EEG psychiatric disorder multiclass classification

Callum Carville

***Abstract* – The aim was to explore a range of machine learning (ML) classifiers to compare and distinguish between psychiatric disorders via EEG recordings as part of a proof-of-concept submission. The EEG resting-state recordings for this study consisted of 945 subjects with a split of 850 patients diagnosed with a major psychiatric disorder and 95 healthy controls. The data included a combination of EEG power spectrum density (PSD) and functional connectivity (FC) of frequency bands. The 9 ML classifiers explored included the support vector machine (SVM), decision tree (DT), cat boost, k-nearest neighbour (KNN), lightgbm, elastic net (EN) logistic regression (LR), naïve bayes (NB), random forest (RF) and xgboost. SVM achieved the highest accuracy score via cross-validation with 86.47% ± 0.3% with a mean standard error rate of 0.004%. The results suggest EEG can have a place in helping predict psychiatric disorders with an appropriate balanced dataset with accuracy, recall and f1-scores well above chance.**

I. Introduction

Electroencephalogram (EEG) [1] is recorded via electrodes along with adhesive gel placed on the scalp of the subjects to monitor electrical activity across the various regions of the brain. Signals also referred to as action potentials are produced by neurons [2] which are the main cells of the human nervous system. Action potentials are responsible for our thoughts, emotions, and physical movements meaning with appropriate data and models it could in theory be possible to decipher and decode them to further improve our understanding of the brain. Neurons are comprised of many components including; Dendrites, Nucleus, Axon, Axon hillock, Myelin sheath, Schwann cells, Terminus, Glial cells, and Nodes of Ranvier.

Metal discs along with gel to improve conductivity to the scalp are used to record EEG activity across the various regions and hemispheres of the brain.

Psychiatric disorders [3] are characterised by a clinically significant disturbance in an individual's cognition, emotion, or behaviour. According to the WHO (World Health Organisation) [3] a bold statement was made with supporting data from GHDx (Global Health Data Exchange) [4] [5] that 1 in 8 people suffer or have lived with a mental disorder during 2019. Those suffering from psychiatric disorders have been linked to premature mortality rates according to a report by the Office for Health Improvement & Disparities (2.5 to 7.2 times more likely to die prematurely), this only emphasises the devastating consequences for those suffering.

A further study by the Nation Library of Medicine [6] [7] found patients suffering from severe mental health / psychiatric disorders tend to have poorer general health which indicates a knock-on effect on an individual's overall well-being.

Observations of the Psychiatrists Global Market show an annual growth during 2022-2023 of 11.4% [8] which was estimated to bring the market from $147.28 billion to $164.01 billion. These stats alone show an alarming onboarding of additional patients instead of a decrease indicating that the processes are failing. The need for a change is now otherwise we are draining resources of the economy, reducing the number of individuals fit to work, and most importantly not solving the disorders which serves as a moral duty.

II. Existing work

Prior to the development of this project, several publications that investigated the classification of psychiatric disorders with EEG via ML. The foundation of this project's research was built upon the study [9] conducted by Seoul National University in South Korea. The team behind this study had the aim to develop ML classifiers to detect and compare psychiatric disorders via EEG, this directly correlated to the aim of this paper's research. The team selected SVM, EN and RF as their classifiers. The dataset was acquired via psychiatrists based on the DSM-IV [10] criteria along with the MINI [11] interview during psychological assessments. The final diagnosis was confirmed by 2 psychiatrics and 2 psychologists based on the review of the original patient diagnosis and psychological assessments conducted 1 month prior and post EEG recording. Subjects were excluded from the study if there was a history of brain trauma or if their IQ was below 70 reflecting mental retardation.

The study contained 945 subjects (850 with a disorder and 95 healthy controls), subjects were both male and females of the age range 18-70 years. The split of the specific disorders was; Acute stress (n=38), Adjustment (n=38), Alcohol use (n=93), Behavioural addiction (n=93), Bipolar (n=67), Depression (n=199), obsessive compulsive disorder (OCD) (n=46), Panic (n=59), Post traumatic stress (PTSD) (n=52), Schizophrenia (n=117) and Social anxiety disorder (n=48). The dataset developed contained 5 minutes of eyes-closed resting-state with 19 or 64 channels acquired with 500-1000Hz sampling rate via the Neuroscan technology. Other data recorded included age, gender, date of EEG recording, education, IQ, disorder, specified disorder, and EEG site recordings.

EN along with the IQ adjustment performed the best with the team also denoting the ideal feature combinations which consisted of Schizophrenia PSD alpha with a 93.83% accuracy, for Trauma and Stress-related disorders the best feature was beta FC achieving a 91.21%, for Anxiety disorders the whole band of PSD reflected the best features with the accuracy of 91.03%, for Mood disorders theta FC was the best feature combination achieving 89.26% accuracy, Addictive disorders had theta PSD as the best features with 85.66% accuracy and for OCD disorder gamma FC were the best features achieving 74.52% accuracy. SVM achieved accuracy scores of 86.02% ± 8.89% and RF = 87.18 ± 8.08%.

The study concludes more severe disorders such as Schizophrenia and PTSD were easier to discriminate against with the team stating this may be a factor of the disorders being associated with altered brain activity. They mentioned the research could be extended further by attempting multi-class classification methods or other classifiers which is where this study comes in.

A second publication [12] focused on the binary classification of patients diagnosed with Depression versus healthy controls via EEG with the ML classifiers SVM, LR and NB. This team conducted a study of 64 subjects, 34 with Depression (18 females and 16 males) and 30 healthy controls (9 females and 21 males). The patients had to meet the diagnostic criteria for DSM-IV [10] whilst the healthy controls were examined for clinical symptoms to exclude the possibility of any mental disability which could skew the results of their research.

