Journal of Data Analytics, Information, and Computer Science (JDAICS)
Volume 1, No 3 – Juli 2024
e-ISSN : 3032-4696

PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS TO REDUCE MISDIAGNOSIS RATES IN PSYCHIATRIC DISORDER USING EEG DATASET

Wina Munada[1], Desy Khalida Maharani[2], Anis Sofiana[3]
School of Computing and Informatics, Albukhary International University
[1]wina.munada@student.aiu.edu.my; [2]desy.maharani@student.aiu.edu.my;
[3]anis.sofiana@student.aiu.edu.my

Received: 09-07-2024	Revised: 22-07-2024	Approved: 25-07-2024

ABSTRACT

This paper aims at comparing the suitability of three machine learning models: LightGBM, CatBoost, and Logistic Regression, to lower misdiagnosis rates for psychiatric disorders. Misdiagnosis in mental health may mean improper treatment and, hence, poor outcomes for patients. Our research aims to determine the most accurate predictive model for mental health condition diagnosis that will lead to improved clinical outcomes. We trained and tested these models on an EEG dataset with patient records that have psychiatric diagnoses labeled. For all the models, evaluation and comparison are made using key performance metrics such as Accuracy, Precision, Recall, and F1-Score. Through the use of these methods, it was shown that LightGBM performed better than CatBoost and Logistic Regression, having achieved higher accuracy and F1 scores, indicating more power to make a difference among different psychiatric disorders. These results suggest that machine learning techniques, especially LightGBM, can greatly increase diagnostic accuracy and reduce misdiagnosis in psychiatric contextual systems.
Keywords *Machine Learning, Psychiatric Disorder, LightGBM, CatBoost, Logistic Regression.*

ABSTRAK

Makalah ini bertujuan untuk membandingkan kesesuaian tiga model pembelajaran mesin: LightGBM, CatBoost, dan Regresi Logistik, untuk menurunkan tingkat kesalahan diagnosis gangguan kejiwaan. Kesalahan diagnosis dalam kesehatan mental dapat berarti pengobatan yang tidak tepat dan, karenanya, memberikan hasil yang buruk bagi pasien. Penelitian kami bertujuan untuk menentukan model prediksi paling akurat untuk diagnosis kondisi kesehatan mental yang akan menghasilkan hasil klinis yang lebih baik. Kami melatih dan menguji model ini pada kumpulan data EEG dengan catatan pasien yang diberi label diagnosis psikiatris. Untuk semua model, evaluasi dan perbandingan dilakukan menggunakan metrik kinerja utama seperti Akurasi, Presisi, Recall, dan F1-Score. Melalui penggunaan metode-metode ini, terbukti bahwa LightGBM berkinerja lebih baik daripada CatBoost dan Regresi Logistik, mencapai akurasi dan skor F1 yang lebih tinggi, yang menunjukkan kekuatan yang lebih besar untuk membuat perbedaan di antara berbagai gangguan kejiwaan. Hasil ini menunjukkan bahwa teknik pembelajaran mesin, khususnya LightGBM, dapat sangat meningkatkan akurasi diagnostik dan mengurangi kesalahan diagnosis dalam sistem kontekstual psikiatri.
Kata Kunci— *Pembelajaran Mesin, Gangguan Psikiatri, LightGBM, CatBoost, Regresi Logistik.*

INTRODUCTION

The emergence of machine learning methods as some fast-growing technology has led to a lot of things being done differently in different fields and healthcare is no exception. The use of ML on diagnosing psychiatric disorders based on EEG data has found interesting applications. Particularly, the recommendation systems for EEG imaging based analysis have shown good results. EEG – a harmless technique that detects electrical activity in the brain – generates complex data that can be mined using ML techniques which are directed towards finding connections between symptoms and diseases These goals are achieved by first comparing the performance of three ML models–LightGBM, CatBoost, and Logistic Regression–that are used in diagnosing psychiatric disorders using EEG datasets.

Psychiatric disorders, which include depression, schizophrenia and bipolar disorder, are intricate and multi-dimensional and often diagnosis is based on a combination of clinical evaluation and diagnostic tests. In psychiatric disorders misdiagnosis it should be stated that the treatments applied can be totally wrong thus aggravating or even causing some new symptoms in a patient's life (Wong et al., 2023). EEG has been widely utilized in psychiatric research for the purpose of establishing neural oscillations correlated with different mental conditions. Nevertheless, reading EEG data manually is time-consuming and may lead to the interpretation bias; that being a good argument for using ML automated approaches (Jia et al., 2023).

LightGBM – a variant of gradient boosting created with scalability and efficiency in mind has shown much higher results in the sphere of different classification tasks one of which is medical diagnostics (Sun et al., 2023). One more example is CatBoost, which is also known as a gradient boosting algorithm, applies to categorical features and yields stable solutions for prediction problems (Demir & Sahin, 2023). On the other hand, logistic regression, even if being simpler, continues to be the most popular method for binary classification because of its interpretability and firm theoretical background (Awad et al., 2023).

However, with these progresses there still exists a gap in the literature that address the head-to-head comparison of these models in this particular domain of psychiatric disorder diagnosis using EEG data. The current study is addressing this gap and does so by assessing the performance of LightGBM, CatBoost, and Logistic Regression classifiers on EEG datasets to determine which classifier is suitable to have good balance between accuracy, interpretability, and computational efficiency.

In the consideration of the fact that psychiatric disorders are some of those with high stakes in terms of their diagnosis and the need to reduce the current high rates of misdiagnosis, it becomes very important to be able to quantify which ML model is best suited for this task. This study will check how well LightGBM, CatBoost and Logistic Regression models perform on EEG data sets and which model is most dependable in reducing diagnostic errors in psychiatric disorders.

