (DL)-based methods such as 1D-CNN and a deep learning model implementation known as ChronoNet (Roy, Kiral-Kornek and Harrer, 2019) for automated diagnosis of Parkinson's and Schizophrenia via electroencephalography (EEG) signals. The k-fold cross-validation method with k = 5 has been used. The algorithm includes preprocessing of the raw dataset extracted from EEG, features extraction and selection, and finally classification. For feature selection, two methods namely statistical features and Average Bandpower using multi taper method. Data was preprocessed by performing ICA analysis to remove eyeblink artefacts and signal noise. Band pass filter and average referencing was used.

A summary of proposed procedure is portrayed in Figure 1



*Figure 1 Methodology*

**Materials and Methods**

Using two public EEG datasets, the proposed methods are tested in this study.

For Parkinson’s Disease, a previously published public data set from another laboratory was taken. 1. The dataset from this study includes raw EEG data from 15 patients with Parkinson's disease (eight females with a mean age of 63.2 years old) taking and not taking dopaminergic medication, as well as 16 healthy participants (nine females, mean age 63.5 years old). All PD patients had been diagnosed by a movement disorder specialist at Scripps clinic in La Jolla, California. Participants were right-handed and provided written consent in accordance with the Institutional Review Board of the University of California, San Diego and the Declaration of Helsinki. For this study only that EEG data was considered which was recorded when the patients were off medication. A counterbalanced order of data collection was used for patients who were taking medication and those who were not. EEG data were acquired using a 32-channel BioSemi ActiveTwo system, sampled at 512 Hz. Resting data were recorded for at least 3 min. During data collection, the participants were seated comfortably and told to fixate on a cross presented on a screen. Additional electrodes were placed lateral to and below the left eye to monitor eye blinks and movements. Other assessments were also conducted, as described in the previous report. 1. which were not analysed here.

For Schizophrenia the dataset of the Institute of Psychiatry and Neurology in Warsaw, Poland, was used (Olejarczyk and Jernajczyk, 2017) 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 this institution, and the data recording was carried out. A signal recording was performed with the eyes closed in 15 min for each case. Recording EEG signals was performed by using standard 10–20 with a sampling frequency of 250 Hz. The used electrodes include Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2.

Preprocessing

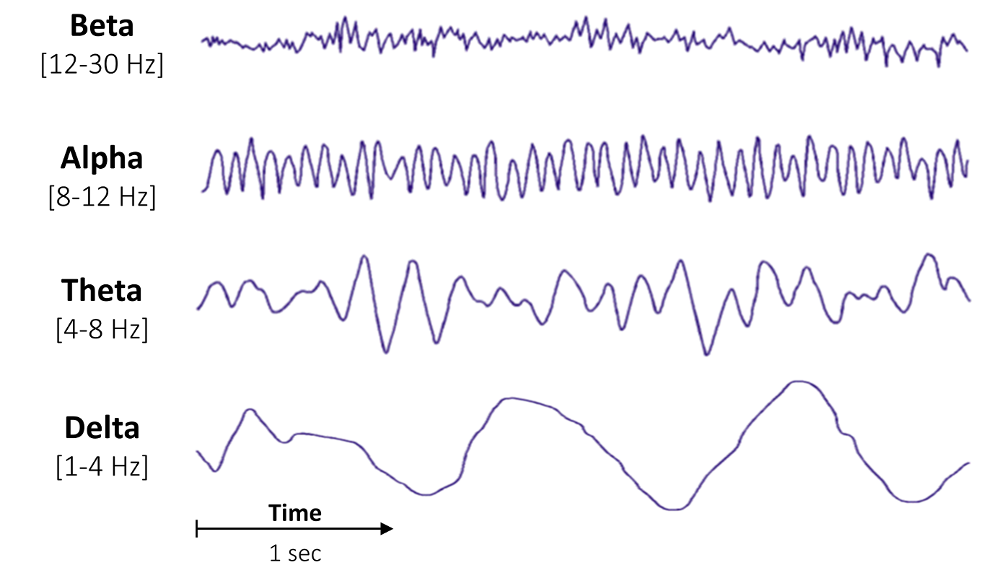
Data was preprocessed in MATLAB application. To summarise, we removed the mean of each electrode and then used an average of all electrodes as a reference. An electrode that was excessively noisy was excluded. High pass filtering at 0.5 Hz was applied to eliminate low frequency drift (using two-way FIR filters). Following independent component analysis, data was examined for artefacts such as eye blinks, muscle activity, electrical noise, and other sources of noise were removed.