The EEG consisted of a 5-minute rest-stating with eyes closed and eyes open as the 2 separate conditions. The data was acquired through 19 electrodes covering the scalp with placements being determined by the 10-20 electrode placement standards. A 50Hz notch filter was applied and a sample rate of 256 per second was used to record the electrical activity of the subject’s brain. The EEG recordings were processed to remove artifacts that could skew results, these artifacts may include blinking, muscle activations, or even the subject’s heart beating.

10-fold cross validation was utilised to evaluate classification performance which meant the training testing split was 90% for training and 10% for testing which is above the suggested practice and opinions on ResearchGate however this was a small dataset meaning data was limited.

The researcher’s study proved a success as they achieved accuracy scores well above the realm of chance with SVM performing the best with an accuracy score of 98%, Logistic Regression reinforced the success of the team with 91.7% accuracy and lastly, Naïve Bayes achieved a score of 93.6% accuracy. They concluded by stating SL could be a promising method for diagnosing patients which could become a generalised approach with further tool developments.

As part of identifying other existing works with EEG ML classification a third article [13] was found where the research performed focused on those suffering from Schizophrenia performing working memory tasks.

This study consisted of 40 diagnosed with Schizophrenia patients along with 12 healthy controls whom all completed a working memory task (Sternberg Working Memory Task) whilst being recorded with EEG. SVM [17] was the classifier choice for this research and the team were able to successfully distinguish subjects suffering from Schizophrenia with 87% accuracy.

The patients diagnosed with Schizophrenia were required to meet the DSM-IV [10] criteria before being eligible to take part in the study. Subjects were also included if they met the age requirements which were marked for subjects to be at least 18 years old, be native English speaking, and have a stable housing situation for a minimum of 30 days. Patients diagnosed with the disorder were excluded if their medication had changed within the last 30 days or if they abused alcohol or drugs, had previous brain trauma, or suffered from mental retardation.

To extract the EEG recordings all participants were placed in front of a 24” monitor at a viewing distance of 1metre in a dimly lit room. EEG was recorded with a 64-channel BioSemi ActiveTwo bio-amplifier with electrodes placed according to the 10-20 system. Extra electrodes were placed at mastoids to act as a reference point, on the outer of both eyes, above and below the right orbit. EEG was acquired with a sampling rate of 1024Hz per second

This study demonstrates how ML architectures along with EEG can be used for feature detection in a binary classification scenario with SVM achieving an 87% accuracy score. Several publications have utilised SVM for EEG classification tasks indicating it complements the data type well.

Other literature were explored for the development of the report with one being focused on observations of how ML architectures have been applied for both diagnostic and predictive use for mental/psychiatric disorders.

The team behind this research [14] made the bold statement that the acquisition and pre-processing of EEG signals were sufficient in many studies but many lacked systematic characterisation of clinical features and many models were inappropriately used with flawed testing metrics. With this 11 suggestions for researchers to improve ML models with EEG studies were made and these are; using clear terminology, being precise when describing clinical samples with identification of cofounding variables, validating the diagnostic procedures against international standards, following EEG standards for recording and processing, explore data augmentation, select a clear model strategy and make sure to test the model, ensure that test data and training data are independent, make sure to identify and balance cofounding variables, select appropriate scoring measures for reporting such as F1-score, analyse and report the influence of hyperparameters and lastly improve transparency via in-depth descriptions of the models and make code publicly available. This publication influenced the project as a set of standards to adhere to. (Please view supporting materials for additional existing work articles and breakdowns).

III. Methodology

The CRISP-ML [15]- [16] methodology was utilised for the project’s life cycle as it complimented the architectures and ML algorithms that were implemented. CRISP-ML methodology consists of seven stages **Figure 1**:

Business Understanding – Within this stage, the problem or opportunity is identified using ML techniques and the project objectives and success criteria would be set.

Data Understanding – Data is collected and observed to improve understanding whilst determining relevance to the problem set in the prior stage. The data is investigated to determine if there are quality issues or limitations which would need to be addressed.

Data Preparation – With a dataset selected, preprocessing is used to clean the data, handle the missing values and outliers with other inconsistencies.

Modelling – ML algorithms are selected based on the datatype and problem type which in this case is a classification problem. The data is split into training, validation, and testing samples (this is important for the evaluation step). The model is then developed with the training data and optimised by fine-tuning the model’s hyperparameters to achieve better performance whilst avoiding overfitting or underfitting. Models are compared with one another, and their performance and accuracy are used as the metrics to determine which model performed the best.

Evaluation – The model’s performance is evaluated against the test set and the success of the model will be determined by the objectives previously set which is to develop a classifier that can accurately distinguish between a range of psychiatric disorders.

Deployment – With the model defined it would next be integrated into a production environment with the appropriate infrastructure and software.

Maintenance – The model would then be monitored and updated to ensure the performance remains optimal and achieve a high level of accuracy.

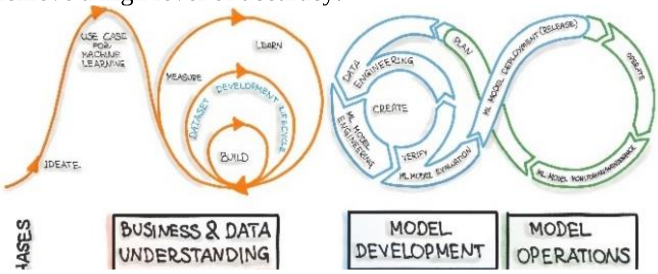

FIGURE 1 - CRISP ML LIFECYCLE DIAGRAM

IV. Technologies used

ADMI PC Intel 4.4GHz QUAD Core, GTX 1650 4GH with 16GB ram PC was the machine used to run the high dimension classifier and was a vital aspect to the project as a regular standard office PC would not have coped with the permutations of the SHAP API [17] or the high dimensions of the various models (additional technologies can be viewed in the supporting documentation).