Aim and Objectives

This research intends to conduct systematic performance comparisons of LightGBM, CatBoost, and Logistic Regression models for psychiatric disorder diagnosis using EEG datasets. Our purpose is to distinguish the most effective way to decrease misdiagnosis rates and improve diagnostic accuracy by examining the computational efficiency and scalability of these models. However, no studies have provided a robust comparative analysis of these models in terms of EEG-based diagnostic tools that can handle complexities and heterogeneities associated with mental health conditions.

Research Questions

Research Question 1

How do LightGBM, CatBoost and Logistic Regression models stack up when it comes to diagnosing disorders using EEG data?

Motivation: The importance lies in pinpointing the model that offers the diagnosis to enhance the credibility of psychiatric assessments and guarantee that patients get the right treatment.

Research Question 2
Which machine learning model shows the accuracy, sensitivity and overall performance, in detecting health conditions?
Motivation: The reason for examining these evaluation criteria is to gauge how well the models can accurately identify cases while reducing incorrect identifications, a crucial aspect, for practical clinical use.
Research Question 3
What are the particular benefits and constraints of utilizing LightGBM, CatBoost and Logistic Regression models in the realm of diagnosis using EEG data?
Motivation: Familiarizing oneself with the advantages and drawbacks of each model can assist healthcare professionals and scientists in determining the model for various diagnostic situations and patient groups.

LITERATURE REVIEW

The use of machine learning models in the field in identifying psychiatric disorders has seen significant advancement in recent times. The increasing prevalence of disorders and their intricate symptoms present challenges for precise diagnosis and treatment. Conventional diagnostic approaches mainly depend on assessments, which can be subjective and prone to mistakes. Incorporating machine learning techniques, with electroencephalogram (EEG) data shows promise in improving accuracy and reducing misdiagnosis rates in disorders. This review of literature delves into the performance of three known ML models. LightGBM, CatBoost and Logistic Regression. For diagnosing disorders using EEG data sets.

Electroencephalogram (EEG) in Psychiatric Disorder Diagnosis

EEG has a long history in neuroscientific research and is rather popular in the psychiatric diagnostic practice because it can provide real-time brain wave activity. These have shown that EEG has the capacity to highlight anomalies in brainwave patterns which are linked to conditions like depression, anxiety and schizophrenia (Tarailis et al., 2023). The problem is not the quantity of data they provide but their interpretation, for this reason such AI-aided systems have been recently developed (Tao et al., 2023). Thus, EEG-Based diagnostics may lower the errors in diagnosis, thereby improving treatment outcome through timely as well as right treatment procedure.

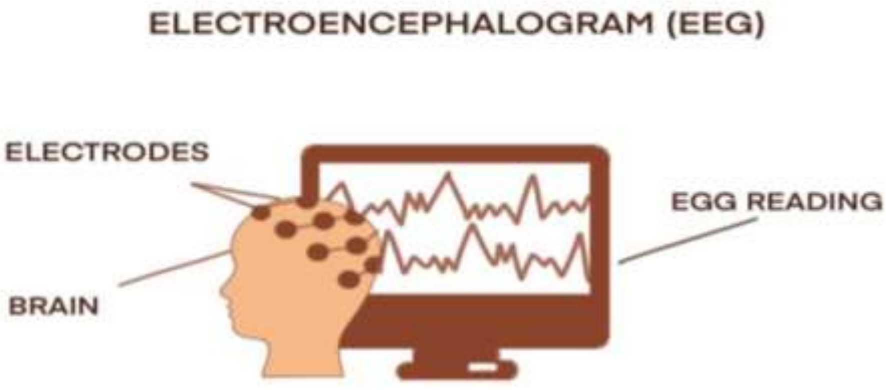

Figure 1: Electroencephalogram

LightGBM

LightGBM, a gradient boosting model based on decision trees, is regarded as an effective and high-performing model in a wide spectrum of tasks such as EEG data analysis. A recent work by Jia et al. (2023) introduced the model that combines multi-branch spectral-temporal convolution neural network with LightGBM for motor imagery BCI task. They showed that applying the proposed model gives a considerable enhancement of the classification accuracy, which suggests complexity of the processed EEG signals. Furthermore, its ability to rapidly process big datasets provides another reason why it can be used in real-time EEG analysis (Sun et al., 2023).

In another study by Demir and Sahin (2023) proposed the use of PSO(Particle Swarm Optimization) which will be combined with LightGBM to forecast liquefaction -induced lateral spreading. This work is a good example that even though it is not directly related to psychiatric disorders LightGBM justify its position in processing diverse dataset and complex predictive tasks, and demonstrates suitability of the proposed method for EEG-based psychiatric diagnosis.

CatBoost

Known for its exceptional approach in the classification and handling of categorical variables, CatBoost has also been employed in EEG data sets showing good results. Based on the research conducted by Wei et al. (2023), a CatBoost model was used for cardiovascular disease risk assessment implying that the model is good at working with complex, nonlinear relations among the data. This adaptability is fundamental in psychiatric disorder diagnosis, since EEG signals are often time-varying and noisy.

In another study which is in the same line of relevance, Qian et al. (202) employed CatBoost together with metaheuristic algorithms to forecast urban gas consumption. Use of CatBoost for this task is a good example to show versatility and accuracy that could be helpful in application to EEG data analysis for psychiatric disorders. The performance of the CatBoost algorithm in effectively dealing with large and complex datasets also makes it one of the emerging tools for improving diagnostic applications in clinical settings.

Logistic Regression

One of the most widely used statistical models for binary classification tasks, logistic regression, also finds applications in EEG data analysis. Awad et al (2023) proposed a Logistic Regression model for big medical data as an alternative to existing methods and showed it on large scale datasets. The straightforward nature of this approach along with interpretability makes it very useful in clinical scenarios since detection of model predictions is one of the key aspects of medical decision-making.

