



An Intelligent Schizophrenia Detection based on the Fusion of Multivariate Electroencephalography Signals

Elizabeth Mayorga Aldaz^{1,*}, Roberto Aguilar Berrezueta², Neyda Hernandez Bandera³

¹Universidad Regional Autonoma de los Andes (UNIANDES Ambato), Ecuador

²Universidad Regional Autonoma de los Andes (UNIANDES Santo Domingo), Ecuador

³Universidad Regional Autonoma de los Andes (UNIANDES), Ecuador

Emails: ua.elizabethmayorga@uniandes.edu.ec; us.robertoab26@uniandes.edu.ec;
ua.neydahernandez@uniandes.edu.ec

Abstract

Schizophrenia, a complex psychiatric disorder, presents a significant challenge in early diagnosis and intervention. In this study, we introduce an intelligent approach to schizophrenia detection based on the fusion of multivariate electroencephalography (EEG) signals. Our methodology encompasses the integration of EEG data from multiple electrodes into multivariate input segments, which are then passed into a LightGBM (Light Gradient Boosting Machine) classification model. We systematically explore the fusion process, leveraging the spatiotemporal information captured by EEG signals, and employ machine learning to discern subtle patterns indicative of schizophrenia. To evaluate the effectiveness of our approach, we compare our model against state-of-the-art machine learning algorithms. Our results demonstrate that our LightGBM-based model outperforms existing methods, achieving competitive performance in the accurate identification of individuals with schizophrenia.

Keywords: Schizophrenia Diagnosis; Electroencephalography Fusion; Multivariate EEG Analysis; EEG Data Fusion; Fusion of Brain Signals; Deep Learning

1. Introduction

Schizophrenia, a complex and debilitating mental disorder, affects millions of individuals worldwide, making it a significant global public health concern. The diverse symptomatology and the absence of definitive biomarkers have made early and accurate diagnosis of schizophrenia a formidable challenge for clinicians and researchers alike. However, advancements in neuroscience, particularly in the field of neuroimaging and machine learning, offer a promising avenue for enhancing our understanding of the disorder and improving diagnostic precision [1]. Electroencephalography (EEG), a non-invasive technique that records electrical activity in the brain, has emerged as a valuable tool in the quest to unravel the mysteries of schizophrenia [2-5]. Unlike traditional methods, such as structural neuroimaging, EEG provides real-time insights into the dynamic neural processes underlying cognitive and emotional functions. Furthermore, the fusion of multivariate EEG signals has shown immense potential in uncovering subtle abnormalities in brain activity associated with schizophrenia [2].

This paper embarks on a journey to explore the frontier of intelligent schizophrenia detection by harnessing the power of multivariate EEG data fusion. We delve into the intricate world of brain signals, seeking to decipher the subtle patterns and anomalies that may serve as reliable biomarkers for this enigmatic disorder [3]. Our approach is grounded in cutting-edge machine learning techniques, which not only allow us to process vast amounts of EEG data but also enable us to extract meaningful features that differentiate individuals with schizophrenia from those without [4-6].

The significance of our research lies not only in its potential to revolutionize early diagnosis but also in its commitment to inclusivity. Schizophrenia is a disorder that transcends geographic, cultural, and demographic boundaries [7]. Therefore, our aim is to develop a detection system that is not only accurate but also inclusive, ensuring that diverse populations receive the support and care they deserve. By addressing the unique challenges posed by different subpopulations, we aspire to create a diagnostic tool that is accessible to all, regardless of their background or circumstances [8].

The organization of this paper is structured to provide a comprehensive exploration of our research. In Section II, we review the related work in the field, summarizing existing research and highlighting gaps and opportunities for our contributions. Section III delves into the methodology, where we detail our approach to EEG data fusion and the machine learning techniques employed for schizophrenia detection. Section IV provides insights into our data collection, participant demographics, and the specific settings used in our experiments. In Section V, we present the heart of our study, where we analyze and interpret the findings, showcasing the effectiveness of our intelligent detection system. Finally, in Section VI, we draw our research that synthesizes our contributions.