Jupyter Notebook was the IDE used to write Python scripts as its interface was user-friendly and the software allowed for the creation of visuals to compliment the research. [18]

Sklearn train_test_split was used to split the array into training and testing data so that the models could be validated for generalisation and determine their accuracy with unseen data. [19]

Sklearn confusion matrix is an API that was used to obtain and evaluate the predictions of the classifiers. It does this by evaluating the quality of the output of the classifier on the dataset. Diagonal numbers represent the quantity of points in which the predicted label matches the true label and vice versa for off-diagonal numbers. [20]

Sklearn classification report API was used to measure the quality of predictions from the algorithm classification performance. The report presents findings for precision, recall, f1 score and support. [21]

Sklearn cross_val_score was used to extract the accuracy score for the model after performing cross-validation of the training and testing data to retrieve a mean average accuracy score that would represent the model's performance.

Sklearn.svm SVC is an algorithm that constructs a hyperplane in the multidimensional feature/weight space to divide the dataset into separate classes. This algorithm was used with several others in an attempt to perform multi-class classification tasks and help determine the disorder of each patient based on EEG recordings. [22]

SHAP API was used to measure feature importance by performing permutation on the models of the various algorithms selected. This API also provided the option for visualisations to help explain and understand the classifiers better. [17]

Optuna was used to find the best hyperparameters for each of the 9 classifiers. The API allows for variations of the model to be run whilst extracting the best results via tuning of hyperparameters and this resulted in an improved overall performance of the model for 6 out of 9 classifiers. [23]

V. Dataset

EEG data recordings of multiple variants of psychiatric disorders with diagnosis are in very limited supply to the public. However, the dataset for this project was retrieved from the highlighted existing work publication as it was made available to the public. This dataset contained a range of psychiatric disorders with other potentially important features including age, gender, date of EEG recording, education, IQ, disorder, specified disorder, and EEG site recordings. The data consisted of 945 subjects with 850 patients diagnosed with major psychiatric disorders and 95 healthy control patients [9]. The EEG signals were retrieved from rest-state, and this was consistent throughout.

The dataset was developed and published by Su Mi Park [24] an employee of ORCiD (an organisation for connecting research and researchers) [25] along with the Department of Psychiatry in Borame Medical Center (Seoul, Republic of Korea), the Department of Statistics in Ewha Womans University (Seoul, Republic of Korea), Department of Psychiatry and Behavioural Science in National University College of Medicine (Seoul, Republic of Korea) and the Institute of Human Behavioural Medicine in National University Medical Research Center (Seoul, Republic of Korea) it was assumed the data recorded can be trusted. Unfortunately, the raw [26] EEG recordings could not be sourced, the researchers who developed the dataset were contacted and there was no response. As a result, the project was forced to work with the pre-processed data which does not include the time series EEG recordings for the channels only the extracted means of each channel per band.

VI. Data preparation

The dataset contained fields that were not marked to act as features/weights in the development of classifiers for this project as EEG recordings [26] along with the diagnosis of the specific disorder were the only requirements. The reason being is that as part of the objectives the aim was to determine if it is possible with EEG recordings [26] to classify an individual with the correct disorder, this way the model's ability to become more dynamic and generalised to the public is possible.

Prior to collecting this dataset, the research team who developed it also performed pre-processing of the EEG signals [26] and removed artifacts [27] from the data. Artifacts are signals recorded by the EEG equipment that are not generated from the brain, these artifacts can skew results when utilising the data, so removal is very important.

With irrelevant data removed from the dataset, a check was performed for null values (values that are missing from the dataset which could impact results) which returned 0 for missing entries, this keeps the data supplied to the model consistent as there are no missing values

The specified disorder column contained data relating to the disorder of the individual that was recorded, I modified the naming conventions of the disorders to remove white spaces in the data as [28] IDE compilers often have difficulty handling spaces and when replacing the titles, I used a camel case approach to the naming conventions to match industry-standard software engineering practices.

String-based specified disorder values were converted into a binary format with the reason being that ML algorithms cannot be applied to string type data.

Finally, there was a check for outliers to determine if there were any discrepancies in the EEG recordings of individuals or instances where there is little supporting data as models are data-driven. There were no outliers found within the dataset.

VII. Data visualisation & understanding

The resulting processed dataset left 946 rows with 1142 columns with 1 column dedicated to the disorder diagnosis, 114 [29]PSD EEG columns (19 electrodes * 6 frequency bands) and 1026 columns representing coherence of EEG (measured between every pair of electrodes for each frequency band making the overall calculation 171 * 6).

The power spectral density (PSD) columns provide a method for representing the distribution of an EEG signals frequency making them more interpretable and their values represent the Watts/Hz. Each PSD measures the signal of power contributed by frequencies within a band.

Coherence measures the synchronisations between signals of two different electrodes and is based on phase consistency. Frequency and phase consistency are derived from the EEG time series data.

The specific disorder column contained a healthy control along with EEG recordings for individuals with PTSD, Schizophrenia, Depression, Social anxiety disorder, Bipolar disorder, OCD, Alcohol use disorder, panic disorder, adjustment disorder, behavioural disorder, and acute stress disorder. [30]- [31]

Figure 2 displays the balance of the disorder count, the main disorder is highlighted on the left side as Depression [30] with the lower-end counts being Acute stress disorder [32] and Adjustment disorder [31]. It can be argued from observations that the data could suffer from impartial results in the ML classifiers as the range of disorders do not have an even distribution of input data that can be used for training and testing. With data of this kind being in limited supply it was decided to make use of all data available as to allow for further potential distinctions and pattern detection when analysing with modern ML algorithms.