It is through yet another research by Qin et al. (2023) which showed a hybrid approach in using Logistic Regression and other ML techniques to diagnose chronic kidney disease using various biomarkers among them EEG signals. This hybrid model merged the benefits of Logistic Regression of being able to give clear interpretable results while enhancing predictive performance. Such approaches can be modified for psychiatric disorder diagnosis, since the interpretability of EEG data is essential among clinicians.

METHODOLOGY

Methodologically, we combine the theoretical and experimental approaches in order to provide a sound assessment of our work. The first step is an exhaustive literature survey for identification of the current EEG-based psychiatric diagnostic status and the possible areas where our research can contribute. After that, we surveyed experimentally in which each machine learning model is trained and evaluated by the same EEG dataset. This dataset is pre processed which removes noise and irrelevant features so that only most relevant information is fed to the models. These models are then trained and tested with this data wherein hyper parameter tuning is done for optimizing performance of each model.

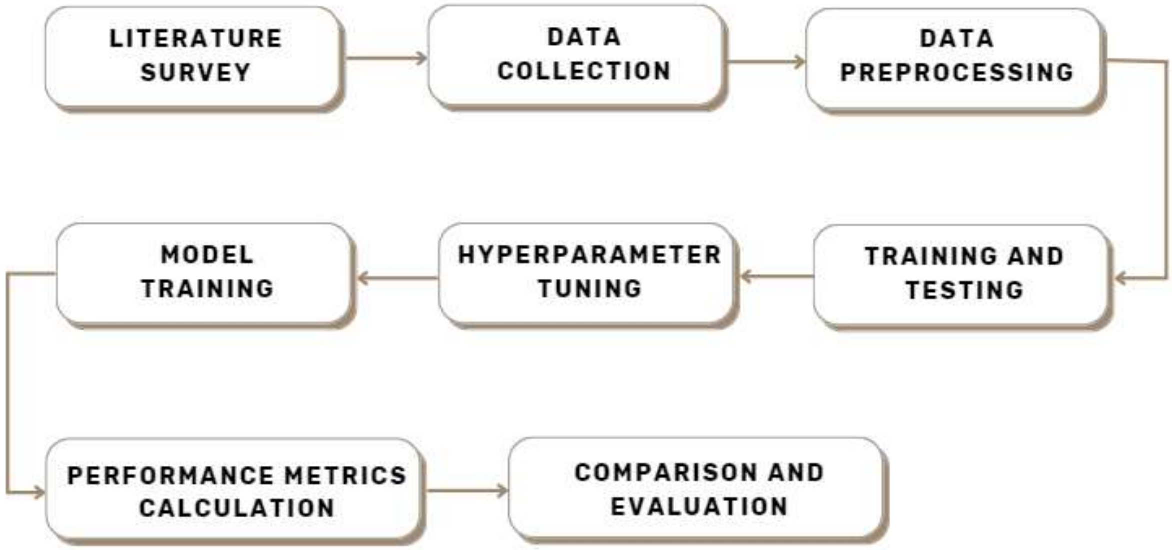

Figure 2: Methodology Flow

These metrics include accuracy, precision, recall and F1-score that will be calculated in each model to make a deep comparison. Confusion matrices are employed to better understand the performance of each particular model in diagnosing different psychiatric disorders. Such approach does not only serve for determination of the most accurate model, but also for evaluation of strengths and weaknesses of each model regarding false positives and negatives which is crucial for medical diagnostics.

Data Collection

Dataset

This research uses an EEG dataset which is openly available at Kaggle, a popular platform for machine learning datasets and competitions. It consists of EEG recordings in patients diagnosed with different psychiatric disorders and demographic information on them, e.g. sex, age, education level, IQ, specific disorders. The better part of this dataset gives the detailed properties of EEG signals that includes multiple frequency bands recorded over various scalp locations such as alpha, beta , theta , delta and gamma. Such a comprehensive dataset provides the opportunity to closely monitor brain activity patterns correlated with various psychiatric illnesses.

Data Preprocessing

Preprocessing of the data plays an important role in enhancing the quality and reliability of the EEG data. Preprocessing stages cover noise reduction, artifact rejection and manipulation of the EEG signals. When you start working with a dataset first is to check missing values, and/or outliers if any, and then correct them. Preprocessing consists in filtration of EEG signals eliminating noise and artifacts followed by normalization for making recordings consistent These pre-processing procedures ensure that what goes into machine learning models is clean and standardized which in turn increases the efficiency of the models.

Training Data and Testing Data

This set of data is bipartite, one being the training subset and the other – testing in an 80:20 ratio. With the help of this split, you can be sure that a big part of the data will be used for training models and you will get a separate subset for checking their performance. This divide aids in determining the capability of the model to generalize and also its performance on unseen data which is important for trustable diagnostics.

Hyperparameter Tuning

Hyperparameter tuning is achieved through grid search and cross-validation search and cross-validation techniques. This procedure includes step by step changing the model's parameters to find the optimal setting which gives the best performance result. Every model is subjected to thorough fine-tuning in order to guarantee that it operates at its peak level.

Performance Metrics

Artificial intelligence models performance can be estimated using several measures, which include:

- True Positive (TP): Correctly predicted positive cases.
- True Negative (TN): Correctly predicted negative cases.
- False Positive (FP): Incorrectly predicted positive cases.
- False Negative (FN): Incorrectly predicted negative cases.