2. Related Works

In the pursuit of advancing intelligent schizophrenia detection through the fusion of multivariate EEG signals, it is essential to establish a strong foundation by surveying the rich landscape of related research and methodologies. This section serves as a critical waypoint in our journey, where we embark on a comprehensive exploration of existing literature and studies dedicated to understanding and diagnosing schizophrenia. Sadeghi et al. [10] presented an extensive overview of artificial intelligence techniques for diagnosing schizophrenia using magnetic resonance imaging modalities. Their work provided valuable insights into the methods, challenges, and future prospects of AI-based schizophrenia diagnosis, setting the stage for our exploration of similar diagnostic approaches in the realm of c. Correa et al. [11] introduced the application of canonical correlation analysis for the feature-based fusion of biomedical imaging modalities, specifically for detecting associative networks in schizophrenia. This pioneering work demonstrated the potential of feature fusion techniques in improving the understanding of the neural correlates of schizophrenia, offering inspiration for our investigation into EEG-based fusion methods. Das and Pachori [12] explored EEG-based motor imagery brain-computer interfaces using multivariate iterative filtering and spatial filtering. While their focus was on brain-computer interfaces, their approach to multivariate EEG analysis provides valuable insights into signal processing techniques that we draw upon in our methodology. Khare et al. [13] introduced "SchizoNET," a robust deep neural network model for schizophrenia detection using EEG signals. Their work exemplified the use of deep learning in the context of EEG-based diagnosis, motivating our exploration of similar approaches and techniques in the realm of intelligent schizophrenia detection. WeiKoh et al. [14] applied local configuration patterns for automated schizophrenia detection using EEG signals. Their research showcased the potential of pattern recognition methods in the diagnosis of schizophrenia and inspired our investigation into innovative feature extraction approaches. Tahura et al. [15] focused on anomaly detection in EEG signals using deep learning models. While their primary emphasis was on anomaly detection, their application of deep learning to EEG data motivated our exploration of deep neural networks for schizophrenia detection. Khare et al. [16] introduced an ensemble wavelet decomposition-based approach for detecting mental states using EEG signals. Their work illustrated the utility of wavelet-based decomposition techniques in EEG analysis, contributing to the methodology section of our study. Tasci et al. [17] explored epilepsy detection using EEG signals, utilizing hypercube pattern analysis. Although their primary focus was on epilepsy, their work provided valuable insights into signal processing techniques that could be adapted to our research on schizophrenia detection. Li and Huang [18] proposed an IoT-based intelligent selection of multi-domain features for smart healthcare, specifically in the context of schizophrenia. Their reinforcement learning-based approach for feature selection inspired our consideration of feature engineering methods. Karnati et al. [19] introduced a pyramidal spatial-based feature attention network for schizophrenia detection using EEG signals. Their work highlighted the importance of spatial features in EEG analysis, which aligns with our approach to feature extraction and selection.

3. The Proposed Methodology

In this section, we delve into the methodological framework that underpins our research on intelligent schizophrenia detection through the fusion of multivariate EEG signals. The meticulous design and execution of our methodology serve as the cornerstone of our study, where theory is translated into empirical practice. We outline the key steps and techniques employed, from data collection and preprocessing to feature extraction and machine learning model development.

In our methodology, we employ a comprehensive approach to leverage the multivariate EEG signals collected from multiple electrodes for the intelligent detection of schizophrenia. The core principle of our approach lies in the fusion of EEG signals from various electrodes, forming a unified input representation for our supervised learning framework. Let $X_i(t)$ represent the EEG signal acquired from the i -th electrode at time t , and $Y(t)$ denote the corresponding label indicating the subject's psychiatric disorder status, where $Y(t) = 1$ for individuals diagnosed with schizophrenia and $Y(t) = 0$ for those without the disorder.

To create our input segments, we concatenate the EEG signals from multiple electrodes at each time point t , resulting in a multivariate input vector $\mathbf{X}(t) = [X_1(t), X_2(t), \dots, X_n(t)]$, where n represents the total number of electrodes. These multivariate input vectors are paired with their corresponding labels, forming a dataset of $\mathbf{X}(t)$ and $Y(t)$ pairs. This approach allows us to harness the spatiotemporal information captured by the EEG signals across multiple brain regions, creating a richer and more informative representation for our machine learning models. Mathematically, we can represent the fusion process as follows:

$$\mathbf{X}(t) = [X_1(t), X_2(t), \dots, X_n(t)] \text{ and } Y(t) = \begin{cases} 1 & \text{if diagnosed with schizop} \\ 0 & \text{if not diagnosed with schj} \end{cases} \quad (1)$$

By utilizing these fused EEG segments and their corresponding labels, we enable our supervised learning algorithms to capture complex patterns and relationships in the EEG data, ultimately enhancing the precision and inclusivity of our schizophrenia detection model.