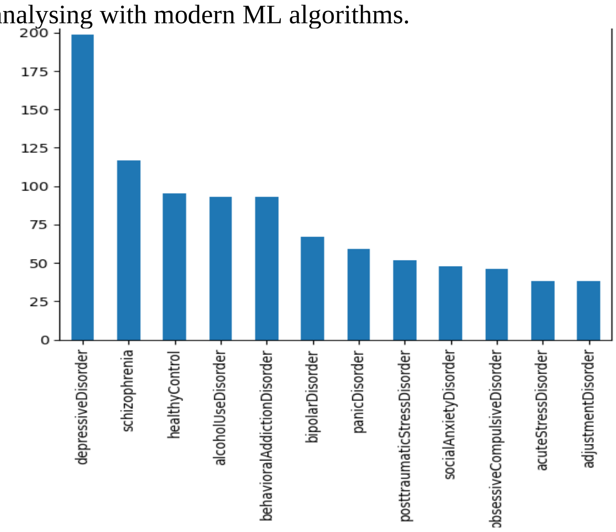

FIGURE 2 - PLOT TO DISPLAY QUANTITY OF DISORDER

The 'Synthetic minority oversampling technique (SMOTE) Tomek' [33] method from imbalanced learn Application programming interface (API) was used to perform oversampling and balance the training data. The screenshot below displays the quantity of each disorder post-processing leaving a much larger overall dataset that is balanced. This API works by selecting a minority class instance at a random point and locating its KNN [34]. The synthetic instance is then developed by selecting one of the KNN at random to form a line/hyperplane segment in the feature space. This more than doubles the number of disorders for the overall dataset by bringing it from 945 to 2400 **Figure 3**.

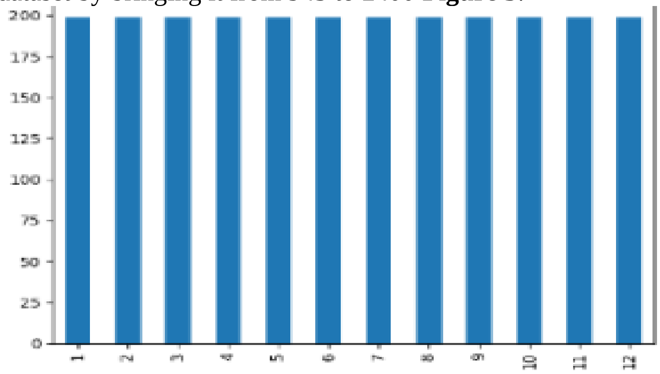

FIGURE 3 - PLOT OF SMOTE QUANTITY OF DISORDER

The EEG data included 5 minutes of eyes-closed resting state with 19 channels acquired via 500 – 1000Hz sampling rate with 0.1 to 100 on-line filters with Neuroscan [35]. The researchers managed to keep electrode impedances below 5 k by applying electrical conductivity gel. 19 channels (FP1, FP2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, and O2) were selected with a basis of the 10 – 20 [36] system in conjunction with a mastoid reference electrode **Figure 4**. (Please read supporting materials for electrode breakdown).

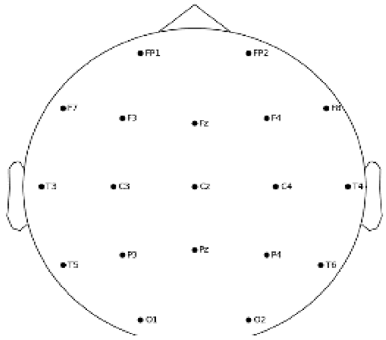

FIGURE 4 - VISUAL OF ELECTRODE PLACEMENT

Figure 5 displays all specific disorders against their waveform activity (red indicating higher electrical activity). This helped to acknowledge how on each waveform [37] there were altering levels of activity in relation to the specific disorder recorded in various brain regions. This indicated that with data exploration and an appropriate classifier it would be possible to extract a model that offers consistent accuracy when performing classification of all disorders.

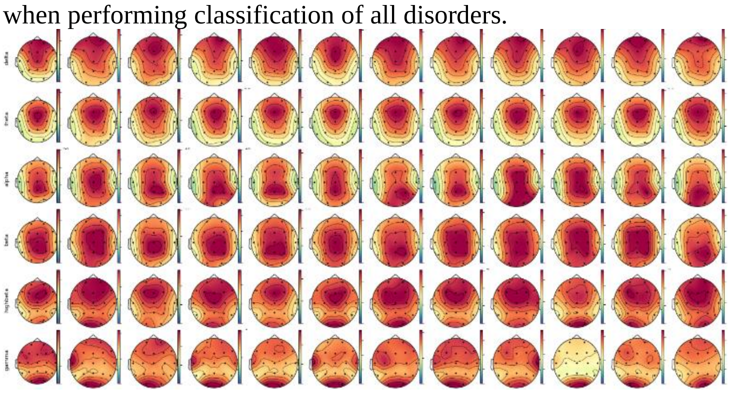

FIGURE 5 - EEG ACTIVITY FOR BRAIN WAVES FOR SPECIFIC DISORDERS

VIII. Data flow

Figure ***6*** below was developed to help illustrate how the data acquired is processed from start to finish for this project. With the business understanding stage in the lifecycle being met by finding an opportunity to build upon existing research the next step was to utilise the dataset to develop an understanding and preparation for classification. With the data prepared it was then passed through a series of steps as can be seen below for the modelling and evaluation. The project was not deployed to a production environment for which an explanation can be found in the conclusion of this report. Without deployment, the maintenance is voided.

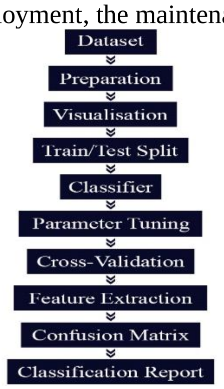

FIGURE 6 - DATA FLOW AND PROJECT FLOW

XI. Model architecture

With the aim of the project requiring the ability to classify a specific disorder from a range of psychiatric disorders [3], selecting an appropriate ML classifier would be vital. Unlike the published paper in which the dataset was obtained from this projects research will be restricted to only using the specific disorder column along with the EEG data as the sole training data as one aim of the research is to determine if it is possible to diagnose an individual with above-chance accuracy via only observing their EEG signals.