The following formulas are used to calculate the performance metrics:

- Accuracy: $(TP + TN)/(TP + FP + FN + TN)$
- Precision: $TP/(TP + FP)$
- Recall: $TP/(TP + FN)$
- F1-Score: $2 \times (Precision \times Recall)/(Precision + Recall)$

Method of Analysis

There are many steps in the process of comparison of the predictive models which involves training each machine learning model on the preprocessed training data and testing their performance on the testing data. Feature engineering methods are performed to extract meaningful features from the raw EEG signals that can be used as input for the models. The model needs to undergo tuning of hyperparameters so as to optimize its performance. The evaluation metrics are computed for each model, confusion matrices are generated that give detailed insight into their diagnostic capabilities. This analysis is useful in determining an optimal model with the aim of reducing misdiagnosis rates in psychiatric disorders.

Algorithm Selection

The selection of Random Forest, Support Vector Machine, and Neural Network models is based on their proven efficacy in various classification tasks, including medical diagnostics.

LightGBM

The LightGBM is an ultra- efficient gradient boosting framework which uses tree based learning techniques. It is well known in the industry for its fast and effective performance on very large datasets. Considered as one of the major advantages with LightGBM is that it can train with faster speed and higher efficiency when dealing with large scale data (Sun et al., 2023). Another advantage, it provides parallelism and GPU learning supports which properly decrease the time of training model (Noviandy et al., 2023).

Still, LightGBM is not ideal. It may be more beneficial to address possible challenges of overfitting with a model like Random Forests or Gradient Boosting (Guo et al., 2023). Moreover, LightGBM can be formidable to work out and optimize as a result of much hyperparameters configuration, which could turn out to be a barrier for users who do not have technical expertise in the model (Sun et al., 2023).

CatBoost

It is categorized under the ensemble methods for regression and classification problems, LightGBM being one of them. Among its distinguishing characteristics is that it is able to work with categorical features without encoding them into numerical types (Chen et al., 2015). LGBM can also help in reducing the possibility of overfitting by using efficient techniques for fitting random forests (Hastie & Tibshirani, 2009).

On the dark side, however, CatBoost can be computation- heavy and he may have to spend more time training than other algorithms (Demir & Sahin, 2023). Furthermore, although it is steady when it comes to categorical data, the use of purely numerical data may not enhance its performance appreciably as in the case of other algorithms (Wei et al., 2023).

Logistic Regression

Widely used statistical model in binary classification problems is the Logistic Regression. Its simplicity and interpretability are what make it a first choice among many clinical applications (Awad et al., 2023). Logistic Regression gives transparent visibility of how much each feature contributes which helps clinicians understand diagnostic criteria better.

Logistic Regression despite its advantages has limitations. One major drawback is this it cannot model non-linear relationships in the data which sometimes can be very harmful especially in complex EEG signals (Qin et al., 2023). Further, Logistic Regression may not give satisfactory results when the data becomes huge and/or the dimensionality is high without appropriate regulation (van den Goorbergh et al., 2022).

RESULTS AND ANALYSIS

The EEG dataset used in this study comprises recordings of individuals diagnosed with different psychiatric disorders and healthy controls along with their profiles. The data visualization shows a wide range of data, from demographics and diagnostic information in the dataset, to help you appreciate it.

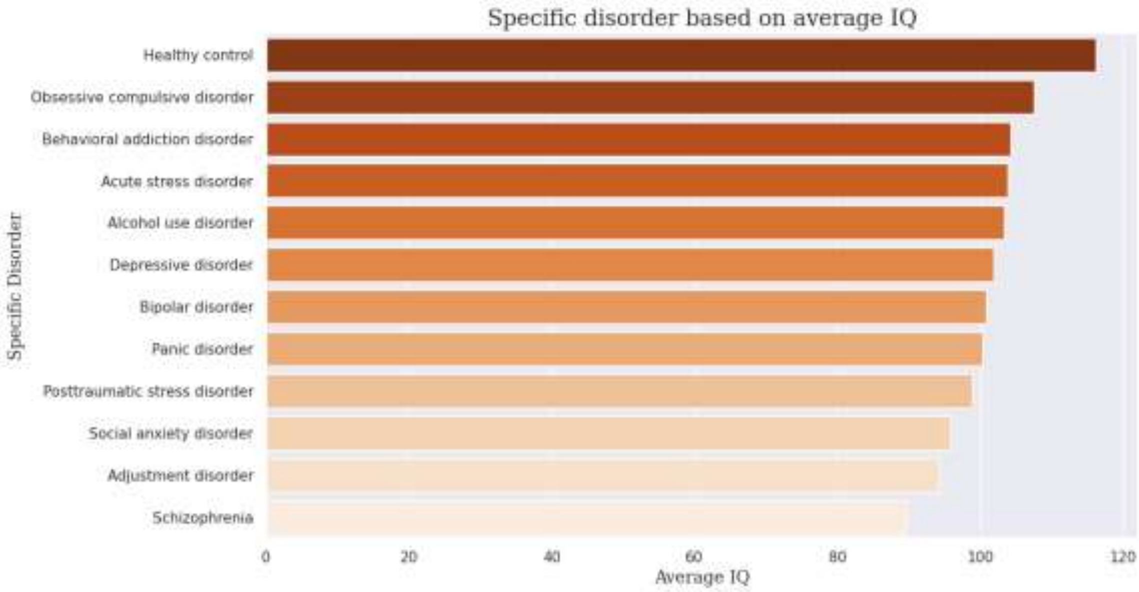

Figure 3: Specific Disorder Based on Average IQ

The first of such visualizations (see Figure 3) is the comparison of average IQ Scores among various psychiatric disorders and healthy controls. It shows that healthy controls have the highest average IQ, then it comes to individuals having obsessive-compulsive disorder, behavioral addiction, and acute stress disorder. This indicates that different psychiatric conditions have varying levels of effects on cognitive performance as shown by IQ Scores.