Data Analysis and Feature Extraction

Feature Extraction aims to reduce the number of features in a dataset by creating new features from the existing ones (and then discarding the original features). These new reduced sets of features should then be able to summarise most of the information contained in the original set of features. For this two methods were implemented, statistical features and features extracted using Average power of EEG bands.

For statistical features . First, each recorded EEG signal( 19 for SZ and 31 for PD) has been divided into overlap-free 5 s frames. After that, mean, standard deviation, variance, minimum, maximum, max/min indices, mean square , root mean square, absolute difference in signal, skewness and kurtosis were calculated for data processing and training.

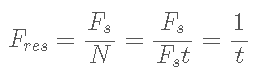
Featured were also extracted by talking the average power of signals in frequency bands such as delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–100 Hz).



*Figure 2 - Frequency bands*

Power spectral density can be used for a wide variety of analyses, but the average band power was chosen since it sums the contribution of each frequency band to the overall power of the signal.

An estimate of the power spectral density is needed first. This was done by averaging successive Fourier transforms of small windows of a signal, using Welch's periodogram. Welch's method improves the accuracy of the classic periodogram. EEG information is time-varying, it's very unlikely that the sign will appear to be a great sum of natural sines. To go back a real spectral estimate, a traditional periodogram calls for the spectral content material of the sign to be stationary (i.e. time-unvarying) over the term considered. Because it is never the case, the periodogram is generally biassed and contains way too much variance. By averaging the periodograms obtained over short segments of the windows, Welch's method allows to drastically reduce this variance. This comes at the cost, however, of a lower frequency resolution. The frequency resolution is defined by:



where Fs is the sampling frequency of the signal, N the total number of samples and t the duration, in seconds, of the signal. In our case we choose a window of 2 seconds.

Machine Learning Methods Used

Support Vector Machine

Support vector machines (SVMs) [11] are a set of supervised learning methods used for classification, regression and outlier detection.

SVMs construct hyperplanes in high-dimensional or infinite-dimensional spaces, which can be used for classification, regression, or other tasks. As a rule of thumb, a good separation is obtained by the hyperplane that has the largest distance to the nearest training data points (so-called functional margin), as, in general, a larger margin reduces generalisation error.

The advantages of support vector machines are:

* Successful in high-dimensional spaces
* Continues to be effective when there are more dimensions than samples. It is also memory efficient because it only uses a portion of the training points (known as support vectors) in the decision function.
* Multiple Kernel functions can be supplied for the decision function, making it versatile. There are common kernels available, but you can also define your own kernels.

The disadvantages of support vector machines include:

* When selecting Kernel functions and regularisation terms, it is essential to avoid overfitting if the number of features is significantly higher than the number of samples.
* Probability estimates are not directly provided by SVMs; instead, they are computed via an expensive five-fold cross-validation method.

K-Nearest Neighbors

*K*-nearest neighbour (KNN) [12] is a classification algorithm where some fixed and small number (*k*) of nearest neighbours (based on a notion of distance) from the training set are located and used together to determine the class of the test instance through a simple majority voting. The class of the test instance is assigned the data class which has the most representatives within the KNN of that point.

Neighbours-based categorization falls under the category of instance-based learning, often known as non-generalizing learning, because it doesn't try to build a broad internal model instead just maintains examples of the training data. A query point is assigned to the data class that has the most representatives among its nearest neighbours after classification is calculated using a simple majority vote of each point's nearest neighbours.

Random Forest

A random forest [13] is a meta estimator that employs averaging to increase predicted accuracy and reduce overfitting after fitting numerous decision tree classifiers to distinct dataset subsamples.