According to Simon Tavasoli (data science advisor and instructor) [38] [39] and Rob Schapire (Princeton University) the best supported ML algorithms for classification are LR [40], DT [41], SVM [22], RF [42], NB [43], and KNN [44]. For the purposes of condensing the report, only 1 algorithm will be highlighted which is SVM [22] (please view supporting materials for in-depth classifier breakdowns).

The SVM [22] (developed by Corinna Cortes and Vladimir Vapnik 1995) classifier locates a hyperplane in a multi-dimensional space. The aim of the hyperplane is to maximise the distance between data points of different classes (disorder in this project). Data points that fall on either side of a hyperplane can be assigned to different classes and the number of dimensions of a hyperplane depends on the number of features/weights.

Unlike LR [45] where the output of the linear function is within the range of 0 – 1 using the sigmoid function, SVM [46] takes the output of the linear function and if the output is greater than 1 it is identified with one class and if it is -1 it is identified with another class meaning in SVM the range of values is between -1 and 1. With the aim of maximising the margin between data points and the hyperplane, a loss function referred to hinge is used.

If the predicted value and actual value are of the same sign, then the cost will be 0 and if not, then a calculation is performed to determine a loss value. A regularisation parameter called the cost function is added to balance the margin maximisation with the loss. With the loss function, a subset of derivatives with respect to the features/weights are extracted to find the gradients, with the use of the gradients the weights can be updated.

Once the model can correctly predict the class of the data point, the classifier will require only updating the gradient from the regularisation parameter. If there is a misclassification with the prediction of a class, the loss and regularisation parameter is included in the algorithm to perform the gradient update.

X. Model performance

TABLE 1 - 70% TRAINING 30% TESTING SPLIT ESTIMATED RESULTS

Model Architecture	Base Model Accuracy	Under Sampling Accuracy	Over Sampling Accuracy	Over Sampling Hyper Tune Accuracy
Cat Boost	21.478%	16.788%	**81.589%**	**82.984%**
Decision Tree	12.323%	11.678%	50.069%	51.324%
K Nearest Neighbour	16.901%	10.218%	54.811%	**82.845%**
Light GBM	20.070%	13.138%	**81.450%**	79.218%
Logistic Regression	19.718%	17.518%	80.613%	79.916%
Naïve Bayes	8.450%	15.328%	25.104%	21.338%
Random Forest	19.014%	6.569%	79.079%	24.825%
Support Vector Machine	17.605%	09.489%	**83.403%**	**84.518%**
XG Boost	22.183%	12.408%	77.405%	78.940%

Comparison of results shown in Table **1** from utilising a training test split of 70% for training and 30% for testing have indicated it is not possible to achieve a reliable multiclass classifier with the limited imbalanced data generated from the study of psychiatric disorders by Seoul National University (the best result being from [47] - [48] XGBoost with 22.1% accuracy). However, with such an investment into a long-standing study by the university, the project was compelled to interrogate the data further in an attempt to understand how this data could still be useful and not made redundant.

Upon further investigation it was noted that the data was heavily unbalanced, this affected the model's ability to differentiate between disorders as the algorithms input training data would automatically heavily lean towards Depression (199 out of 945) or Schizophrenia (117 out of 945) as they dominated in terms of quantity of recordings.

To combat the imbalanced classes two techniques were implemented under and over sampling. Unfortunately, the implementation of under sampling (reduced all classes to an equal quantity of 35 instances) with the Imbalanced learn API lost accuracy on all classifiers excluding Naïve bayes [49] [50] which gained an estimated 6% accuracy bringing its overall total to 15.3%. With under-sampling having no real impact to drastically improving overall model accuracy it was clear that not only was there an issue with the balance of data but also the limited quantity.

This was confirmed when implementing the over-sampling technique with the SMOTE API [33] which increased all class instances to an even distribution of 200 per class making a grand total of 2400 instances. How the new instances for each class were generated was via the KNN [34] [44] algorithm and generating new artificial data on a hyperplane that would not upset the overall feature space. This technique exposed the differentials between classes drastically and increased model performance in the best-performing techniques such as Cat Boost [51]- [52] from 21.47% to 81.58% and SVM from 17.6% to 83.4%. This technique exposed class differentials meaning the project's initial goals could still be met which included the development of a POC for a ML classifier that uses EEG [1] data to perform multi-class classification on a range of disorders with a high level of accuracy. Satisfied with the results from over-sampling on base models without hyper tuning the next step was to investigate if there was potential for increased gains via hyperparameter tuning.

Optuna API [23] was used to tune the hyperparameters via a test and trial process that would modify the parameters between the ranges provided and then test against accuracy after each trial. As there were 9 classifiers and the project was time restricted the number of trials per model was set to 30, ideally with more time the number of trials would have been set to a minimum of the squared number of parameters by their range. Results from the trials did not impact the top-performing classifiers significantly but did add an extra estimated 1% accuracy to the Cat Boost [51]- [52] and SVM [22] models and a significant improvement from 54.81% to 82.84% in KNN [34] [44] .

SVM [22] performed the best with the [1] EEG data achieving 84.51% accuracy which is a testament to the amount of research and work carried out to find differentials when exploring and interrogating the dataset. Cat Boost. [51]-[52] followed closely in second place with 82.98% accuracy further emphasising with the correct set of classifiers patterns and relationships in EEG data can be identified.

XI. Model confusion matrix

The confusion matrix [20] is used to derive additional understanding from the classifier's strengths and weaknesses. The confusion matrix was created by applying the true (rows) value against the predicted values (columns) of the model to evaluate areas of confusion. All values off the diagonal represent an error / a form of confusion by the classifier.