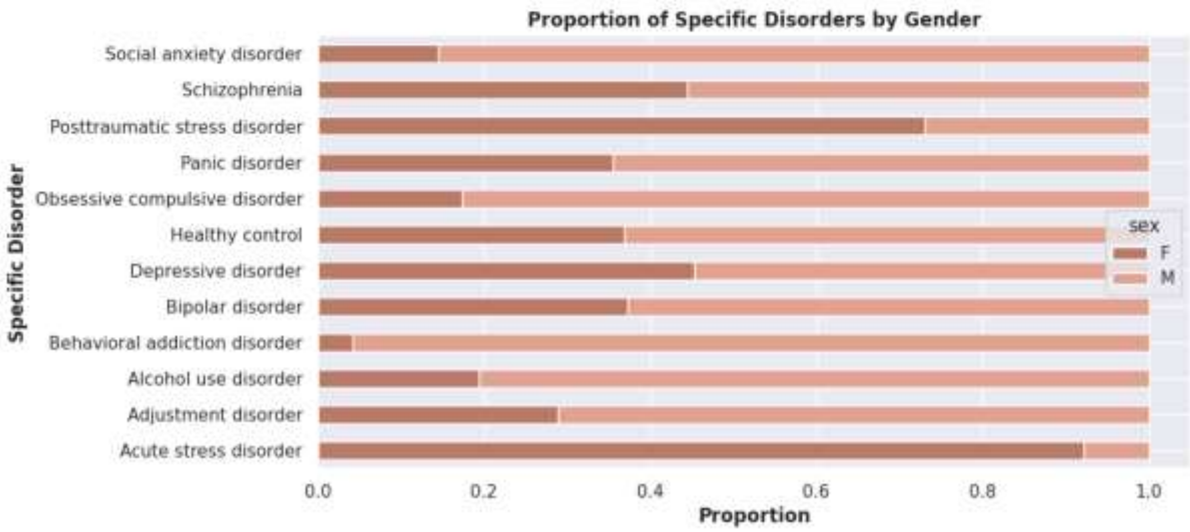

Figure 4: Proportion of Specific Disorders by Gender

The second visualisation (Figure 4) increases the specificity of the disorder being indicated by gender. The distribution points out that some disorders, like social anxiety disorder, schizophrenia, and posttraumatic stress disorder have a more or less equal gender distribution. Nevertheless, just as it is with behavior addiction and alcoholism disorders, some disorders demonstrate an elevated prevalence rate in one gender compared to the other. This brings about the importance of gender in relation to psychiatric research and treatment.