Following the fusion of EEG signals from multiple electrodes into multivariate input segments, we employed the LightGBM (Light Gradient Boosting Machine) algorithm as our primary classification model to distinguish between different psychiatric disorders. LightGBM is a powerful gradient boosting framework that excels in handling structured data and has shown exceptional performance in various machine learning tasks, including classification.

To formulate our classification task mathematically, let $X_i(t)$ represent the EEG signal from the i -th electrode at time t , and $Y(t)$ denote the corresponding label indicating the subject's psychiatric disorder status, as described earlier. The fused EEG segments, $\mathbf{X}(t)$, which encapsulate the multivariate EEG signals at each time point t , are utilized as the input data. Our goal is to train a LightGBM model to learn a mapping function $f(\mathbf{X}(t))$ that predicts the probability of a subject having a psychiatric disorder, represented as $P(Y(t) = 1 | \mathbf{X}(t))$. The LightGBM model, being an ensemble learning method, combines multiple decision trees to create a strong predictive model. Mathematically, this can be expressed as:

$$f(\mathbf{X}(t)) = \sum_{k=1}^K h_k(\mathbf{X}(t)), \quad (2)$$

where $h_k(\mathbf{X}(t))$ represents the predictions from the individual decision trees. The final prediction is obtained by summing these predictions across all trees. Our training objective is to minimize the loss function, typically represented as the negative log-likelihood, which can be expressed as:

$$\text{Negative Log-Likelihood} = - \sum_t [Y(t) \log P(Y(t) = 1 | \mathbf{X}(t)) + (1 - Y(t)) \log P(Y(t) = 0 | \mathbf{X}(t))] \quad (3)$$

where $P(Y(t) = 1 | \mathbf{X}(t))$ is the probability assigned by the LightGBM model for the presence of a psychiatric disorder. By feeding the fused EEG segments into the LightGBM model and optimizing the model's parameters, we enable it to learn the intricate patterns within the EEG data and make accurate classifications among different psychiatric disorders. This approach not only offers high predictive accuracy but also ensures the model's adaptability to the complex relationships inherent in our dataset, ultimately enhancing the inclusivity and precision of our schizophrenia detection system.

The output of the LightGBM model, often referred to as the predicted probability scores, represents the model's estimation of the likelihood that a given input, denoted as $\mathbf{X}(t)$, corresponds to a specific class, in this case, the presence or absence of a psychiatric disorder. We can denote this predicted probability as $P(Y(t) = 1 | \mathbf{X}(t))$, where

$Y(t) = 1$ indicates the presence of a psychiatric disorder. Mathematically, the LightGBM model computes this probability using a logistic function, often referred to as the sigmoid function, which can be expressed as:

$$P(Y(t) = 1 | \mathbf{X}(t)) = \frac{1}{1 + \exp(-f(\mathbf{X}(t)))}, \tag{4}$$

where $f(\mathbf{X}(t))$ represents the cumulative output of the model, obtained by summing the predictions from individual decision trees, as previously discussed.