A variation on the bagging approach, random forest (RF) uses averaging to increase forecast accuracy and reduce overfitting by fitting multiple DT classifiers to different subsamples of the dataset. In contrast to bagging, RF also entails picking a subset of input characteristics (columns or variables) at each point where the trees are divided. It forces each DT in the ensemble to be more distinctive by condensing the features to a random subset that may be taken into account at each split point.

Logistic Regression

A statistical technique for forecasting binary classes is logistic regression[14]. The result or goal variable has a binary nature. There are just two conceivable classes when something is dichotomous. It can be applied, for instance, to issues with cancer detection. It determines the likelihood that an event will occur. When the target variable is categorical, linear regression is applied in a specific way. A log of the odds is used as the dependent variable. Using a logit function, logistic regression makes predictions about the likelihood that a binary event will occur.

Linear Regression Equation:



Where, y is a dependent variable and x1, x2 ... and Xn are explanatory variables.

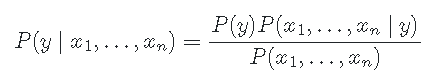
Logistic regression's characteristics include :

* The dependent variable's conformance to the Bernoulli distribution.
* A maximum likelihood approach is used for estimation.
* Model fitness is determined using Concordance and KS-Statistics; there is no R Square.

Naive Bayes

A supervised learning approach called Naive Bayes [15] applies Bayes' theorem under the "naive" assumption that given the value of the class variable, every pair of features is conditionally independent of one another. A group of supervised learning algorithms known as naive Bayes methods utilise Bayes' theorem with the "naive" assumption that each pair of features is conditionally independent given the value of the class variable.

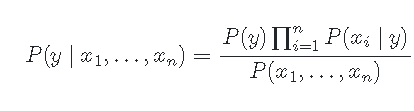
Bayes’ theorem states the following relationship, given class variable y and dependent feature vector x1 through xn :



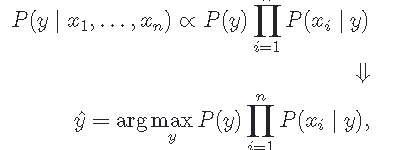
Using the naive conditional independence assumption that:



for all i , this relationship is simplified to:



Since is constant given the input, we can use the following classification rule:



and we can use Maximum A Posteriori (MAP) estimation to estimate and ; the former is then the relative frequency of class in the training set.

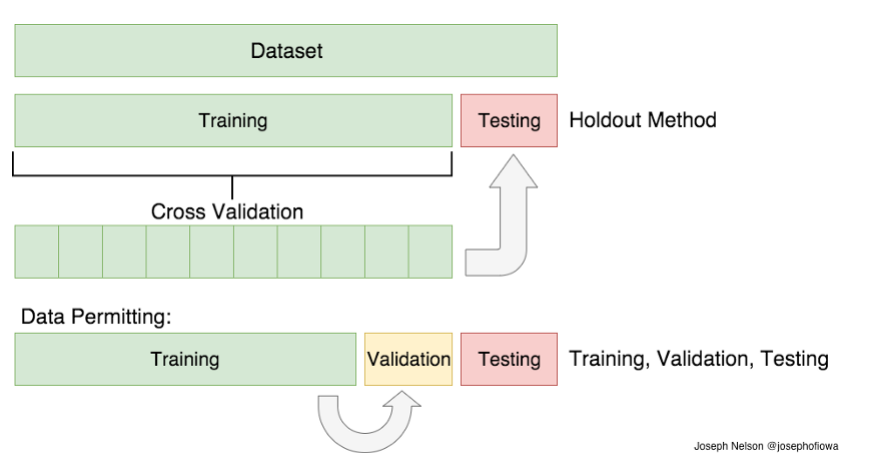
The different naive Bayes classifiers differ mainly by the assumptions they make regarding the distribution of .

In comparison to more complex techniques, naive Bayes learners and classifiers can operate at lightning speeds. Each distribution can be individually estimated as a one-dimensional distribution due to the decoupling of the class conditional feature distributions. This in turn aids in resolving issues brought on by the dimensionality curse.