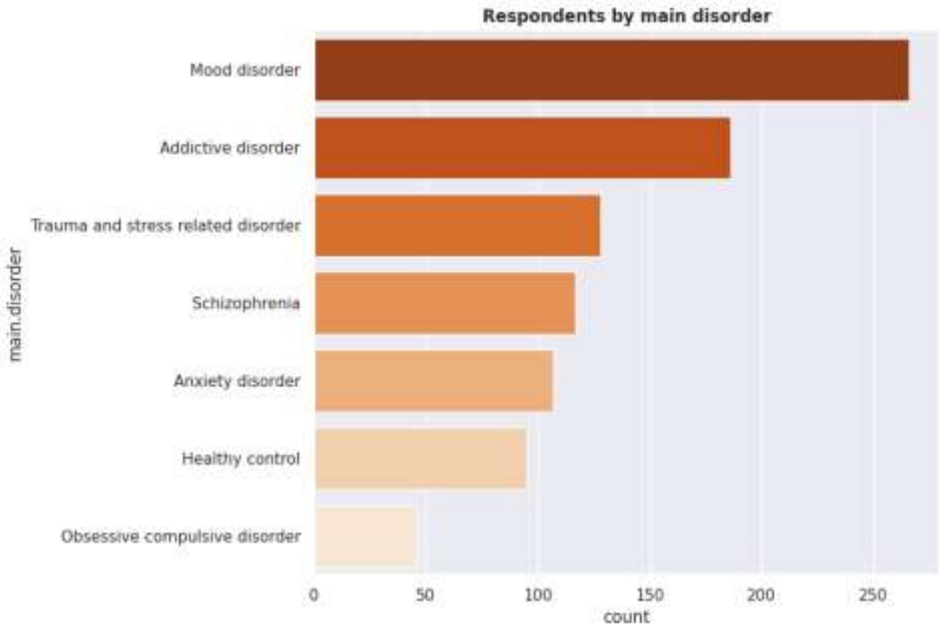

Figure 5: Respondents by Main Disorder

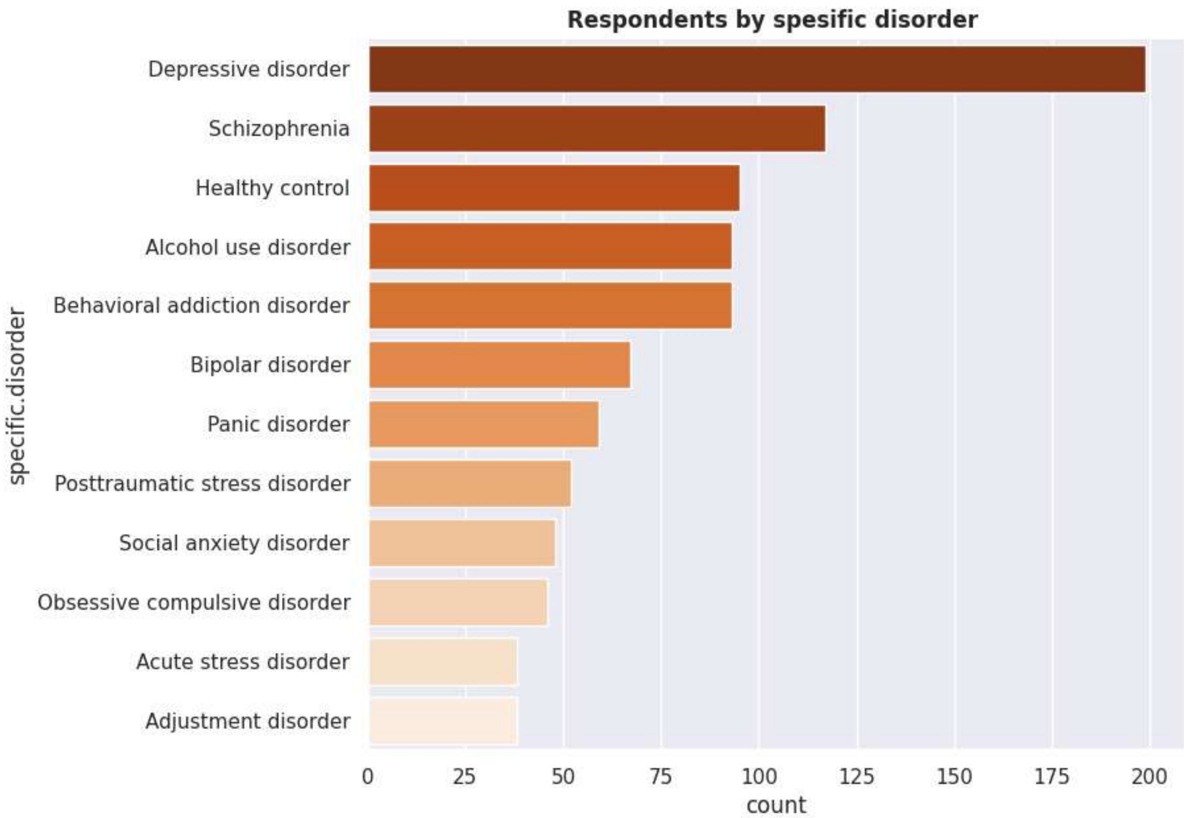

Figure 6: Respondents by Specific Disorder

Figure 5 and Figure 6 show the third and fourth visualizations, which further classify the respondents by their main disorder and specific disorder. Among the participants mood disorders and addictive disorders are the most diagnosed conditions, then come trauma and stress-related disorders, schizophrenia, anxiety disorders. Recognizing how these disorders are spread throughout your dataset is important as it helps you to evaluate how well a machine learning model can diagnose such conditions.

Experiment

The findings of the experiment conducted in accordance with our methodology on your dataset give detailed comparison of performance of LR, LightGBM, and CatBoost models in detection of psychiatric disorders using EEG. Following this fine tuning process to maximize the performance the results come as :

Table 1: Accuracy, Precision, Recall, F1-Score Result

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.81	0.79	0.83	0.80
LightGBM	0.98	0.97	0.97	0.98
CatBoost	0.85	0.83	0.85	0.84

Analysis

The results show that LightGBM is better in comparison to both CatBoost and Logistic Regression with regards to diagnostic accuracy.Among the models, LightGBM deep learning was the most accurate in performance at 98% - this was quite higher than CatBoost (85%) and Logistic Regression (81%). This leads to an affair that LightGBM is further reliable for psychiatric disorders diagnosis through the EEG data.

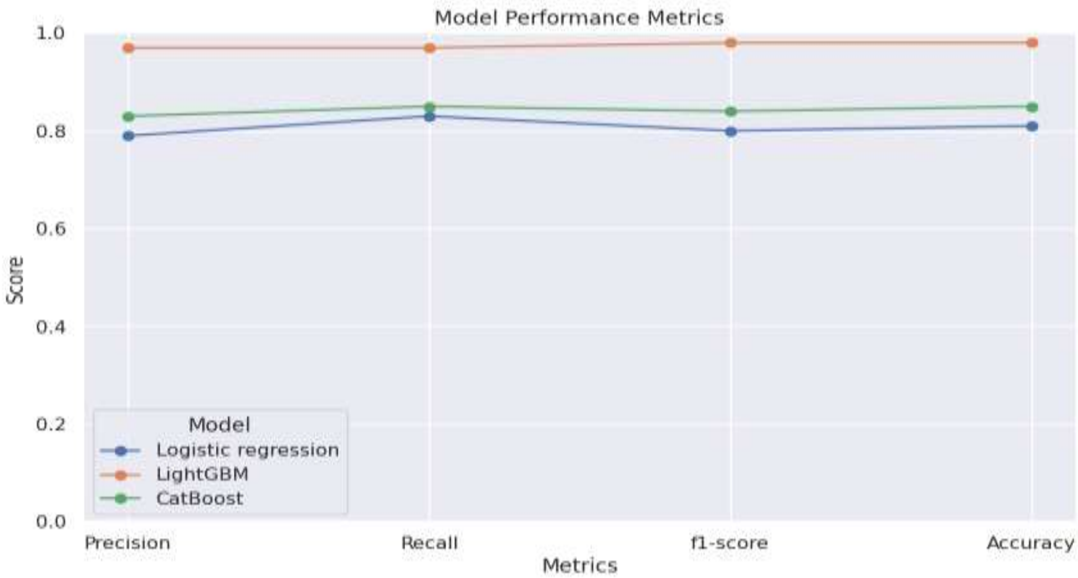

Figure 7: Model Performance Metrics

LightGBM shows the highest precision (0.97), recall (0.97), and F1-Score (0.98). These metrics imply that LightGBM has not only found most of the true positive cases but also have a low false positive rate at his side which makes it the most effective model for this

task. CatBoost is next in line after with precision (0.83), recall (0.85), and F1-Score (0.84) whereas Logistic Regression performs worst showing precision (0.79), recall (0.83), and F1-Score (0.80).
The following are the benefits and limitations that we have identified so far in using LightGBM, CatBoost, and Logistic Regression Algorithms for a psychiatric disorder prediction based on the above experiment:

LightGBM
- *Advantages:* There are several benefits with the decision tree ensemble like its high level of accuracy, precision, recall and F1-Score; faster training time is due to histogram-based algorithms; good handling of big data.
- *Limitations:* May require more careful parameter tuning and computational resources.