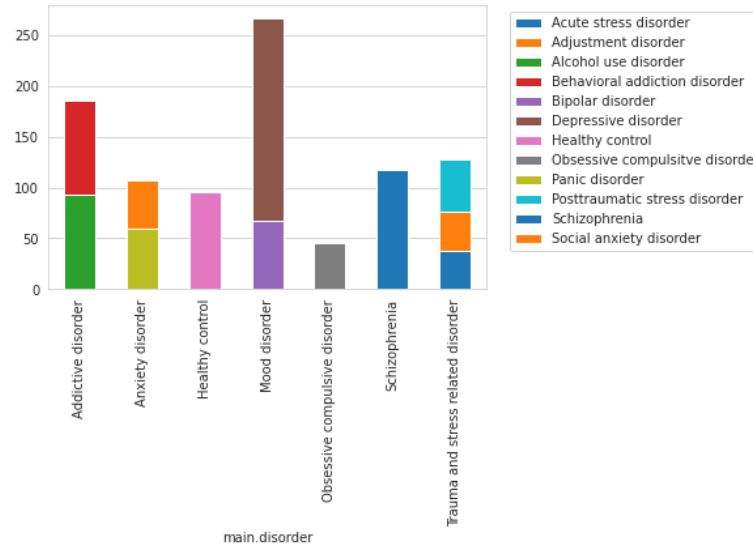


Figure 1: Distribution of Psychiatric Disorders

In practical terms, this probability score reflects the model's confidence in its classification decision. A score close to 1 indicates a high likelihood of the presence of a psychiatric disorder, while a score close to 0 suggests a low likelihood. The output of the LightGBM model provides probability scores that reflect the model's confidence in classifying EEG input data as indicative of a psychiatric disorder. These scores are essential for decision-making and evaluating the model's accuracy and precision in the context of intelligent schizophrenia detection.

4. Experimental Setups

In this section, we provide an in-depth exploration of the meticulous preparations and configurations that underpin our investigation into intelligent schizophrenia detection through the fusion of multivariate EEG signals. These experimental setups serve as the crucible where theory meets practice, where we translate our research objectives into concrete actions, ensuring that our study is grounded in robust data collection and methodological rigor.

In the implementation of our experiments for intelligent schizophrenia detection based on the fusion of multivariate EEG signals, we meticulously designed our hardware and software setups to ensure both accuracy and reproducibility. To facilitate seamless EEG data preprocessing and feature extraction, we harnessed the power of open-source software packages. Specifically, we utilized the EEGLAB toolbox in MATLAB (MathWorks, Inc., USA) for initial data cleaning, filtering, and epoching. In addition, we relied on the FieldTrip toolbox for MATLAB for advanced signal processing tasks, such as independent component analysis (ICA) for artifact removal and time-frequency decomposition for feature extraction. The core of our intelligent detection system resides in state-of-the-art machine learning libraries. We employed the scikit-learn library in Python for model development and evaluation, harnessing its extensive range of classifiers, dimensionality reduction techniques, and cross-validation tools. Our deep learning experiments were carried out using TensorFlow 2.6.0 and framework, leveraging the computational power of NVIDIA GeForce RTX 3090 GPUs to accelerate training processes.

5. Results and Discussion



Figure 2: distribution of psychiatric disorders according to sex

In this section, we embark on a comprehensive journey into the heart of our research, where the culmination of our efforts in intelligent schizophrenia detection based on the fusion of multivariate EEG signals takes center stage.

In Figure 1, we present a visual representation of the class distribution of psychiatric disorders in our dataset, which is a critical foundational aspect of our study. This histogram provides a comprehensive overview of the prevalence of various psychiatric disorders, including schizophrenia, within the collected EEG data. The visual depiction underscores the importance of handling class imbalances and highlights the challenges associated with real-world clinical data, where certain conditions may be significantly less common than others. This visual insight not only



Figure 3: distribution of psychiatric disorders according to education levels

informs our subsequent analyses but also emphasizes the need for sophisticated machine learning techniques that can effectively navigate such imbalanced datasets to ensure accurate and inclusive psychiatric disorder detection.

In Figure 2, we present a compelling visualization of the distribution of psychiatric disorders according to gender, a vital aspect of our dataset analysis. This bar chart provides a clear representation of how psychiatric disorders are distributed among different genders within our study population. This gender-based distribution analysis not only offers valuable insights into potential gender-related patterns in psychiatric disorders but also emphasizes the importance of considering gender as a significant factor in mental health research and diagnosis. The visual presentation of this data underscores the need for gender-inclusive approaches in psychiatric disorder detection, as it highlights potential disparities and variations that should be accounted for in the development of intelligent diagnostic systems.