Note on training

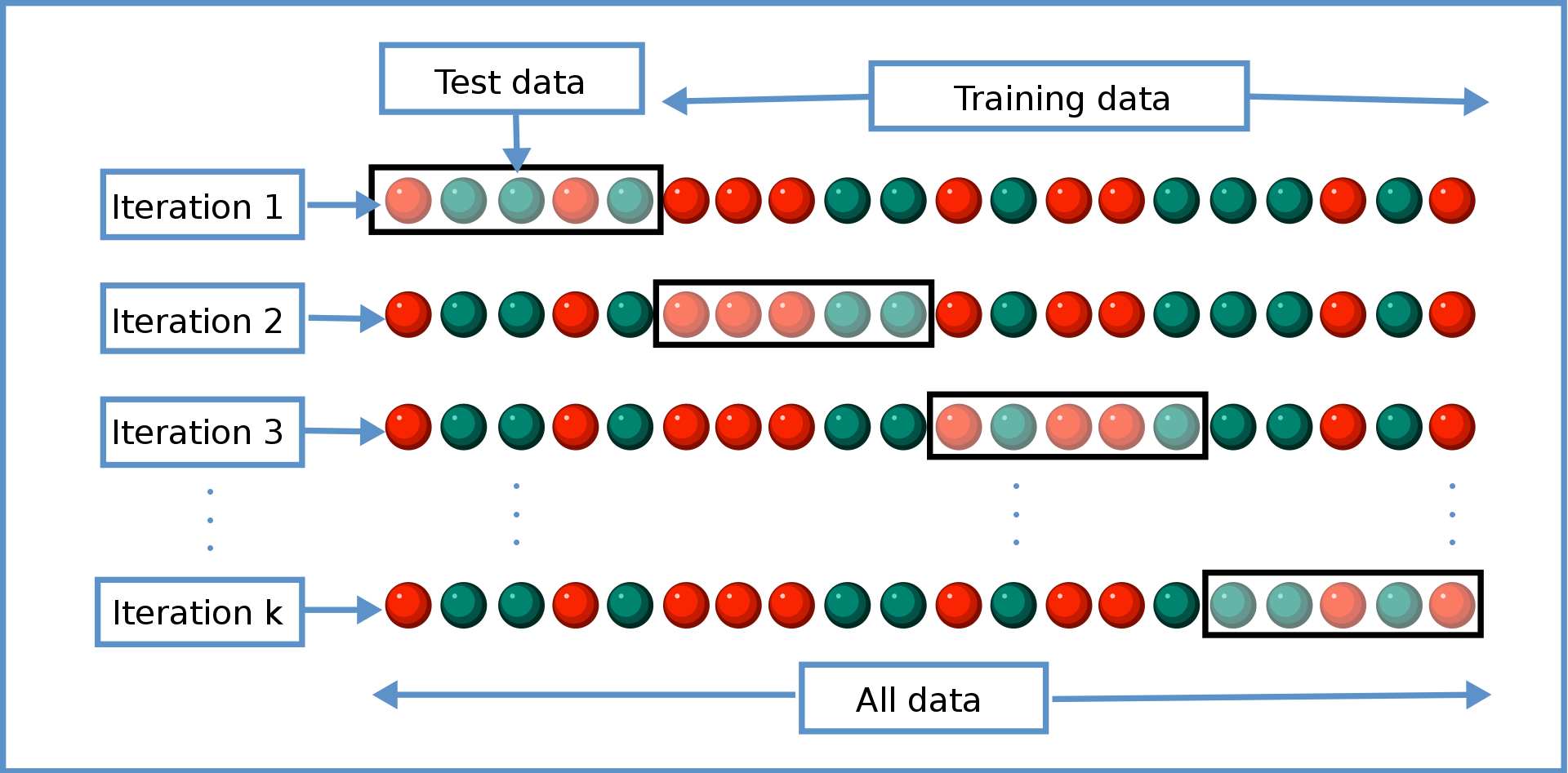
For all the above mentioned ML Algorithms, hyperparameter tuning was applied. A hyperparameter is a parameter of the model that is set before the start of the learning process. As different Machine Learning models have different hyperparameters, an exhaustive Grid Search technique for hyperparameter optimization was applied. An exhaustive grid search takes in as many hyperparameters and tries every single combination of the parameters as well as many cross validations. It is a good way to determine the best hyperparameter values to use, but it quickly becomes time consuming with every additional parameter value therefore only the most likely parameters were chosen.