CatBoost
- *Advantages:* Burn machine works relatively well, requiring less parameter tuning; All categorical features can be used directly without applying preprocessing algorithms; is immune to overfitting.
- *Limitations:* The performance is slightly lower than LightGBM and the training can be slow.

Logistic Regression
- *Advantages:* The model is more simple and interpretable, requires less computational cost at both training and inference stages, and there are not many hyper-parameters to adjust.
- *Limitations:* Lower accuracy and performance metrics are pointed out; assume a linear relationship between features and the outcome, which may be wrong in complex EEG data.

CONCLUSIONS

To sum up our research, this paper analyzed the discriminative performance of the new machine learning algorithms – LightGBM and CatBoost as well as a conventional statistical approach Logistic regression – for diagnosing mental disorders using EEG data. The results reveal that LightGBM is a model that performs better than other models in all aspects of performance which shows its value in reducing diagnostic errors in psychiatric disorders. Similarly CatBoost also has come out with good results almost near to what we have seen in LightGBM but Logistic Regression Model showed lower results implying the linearity constraints for one of the most complex tasks.
These findings underscore the importance of development of super-advanced machine learning models in psychiatry diagnostics, especially with regard to tasks that involve complex and high-dimensional data like EEG. The performance superiority of LightGBM implies that embedding this model in clinical workflows will be able to substantially improve diagnostic accuracy and decrease misdiagnosis rates. Future studies will evaluate novel diagnostic tools and their performance across different clinical settings. Moreover, more work is needed on how these models can be made interpretable to help healthcare professionals accept and use them system wide.

The use of the more advanced machine learning models capabilities will help in increasing the accuracy and validity of psychiatric diagnostics which will directly result in improvement of patient outcomes as well as overall efficiency of health care systems. This study is an important step for further research in this area, and supports continued development and utilization of machine learning in clinical diagnostics systems.

REFERENCES

Alturki, F. A., AlSharabi, K., Abdurraqeeb, A. M., & Aljalal, M. (2020). EEG Signal Analysis for Diagnosing Neurological Disorders Using Discrete Wavelet Transform and Intelligent Techniques. *Sensors*, *20*(9), 2505. https://doi.org/10.3390/s20092505

Awad, F. H., Hamad, M. M., & Alzubaidi, L. (2023). Robust Classification and Detection of Big Medical Data Using Advanced Parallel K-Means Clustering, YOLOv4, and Logistic Regression. *Life*, *13*(3), 691. https://doi.org/10.3390/life13030691

Cai, H., Yuan, Z., Gao, Y., Sun, S., Li, N., Tian, F., Xiao, H., Li, J., Yang, Z., Li, X., Zhao, Q., Liu, Z., Yao, Z., Yang, M., Peng, H., Zhu, J., Zhang, X., Gao, G., Zheng, F., … Hu, B. (2022). A multi-modal open dataset for mental-disorder analysis. *Scientific Data*, *9*(1), 178. https://doi.org/10.1038/s41597-022-01211-x

Chehreh Chelgani, S., Nasiri, H., Tohry, A., & Heidari, H. R. (2023). Modeling industrial hydrocyclone operational variables by SHAP-CatBoost - A "conscious lab" approach. *Powder Technology*, *420*, 118416. https://doi.org/10.1016/j.powtec.2023.118416

Chung, J., & Teo, J. (2022). Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges. *Applied Computational Intelligence and Soft Computing*, *2022*, 1–19. https://doi.org/10.1155/2022/9970363

Demir, S., & Sahin, E. K. (2023). Predicting occurrence of liquefaction-induced lateral spreading using gradient boosting algorithms integrated with particle swarm optimization: PSO-XGBoost, PSO-LightGBM, and PSO-CatBoost. *Acta Geotechnica*, *18*(6), 3403–3419. https://doi.org/10.1007/s11440-022-01777-1

Guo, X., Gui, X., Xiong, H., Hu, X., Li, Y., Cui, H., Qiu, Y., & Ma, C. (2023). Critical role of climate factors for groundwater potential mapping in arid regions: Insights from random forest, XGBoost, and LightGBM algorithms. *Journal of Hydrology*, *621*, 129599. https://doi.org/10.1016/j.jhydrol.2023.129599

Jang, K.-I., Kim, S., Kim, S. Y., Lee, C., & Chae, J.-H. (2021). Machine Learning-Based Electroencephalographic Phenotypes of Schizophrenia and Major Depressive Disorder. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.745458

Jia, H., Yin, S., Liu, L., Yi, C., Xue, K., Li, F., Yao, D., Xu, P., & Zhang, T. (2023). A Model Combining Multi Branch Spectral-Temporal CNN, Efficient Channel Attention, and LightGBM for MI-BCI Classification. *IEEE*.