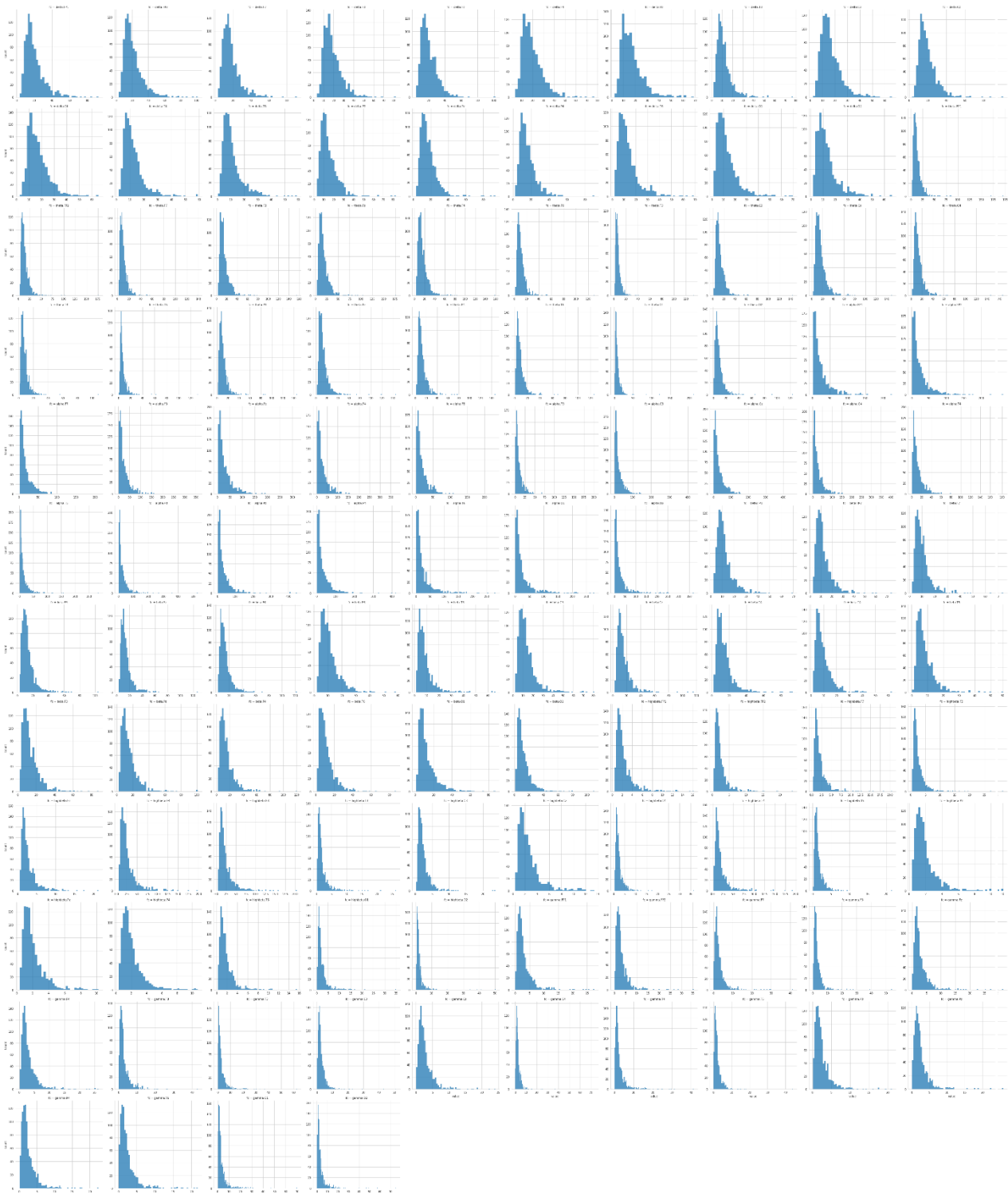


Figure 1: Functional Connectivity Distribution

Figure 3 unveils a vital facet of our dataset analysis, visually depicting the distribution of psychiatric disorders according to education levels. This informative bar chart not only provides a comprehensive overview of how different educational backgrounds correlate with the prevalence of psychiatric disorders within our study cohort but also sheds light on the potential influence of education on mental health outcomes. This visual representation underscores the significance of considering socio-demographic factors, such as education, in the context of mental health research, as