In order to avoid the caveats of the train/test split method, Cross-validation was applied on the data set. It uses more subsets than train/test split yet is fairly comparable to it. In other words, the data is divided into k subsets, and one of those subsets is trained on. Tests are conducted on the final subset. It is carried out for every subset.



*Figure 3 - Cross Validation*

Among the several cross validation techniques, the K-Folds Cross Validation methodology was employed. We divided our data into k distinct subsets for K-Folds Cross Validation (or folds). Our data is divided into k-1 subsets for training, and the final subset (or fold) is used for testing. We next conclude our model by averaging the model against each of the folds. Then we compare it to the test set.



*Figure 4- K Cross validation*

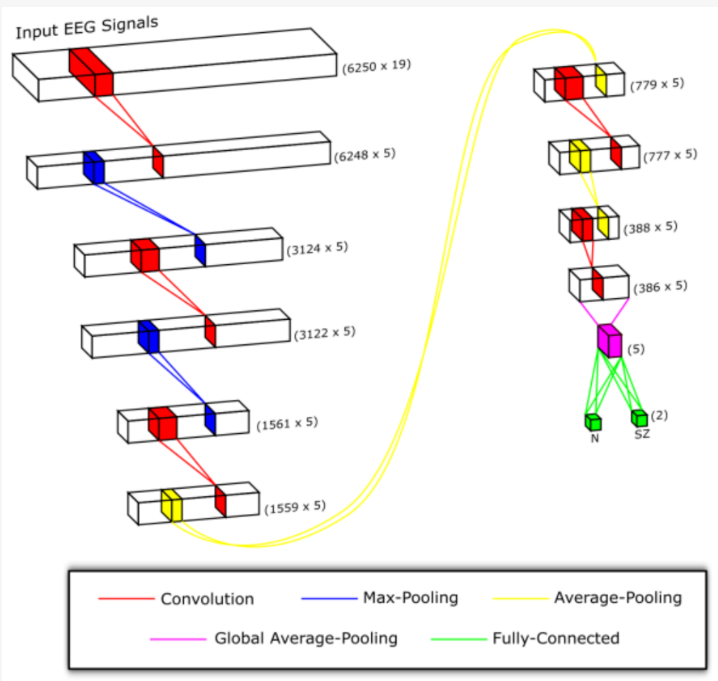
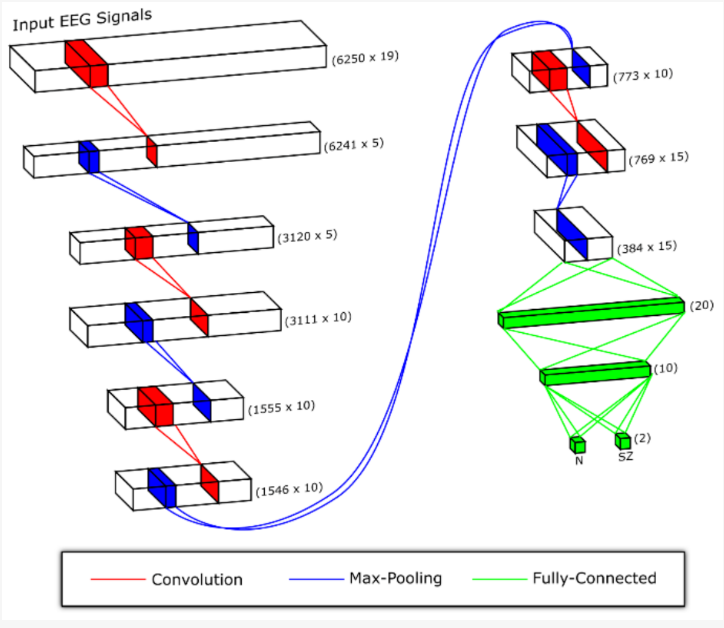
Deep Learning Techniques