When observing the confusion matrix *Figure 7* for the SVM [22] model it is clear a large proportion of the confusion came from classifying the Depression disorder, as every cell off the diagonal for the true value of Depression contained a value indicating a repeat error (cells associated with 5). This repeat pattern could be speculated that either Depression shares a lot of commonalities with other disorders or that the data set is not sufficient enough to reach higher levels of discrimination.

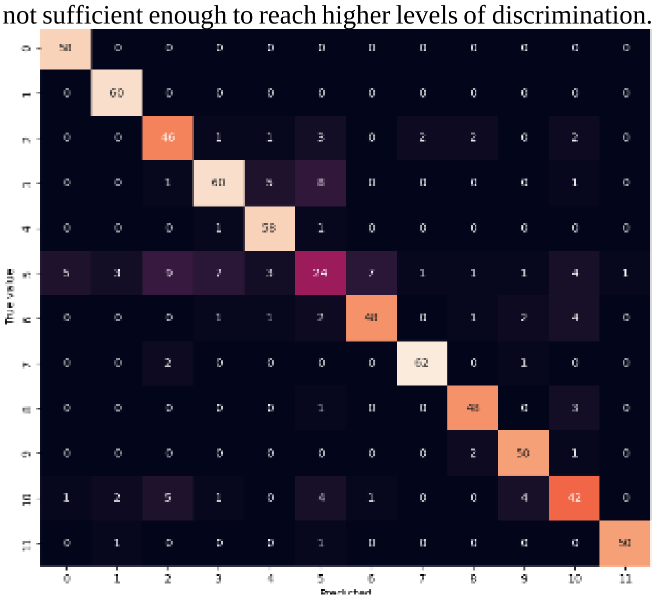

FIGURE 7 - CONFUSION MATRIX FOR SVM

XII. Model classification report

The classification report [53] consists of a Precision, Recall and F1-score which all measure the model's accuracy. Precision measures the number of positive instances that were correctly identified, Recall measures the proportion of actual instances correctly identified and F1-score is the mean of the precision and recall scores. The number values (1-12) on the left side represent each of the disorders that have been classified along with the healthy controls.

When observing the SVM [22] classification report *Figure 8* it further emphasises how the model had difficulty classifying Depression [30] disorder against others and in turn, seems to have impacted Alcohol use disorder [54] and Schizophrenia [55]. All other disorders achieved a respectable level of accuracy in the report reflecting in an above-average model.

	precision	recall	f1-score	support
1	0.91	1.00	0.95	58
2	0.91	1.00	0.95	60
3	0.73	0.81	0.77	57
4	0.85	0.80	0.82	75
5	0.85	0.97	0.91	60
6	0.55	0.36	0.44	66
7	0.86	0.81	0.83	59
8	0.95	0.95	0.95	65
9	0.89	0.92	0.91	52
10	0.86	0.94	0.90	53
11	0.74	0.70	0.72	60
12	0.98	0.96	0.97	52
accuracy			0.85	717
macro avg	0.84	0.85	0.84	717
weighted avg	0.84	0.85	0.84	717

FIGURE 8 - CLASSIFICATION REPORT FOR SVM

XIII. Cross validation

As the development of the classifiers came at a later stage in the project it was noted that they were all developed with the train_test_split method [97] with a random_state parameter set, this meant every time the classifier was trained it was trained from new but with the same 70% of training data to allow for consistency which made it possible to access results at all stages of the classifier development from base to where over-sampling and hyperparameters were tuned without bias in the training data. One downfall of this is the model's true potential may not yet have been realised as the 30% of data used for testing may have contained further information which could help in the classification of psychiatric disorders.

The solution to this was to implement repeated K fold cross validation [97] [98]. Repeated K-fold cross-validation is a strategy to improve the model's estimated performance as it repeats the cross-validation method several times to compute a mean result across all K folds and reduce noise. Being conscious of time the number of folds for cross-validation was limited to 5 giving a training and test split of 80/20% which from opinions on ResearchGate seems to be the most common option. Repeats were set to 5 to reveal the mean for classifier accuracy more precisely and the standard error that is estimated to be expected from the model. This resulted in a peak of 86.77% accuracy with the SVM classifier.

XIV. Feature extraction

SHAP API performed permutations on all classifiers with 2 experiments to retrieve the top 10 and 100 best-performing features/weights Table 2. With these two sets of features, the classifiers were retrained with the intention of achieving high accuracy scores, a faster time for training as less data would be used and quicker convergence making the overall classifier lighter and faster. This experiment requires further investigation as the accuracy score plummeted and permutation time was costly with the SVM classifier being estimated to require more than a week to process and with time being limited the experiment was cut short. Cat Boost achieved the top accuracy score of 22.53% with 100 features, this further emphasises the requirement for further investigation with classifier performance depleting.

TABLE 2 - SHAP API 100 VS 10 FEATURE PERFORMANCE COMPARISON

Model Architecture	Over Sampling Hyper Tuned Accuracy 10 Features	Over Sampling Hyper Tuned Accuracy 1 Features
Cat Boost	20.774%	**22.535%**
Decision Tree	14.788%	13.380%
K Nearest Neighbour	18.309%	10.915%
Light GBM	18.309%	**22.183%**
Logistic Regression	**21.478%**	19.014%
Naïve Bayes	10.915%	10.211%
Random Forest	**22.183%**	**21.478%**
Support Vector Machine	Requires further computation time	Requires further computation time
XG Boost	15.845%	21.126%

XV. Limitations

As previously highlighted the data was not provided in the raw format which would be time-series recordings. EEG recordings are limited in terms of public availability making it difficult to source and compare against other potential recordings to determine the quality of data. The formatting of the recordings heavily impacts the performance of the classifier as artifact removal which requires human intervention can lead to the removal of key differentials and vital information.