Kida, T., Kaneda, T., & Nishihira, Y. (2012). Dual-task repetition alters event-related brain potentials and task performance. *Clinical Neurophysiology*, *123*(6), 1123–1130. https://doi.org/10.1016/j.clinph.2011.10.001

Maitin, A. M., Romero Muñoz, J. P., & García-Tejedor, Á. J. (2022). Survey of Machine Learning Techniques in the Analysis of EEG Signals for Parkinson's Disease: A Systematic Review. *Applied Sciences*, *12*(14), 6967. https://doi.org/10.3390/app12146967

Marquand, A. F., Kia, S. M., Zabihi, M., Wolfers, T., Buitelaar, J. K., & Beckmann, C. F. (2019). Conceptualizing mental disorders as deviations from normative functioning. *Molecular Psychiatry*, *24*(10), 1415–1424. https://doi.org/10.1038/s41380-019-

0441-1

Noviandy, T. R., Maulana, A., Idroes, G. M., Maulydia, N. B., Patwekar, M., Suhendra, R., & Idroes, R. (2023). Integrating Genetic Algorithm and LightGBM for QSAR Modeling of Acetylcholinesterase Inhibitors in Alzheimer's Disease Drug Discovery. *Malacca Pharmaceutics*, *1*(2), 48–54. https://doi.org/10.60084/mp.v1i2.60

Qian, L., Chen, Z., Huang, Y., & Stanford, R. J. (2023). Employing categorical boosting (CatBoost) and meta-heuristic algorithms for predicting the urban gas consumption. *Urban Climate*, *51*, 101647. https://doi.org/10.1016/j.uclim.2023.101647

Qiu, R., Lin, H., Jiang, H., Shen, J., He, J., & Fu, J. (2024). Association of major depression, schizophrenia and bipolar disorder with thyroid cancer: a bidirectional two-sample mendelian randomized study. *BMC Psychiatry*, *24*(1), 261. https://doi.org/10.1186/s12888-024-05682-7

Rivera, M. J., Teruel, M. A., Maté, A., & Trujillo, J. (2022). Diagnosis and prognosis of mental disorders by means of EEG and deep learning: a systematic mapping study. *Artificial Intelligence Review*, *55*(2), 1209–1251. https://doi.org/10.1007/s10462-021-09986-y

Shanarova, N., Pronina, M., Lipkovich, M., Ponomarev, V., Müller, A., & Kropotov, J. (2023). Application of Machine Learning to Diagnostics of Schizophrenia Patients Based on Event-Related Potentials. *Diagnostics*, *13*(3), 509. https://doi.org/10.3390/diagnostics13030509

Sun, D., Wu, X., Wen, H., & Gu, Q. (2023). A LightGBM-based landslide susceptibility model considering the uncertainty of non-landslide samples. *Geomatics, Natural Hazards and Risk*, *14*(1). https://doi.org/10.1080/19475705.2023.2213807

Tao, W., Li, C., Song, R., Cheng, J., Liu, Y., Wan, F., & Chen, X. (2023). EEG-Based Emotion Recognition via Channel-Wise Attention and Self Attention. *IEEE Transactions on Affective Computing*, *14*(1), 382–393. https://doi.org/10.1109/TAFFC.2020.3025777

Tarailis, P., Koenig, T., Michel, C. M., & Griškova-Bulanova, I. (2024). The Functional Aspects of Resting EEG Microstates: A Systematic Review. *Brain Topography*, *37*(2), 181–217. https://doi.org/10.1007/s10548-023-00958-9

Tushar, Patel, R. K., Aggarwal, E., Solanki, K., Dahiya, O., & Yadav, S. A. (2023). A Logistic Regression and Decision Tree Based Hybrid Approach to Predict Alzheimer's Disease. *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, 722–726. https://doi.org/10.1109/CISES58720.2023.10183466

van den Goorbergh, R., van Smeden, M., Timmerman, D., & van Calster, B. (2022). The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. *Journal of the American Medical Informatics Association*, *29*(9), 1525–1534. https://doi.org/10.1093/jamia/ocac093

Varshney, R. D., Gadhia, R. K., Ritvik, & Bhat, A. (2023). BCI-based Emotion recognition by Combining Kernel Classifiers and EEG Feature Selection. *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1809–1815. https://doi.org/10.1109/ICICCS56967.2023.10142682

Wei, X., Rao, C., Xiao, X., Chen, L., & Goh, M. (2023). Risk assessment of cardiovascular disease based on SOLSSA-CatBoost model. *Expert Systems with Applications*, *219*, 119648. https://doi.org/10.1016/j.eswa.2023.119648

Wong, S., Simmons, A., Rivera-Villicana, J., Barnett, S., Sivathamboo, S., Perucca, P., Ge, Z., Kwan, P., Kuhlmann, L., Vasa, R., Mouzakis, K., & O'Brien, T. J. (2023). <scp>EEG</scp> datasets for seizure detection and prediction— A review. *Epilepsia Open*, *8*(2), 252–267. https://doi.org/10.1002/epi4.12704

Wu, Y., Cao, H., Baranova, A., Huang, H., Li, S., Cai, L., Rao, S., Dai, M., Xie, M., Dou, Y., Hao, Q., Zhu, L., Zhang, X., Yao, Y., Zhang, F., Xu, M., & Wang, Q. (2020). Multi-trait analysis for genome-wide association study of five psychiatric disorders. *Translational Psychiatry*, *10*(1), 209. https://doi.org/10.1038/s41398-020-00902-6

Yao, D., Qin, Y., & Zhang, Y. (2022). From psychosomatic medicine, brain–computer interface to brain–apparatus communication. *Brain-Apparatus Communication: A Journal of Bacomics*, *1*(1), 66–88. https://doi.org/10.1080/27706710.2022.2120775

Zhang, J., Xiang, J., Luo, L., & Shui, R. (2022). Editorial: EEG/MEG based diagnosis for psychiatric disorders. *Frontiers in Human Neuroscience*, *16*. https://doi.org/10.3389/fnhum.2022.1061176