Convolutional Neural Network

The CNN is a sophisticated network made up of numerous disguised layers and parameters. Convolution, max pooling, and fully connected layers make up the network's three primary tiers. The convolutional layer of the CNN goes through a training procedure in which it employs various kernel sizes to understand the input information. Features are retrieved from the input signals during convolution, and the feature maps for the following layer are then created. The batch normalisation layer is then used to exercise the middle layers in order to normalise the training data. This aids in accelerating and enhancing the learning process. The feature map gets smaller as a result of max pooling because it only returns the highest value from each kernel. The top features of the input data are depicted in the output from the convolutional and pooling layers. The input data is subsequently divided into several groups by the fully-connected layer using the training data as a basis. Each neuron in the fully connected and max pooling layers is connected, allowing the output to predict the result of the input signal as normal or not with accuracy.

A Deep Convolutional Neural Network Model for Automated Diagnosis of Schizophrenia designed by Oh, S.L.; Vicnesh, J.; Ciaccio, E.J.; Yuvaraj, R.; Acharya,[16] was implemented and tested. Features were extracted automatically at the convolution stage, with the most significant features extracted at the max-pooling stage, and the fully connected layer is utilised to classify the signals.

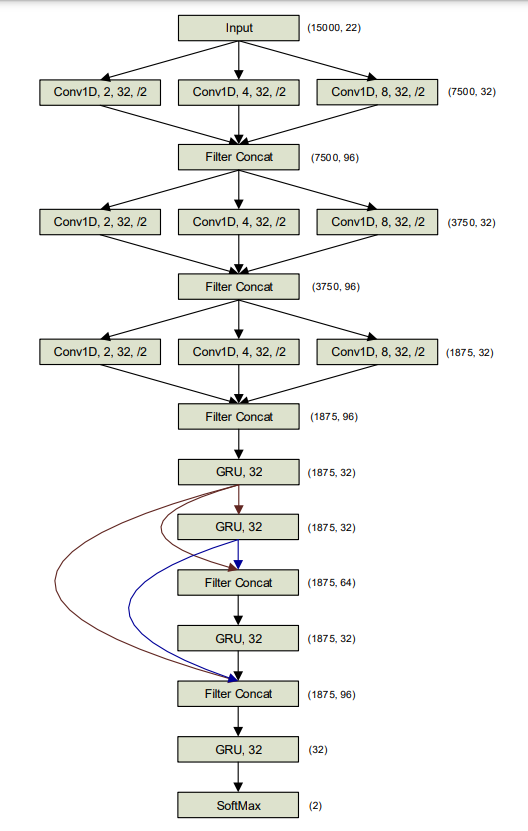
These figures highlight the architectures proposed in this study.



*Figure 5 1dCNN Architecture*

ChronoNet: A Deep Recurrent Neural Network for Abnormal EEG Identification

A novel recurrent neural network (RNN) architecture termed ChronoNet designed by Roy, Kiral and Harrer [17] was implemented and tested using the Keras Framework in python. ChronoNet is formed by stacking multiple 1D convolution layers followed by deep gated recurrent unit (GRU) layers where each 1D convolution layer uses multiple filters of exponentially varying lengths and the stacked GRU layers are densely connected in a feed-forward manner. ChronoNet directly takes time-series EEG as input and learns meaningful representations of brain activity patterns. The complete architecture of the model can be found by following the paper. The system framework of that study is highlighted in the illustration.



*Figure 6- ChronoNet Architecture*

**Results and Conclusion**

Conventional machine learning techniques were trained using features extracted by statistical methods for both the Schizophrenia and Parkinson’s Disease data sets The tested algorithms include Support Vector Machine, k-nearest neighbours, Gaussian Naive Bayes, Random Forest and Logistic Regression.Epoch length of 5 seconds was used and the cross validation folds were set to 40 and all the algorithms were hyperparameter tuned using the GridSearch technique of python. The results of the experiment can be examined in the following tables 1 and 2.