Other limitations are involved in the acquiring of EEG data [26] for psychiatric disorders [3] which rely on professional diagnosis from both a psychiatrist or psychologist to confirm in accordance with the DSM-IV [10] standards. With trained professionals determining the diagnosis for the disorder there is always the possibility of human error and misdiagnosing the patient.

The patients recorded for the dataset did not have their prescriptions restricted which can heavily impact the neural activity of the subject being recorded with EEG.

XVI. Ethical & societal challenges

Several challenges surround the acquirement of EEG recordings [26] including the potential for a patient to suffer from an episode during a recording if they suffer with Schizophrenia [56] which may put the patient and others in danger, additionally panic attacks may occur from social anxiety by interacting with the team making the recordings or placing the electrodes on the scalp of the patient. A patient may also suffer from an allergy attack if they are not compatible with the adhesive gel or the material of the electrodes being placed on the scalp.

Patient confidentiality and GDPR [57] regulations would also require compliance in order to make use of the data with consent from the patient to use the data for its intended purpose otherwise those performing the recordings may face legal consequences.

XIV. Discussion & conclusion

Results from torturing data did reveal hidden relationships and differentials helping to meet a POC set of requirements and prove that multiclass classification of Psychiatric disorders is possible with a balanced dataset that contains sufficient levels of available recordings.

However, in this instance the dataset from Seoul University was not sufficient to develop a classifier fit for generalisation as it was both unbalanced and limited in terms of the number of disorders recorded. Development with SVM [22] and Cat Boost [51] [58] classifiers did reveal a potential avenue to explore once better public EEG data is made available as both indicated above chance levels of accuracy and precision during the classification stage. The KNN classifier also indicated how a much simpler algorithm could achieve above-average accuracy in keeping with the best-performing classifiers of this research.

Overall, the project shows a lot of promise by meeting POC (read supporting materials for more information) requirements to inspire further exploration in this area of research with the aim of an improved understanding of both psychiatric disorders with alternative classifiers for diagnosis by utilising [26] EEG. The research carried out is reproducible as all code and experiments were provided with the report long and in-depth breakdowns of classifiers selected for research can be viewed in the supporting materials, this all helps to meet the standards mentioned in the article by Mateo de Bardeci and his team [14]. The project unfortunately did not achieve 95% accuracy in classification and is not fit to generalise as a result of limitations with the dataset.

XVII. Future research

To further expand this project additional trials would be run during the hyperparameter tuning stage as the trials were limited thus not fully exploiting the potential of the classifiers and this could be the difference in achieving higher accuracy and precision scores. During the cross-validation stage, a limited time meant the number of repeat experiments was restricted to 5 which ideally would have been 10 to fully explore the 9 classifiers with the dataset. There are other cross-validation methodologies that could have been explore with more time e.g., Stratified K-fold or leave-one-out. The SHAP API needs to be explored further for feature extraction to determine what data is redundant via achieving a high level of accuracies which from the experiments in this research was not achieved directly related to the number of features selected (both 10 and 100 of the top features were utilised). The dataset was also limited and unbalanced so with a fundraiser additional subjects could be recorded not only to balance the dataset but improve the amount of data the classifier can learn from. Lastly, more extreme testing of the classifier would take place to verify its fitness for generalisation. [59]

References

[1] NHS, "Electroencephalogram (EEG)," [Online]. Available: https://www.nhs.uk/conditions/electroencephalogram/.

[2] P. Dave, "Neural Conduction, Action Potential and Synaptic transmission," [Online]. Available: https://www.youtube.com/watch?v=zHJ3h675nNk&list=LL&index=46&t=3s.

[3] World health organisation, “Mental Disorders,” [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/mental-disorders. [Accessed 20 08 2023].

[4] GHDx, “GHDx,” [Online]. Available: https://vizhub.healthdata.org/gbd-results/.

[5] UN News, “Nearly one billion people have a mental disorder: WHO,” [Online]. Available: https://news.un.org/en/story/2022/06/1120682.

[6] C. C. &. Consulting, “HOW SELF-ESTEEM RELATES TO PHYSICAL & MENTAL HEALTH,” [Online]. Available: https://www.cognition.org.uk/coaching-hypnotherapy/blog/how-self-esteem-relates-to-physical-mental-health#:~:text=Self-esteem%20is%20related%20to,poor%20physical%20and%20mental%20health.

[7] Gov.uk, “Premature mortality in adults with severe mental illness (SMI),” [Online]. Available: https://www.gov.uk/government/publications/premature-mortality-in-adults-with-severe-mental-illness/premature-mortality-in-adults-with-severe-mental-illness-smi.

[8] Research and markets, “Psychiatrist Global Market Report,” [Online]. Available: https://www.researchandmarkets.com/report/psychiatry.

[9] B. J. D. Y. O. Su Mi Park, “Identification of Major Psychiatric Disorders From Resting-State Electroencephalography Using a Machine Learning Approach,” [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.707581/full.

[10] K. A. P. D. B. K. F. J. Z. S. a. K. S. K. Dan J. Stein, “What is a Mental/Psychiatric Disorder? From DSM-IV to DSM-V,” [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3101504/.

[11] Y. L. K. H. S. P. A. J. J. E. W. T. H. R. B. G. C. D. D V Sheehan 1, “The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10,” [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/9881538/.

[12] S. S. A. A. M. A. M. Y. &. A. S. M. Wajid Mumtaz, “A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD),” [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28702811/.

[13] J. B. R. J. J. G. K. &. C.-M. A. C. Jason K Johannesen, “Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults,” [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27375854/.

[14] C. T. I. S. O. Mateo de Bardeci, “Deep learning applied to electroencephalogram data in mental disorders: A systematic review,” [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301051121001101.

[15] R. Wirth, “CRISP-DM,” [Online]. Available: https://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf.

[16] WSSU.Edu, “Research methologies,” [Online]. Available: https://www.wssu.edu/about/offices-and-departments/office-of-sponsored-programs/pre-award/_Files/documents/develop-quantitative.pdf.