| Model | Accuracy | Precision | F1-Score | Recall |
| --- | --- | --- | --- | --- |
| SVM | 0.94 | 0.94 | 0.95 | 0.96 |
| K-NN | 0.89 | 0.89 | 0.90 | 0.91 |
| GaussianNB | 0.64 | 0.72 | 0.44 | 0.37 |
| Random Forest | 0.95 | 0.95 | 0.95 | 0.96 |
| Logistic Regression | 0.90 | 0.85 | 0.80 | 0.81 |

*Tabel 1- Parkinson’s Disease Data Set - Features: Statistical*

| Model | Accuracy | Precision | F1-Score | Recall |
| --- | --- | --- | --- | --- |
| SVM | 0.78 | 0.83 | 0.78 | 0.76 |
| K-NN | 0.62 | 0.64 | 0.67 | 0.71 |
| GaussianNB | 0.51 | 0.54 | 0.15 | 0.12 |
| Random Forest | 0.91 | 0.92 | 0.93 | 0.94 |
| Logistic Regression | 0.85 | 0.89 | 0.86 | 0.86 |

*Tabel 2-* Schizophrenia Data Set  *- Features: Statistical*

For both the data sets the results were similar but the classification algorithms performed much better over all for the detection of PD compared to Schizophrenia. Random Forest out performed the rest in both the cases. Logistic Regression performing slightly worse to come second. Random Forest had scores of 0.91, 0.92, 0.93, 0.94 and 0.95, 0.95, 0.95, 0.96 for accuracy, precision, f1- score and recall for PD and SZ data sets respectively. Gaussian Naive Bayes had the worst classification accuracy in both the cases with scores as low as 0.51, 0.54, 0.15, 0.12 and 0.64, 0.72, 0.44, 0.37.

The same algorithms were also trained on extracted features of Average Band Power of the EEG channels. But overall the performance was worse compared to the results of the previous techniques. This time also the hyperparameter tuning was done and cross validation fold technique was used. Window size of 2 seconds was used. As shown in the following tables 3 and 4, the results are summarised.

| Model | Accuracy | Precision | F1-Score | Recall |
| --- | --- | --- | --- | --- |
| SVM | 0.67 | 0.73 | 0.61 | 0.59 |
| K-NN | 0.54 | 0.56 | 0.53 | 0.54 |
| GaussianNB | 0.60 | 0.43 | 0.39 | 0.29 |
| Random Forest | 0.71 | 0.73 | 0.69 | 0.72 |
| Logistic Regression | 0.66 | 0.64 | 0.48 | 0.44 |

*Tabel 3-* Parkinson’s Disease Data Set  *- Features: Average Band Power*

| Model | Accuracy | Precision | F1-Score | Recall |
| --- | --- | --- | --- | --- |
| SVM | 0.54 | 0.24 | 0.12 | 0.10 |
| K-NN | 0.57 | 0.63 | 0.58 | 0.59 |
| GaussianNB | 0.52 | 0.08 | 0.06 | 0.06 |
| Random Forest | 0.69 | 0.71 | 0.68 | 0.72 |
| Logistic Regression | 0.58 | 0.60 | 0.36 | 0.30 |

*Tabel 4-* Schizophrenia Data Set  *- Features: Average Band Power*

The results were underwhelming this time for both the cases with Random Forest again performing better than the rest but by a small margin only.

The One dimensional convolutional neural network was tested on the Schizophrenia data set. It showed promising results with accuracy close to 0.81 with validation accuracy of 90% and loss of 0.30.

Chorono-net implementation was also tested on both the data sets to evaluate its performance. It is originally designed to detect abnormal EEG Signals. Our results were a little less impressive than those in the paper. Our accuracy for the PD data set was 0.82, and for the SZ data set was 0.81

This study shows that even simple Machine Learning techniques can be very helpful in speeding up the diagnostic process. Techniques such as Random Forest can give promising results while taking less resources and time compared to other sophisticated deep learning models. This can be an especially helpful tool for Psychiatrists and doctors in places like the developing countries where the number of mental health workers are severely limited.

**Limitations of the study**

This study was conducted as a solo as a summer internship project with serious time limitations, lack of funding. Furthermore, it had a small data set, especially for Parkinson's Disease, which had just 26 patients. A better and a larger data set could definitely improve the accuracy of the models trained with less chance of over fitting. Also the time constrained only allowed the testing of limited methodologies. In the future with better funding and larger datasets, further approaches can undoubtedly be tested and investigated to improve the efficiency of diagnosing mental disorders with higher accuracy.

References

1. Lewis DA, Levitt P. Schizophrenia as a disorder of neurodevelopment. Annu Rev Neurosci. 2002;25:409-32. doi: 10.1146/annurev.neuro.25.112701.142754. Epub 2002 Mar 22. PMID: 12052915.
2. Kalia, L.V. and Lang, A.E. (2015) Parkinson’s Disease. Lancet, 386, 896-912.

<https://doi.org/10.1016/S0140-6736(14)61393-3>

1. Poewe, W.; Seppi, K.; Tanner, C.; Halliday, G.M.; Brundin, P.; Volkmann, J.; Schrag, A.E.; Lang, A.E. Parkinson disease. Nat. Rev. Dis. Primers 2017, 3, 17013.
2. Afonso, L.C.S.; Rosa, G.H.; Pereira, C.R.; Weber, S.A.; Hook, C.; Albuquerque, V.H.C.; Papa, J.P. A recurrence plot-based approach for Parkinson’s disease identification. Future Gener. Comput. Syst. 2019, 94, 282–292
3. 16. Rios-Urrego, C.D.; Vásquez-Correa, J.C.; Vargas-Bonilla, J.F.; Nöth, E.; Lopera, F.; Orozco-Arroyave, J.R.
4. Ly, Q.T.; Handojoseno, A.A.; Gilat, M.; Chai, R.; Martens, K.A.E.; Georgiades, M.; Naik, G.R.; Tran, Y.; Lewis, S.J.; Nguyen, H.T.Detection of turning freeze in Parkinson’s disease based on S-transform decomposition of EEG signals.
5. Ly, Q.T.; Handojoseno, A.A.; Gilat, M.; Chai, R.; Martens, K.A.E.; Georgiades, M.; Naik, G.R.; Tran, Y.; Lewis, S.J.; Nguyen, H.T. Detection of gait initiation failure in Parkinson’s disease based on wavelet transform and support vector machine.
6. Yuvaraj, R.; Rajendra Acharya, U.; Hagiwara, Y. A novel Parkinson’s disease diagnosis index using higher-order spectra features in EEG signals. Neural Comput. Appl. 2018, 30, 1225–1235
7. Oh, S.L.; Hagiwara, Y.; Raghavendra, U.; Yuvaraj, R.; Arunkumar, N.; Murugappan, M.; Acharya, U.R. A deep learning approach for Parkinson’s disease diagnosis from EEG signals. Neural Comput. Appl. 2020, 32, 10927–10933
8. Khare, S.K.; Bajaj, V.; Acharya, U.R. Detection of Parkinson’s disease using automated tunable Q wavelet transform technique with EEG signals. Biocybern. Biomed. Eng. 2021, 41, 679–689
9. Cortes, C., and Vapnik, V. (1995). Support-vector networks. Mach. Learn. 20, 273–297. doi: 10.1007/BF00994018
10. Cover, T., and Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transac. Inform. Theory 13, 21–27. doi: 10.1109/TIT.1967.1053964
11. Breiman, L. (2001). Random forests. Mach. Learn. 45, 5–32. doi: 10.1023/A:1010933404324
12. Tolles, Juliana; Meurer, William J (2016). "Logistic Regression Relating Patient Characteristics to Outcomes". JAMA. 316 (5): 533–4. doi:10.1001/jama.2016.7653. ISSN 0098-7484. OCLC 6823603312. PMID 27483067.
13. Zhang, H. (2004). The optimality of naive Bayes. AA 1:3
14. Oh, S.L.; Vicnesh, J.; Ciaccio, E.J.; Yuvaraj, R.; Acharya, U.R. Deep Convolutional Neural Network Model for Automated Diagnosis of Schizophrenia Using EEG Signals. Appl. Sci. 2019, 9, 2870. https://doi.org/10.3390/app9142870
15. Roy, Kiral, Harrer, 2019/01/01, ChronoNet: A Deep Recurrent Neural Network for Abnormal EEG Identification