[17] W. ROSINSKI, “SHAP API feature importance,” [Online]. Available: https://www.kaggle.com/code/wrosinski/shap-feature-importance-with-feature-engineering.

[18] Jupyter, “Jupyter,” [Online]. Available: https://jupyter.org/.

[19] sckit -learn org, “train test split,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html.

[20] scikit learn, “Confusion Matrix,” [Online]. Available: https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html.

[21] scikit learn, “Classification report,” [Online]. Available: https://www.scikit-yb.org/en/latest/api/classifier/classification_report.html.

[22] scikit learn, “SVM,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html.

[23] Optuna, “Optuna,” [Online]. Available: https://optuna.org/.

[24] OSF Home, “Su Mi Park,” [Online]. Available: https://osf.io/pm3yt/.

[25] ORCid, “ORCid,” [Online]. Available: https://orcid.org/0000-0001-5200-9507.

[26] Mayoclinic, “EEG explained,” [Online]. Available: https://www.mayoclinic.org/tests-procedures/eeg/about/pac-20393875.

[27] National Library of Medicine, “Common Artifacts During EEG Recording,” [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK390358/.

[28] JetBrains, “JetBrains Pycharm,” [Online]. Available: https://www.jetbrains.com/pycharm/.

[29] SIEMENS, “Power Spectral Density,” [Online]. Available: https://community.sw.siemens.com/s/article/what-is-a-power-spectral-density-psd.

[30] NHS, “Depression In adults,” [Online]. Available: https://www.nhs.uk/mental-health/conditions/depression-in-adults/overview/.

[31] National Library of Medicine, “Adjustment Disorder,” [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3701359/.

[32] National Library of Medicine, “Acute Stress Disorder,” [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK560815/.

[33] J. Brownlee, “SMOTE for imbalanced classification,” [Online]. Available: https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/.

[34] IBM, “What is the k-nearest neighbors algorithm?,” [Online]. Available: https://www.ibm.com/topics/knn.

[35] NeuroSky, “NeuroSky explaing waves,” [Online]. Available: https://neurosky.com/2015/05/greek-alphabet-soup-making-sense-of-eeg-bands/#:~:text=In%20our%20previous%20blog%2C%20we,waveform%20captured%20at%20an%20electrode.

[36] Research Gate , “10-20 Electrode Placement rules,” [Online]. Available: https://www.researchgate.net/figure/The-placement-positions-of-EOG-Fp1-and-Fp2-electrodes-E1-and-E2-are-the-left-and-right_fig1_335074690.

[37] S. M. O. S. M. D. H. &. J. V. M. Steven X. Moffett, “Dynamics of high frequency brain activity,” [Online]. Available: https://www.nature.com/articles/s41598-017-15966-6.

[38] S. Tavasoli, “Simon Tavasoli,” [Online]. Available: https://www.linkedin.com/in/simontavasoli/?originalSubdomain=ca.

[39] S. Tavasoli, “Top machine learning algoritms,” [Online]. Available: https://www.simplilearn.com/10-algorithms-machine-learning-engineers-need-to-know-article.

[40] sci kit learn, “Linear Regression,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html.

[41] Scikit learn, “Decision Trees,” [Online]. Available: https://scikit-learn.org/stable/modules/tree.html.

[42] IBM, “What is random forest?,” [Online]. Available: https://www.ibm.com/topics/random-forest.

[43] Z. Zhang, “Naive Bayes Explained,” [Online]. Available: https://towardsdatascience.com/naive-bayes-explained-9d2b96f4a9c0.

[44] Scikit learn, “K nearest neighbours,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html.

[45] IBM, “What is logistic regression?,” [Online]. Available: https://www.ibm.com/topics/logistic-regression.

[46] scikit learn, “SVM,” [Online]. Available: https://scikit-learn.org/stable/modules/svm.html.

[47] D. Leventis, “XGBoost Mathematics Explained,” [Online]. Available: https://dimleve.medium.com/xgboost-mathematics-explained-58262530904a.

[48] Simplilearn, “What is XGBoost? An Introduction to XGBoost Algorithm in Machine Learning,” [Online]. Available: https://www.simplilearn.com/what-is-xgboost-algorithm-in-machine-learning-article.

[49] IBM, “What are Naïve Bayes classifiers?,” [Online]. Available: https://www.ibm.com/topics/naive-bayes.

[50] Upgrad, “Naive Bayes Explained,” [Online]. Available: https://www.upgrad.com/blog/gaussian-naive-bayes/.

[51] Catboost, “Catboost,” [Online]. Available: https://catboost.ai/.

[52] M. Yawar, “CatBoost-ML,” [Online]. Available: https://www.codingninjas.com/studio/library/catboost-ml.

[53] Muthukrishnan, “Classification Report explained,” [Online]. Available: https://muthu.co/understanding-the-classification-report-in-sklearn/.

[54] National Institute on Alcohol Abuse and Alcoholism, “Alcohol Use Disorder,” [Online]. Available: https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/understanding-alcohol-use-disorder.

[55] National Institute of Mental Health, “Schizophrenia,” [Online]. Available: https://www.nimh.nih.gov/health/topics/schizophrenia#:~:text=What%20is%20schizophrenia%3F,for%20their%20family%20and%20friends.

[56] NHS, “Social Anxietyt Disorder,” [Online]. Available: https://www.nhs.uk/mental-health/conditions/social-anxiety/.

[57] GDPR, “GDPR,” [Online]. Available: https://gdpr-info.eu/.

[58] A. Oppermann, “What Is CatBoost?,” [Online]. Available: https://builtin.com/machine-learning/catboost.

[59] Google Developers, “Generalisation,” [Online]. Available: https://developers.google.com/machine-learning/crash-course/generalization/video-lecture#:~:text=Generalization%20refers%20to%20your%20model's,Time%3A%205%20minutes%20Learning%20Objectives.