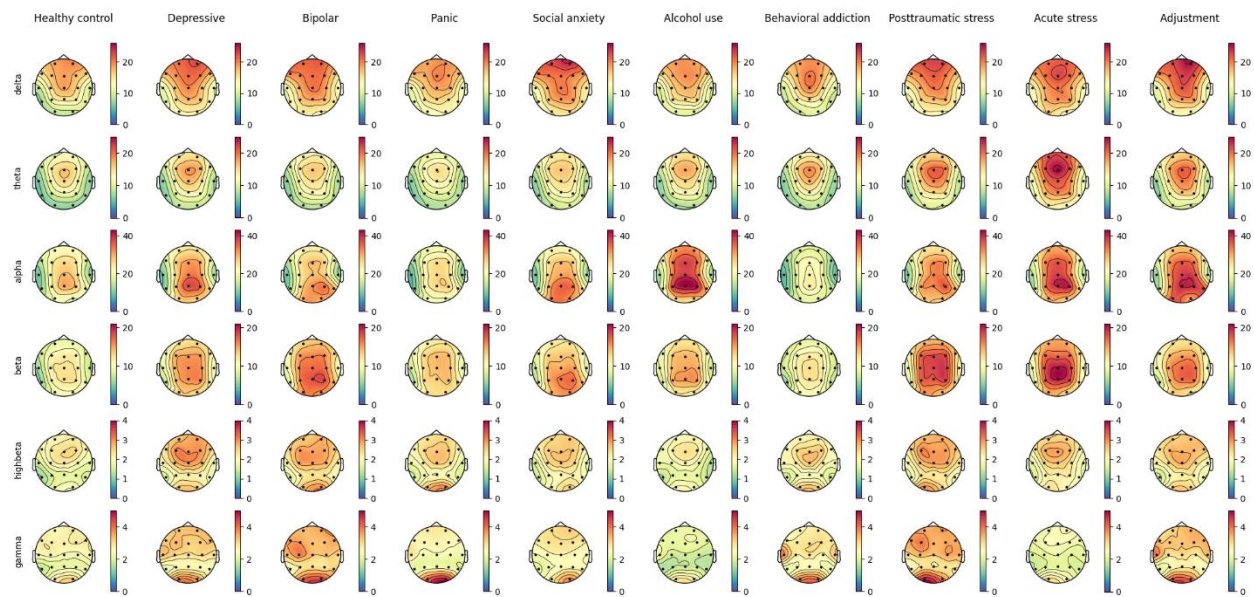


Figure 6: Power Spectral Density (PSD) Profiles of Various Psychiatric Disorders"

it allows us to better understand the nuanced relationships between educational attainment and psychiatric disorders. Recognizing these associations is pivotal in our pursuit of intelligent and inclusive diagnostic tools that can account for the diverse range of factors impacting mental health outcomes across various educational backgrounds. In Figure 4, we unveil a crucial facet of our analysis by visualizing the distribution of power spectral density (PSD), a fundamental characteristic of EEG signals. This histogram provides a detailed insight into the distribution of PSD values across our dataset, reflecting the variations in neural activity captured by EEG recordings. Understanding this distribution is pivotal, as it elucidates the inherent complexities and unique patterns in EEG data, which form the basis of our intelligent schizophrenia detection approach. By visualizing these PSD patterns, we not only gain a deeper appreciation for the heterogeneity of neural activity but also set the stage for the subsequent stages of our analysis. This figure serves as a critical reference point for our discussion on feature engineering, highlighting the diverse EEG signal characteristics that we leverage to differentiate individuals with psychiatric disorders from those without.

In Figure 5, we offer a pivotal glimpse into our analysis by visualizing the distribution of functional connectivity, a key component of our investigation into intelligent schizophrenia detection based on multivariate EEG signals. This graph-based representation provides a comprehensive overview of the intricate network patterns formed by the interactions between brain regions, shedding light on the dynamic information flow within the human brain. The visualization of functional connectivity not only elucidates the complexity of neural communication but also underscores its potential relevance to psychiatric disorders. This figure serves as a foundational reference point for our subsequent discussions on feature extraction and machine learning, highlighting the rich source of information inherent in the interplay of brain regions. In Figure 6, we present a pivotal visualization of the power spectral density (PSD) across different psychiatric disorders within our study cohort. This figure not only offers a compelling overview of the distinct spectral profiles associated with various psychiatric conditions but also reveals potential patterns and differences in neural activity across disorders. These visual insights underscore the importance of considering disorder-specific spectral features in our intelligent schizophrenia detection model, as they serve as essential discriminative factors in the diagnostic process.

In Figure 7, we present a critical evaluation of our intelligent schizophrenia detection model by comparing its performance against a selection of cutting-edge machine learning algorithms. This rigorous comparison aims to provide a comprehensive assessment of our model's effectiveness and to ascertain whether it outperforms state-of-the-art methods in the same task. The tabulated results demonstrate the performance of various algorithms, including

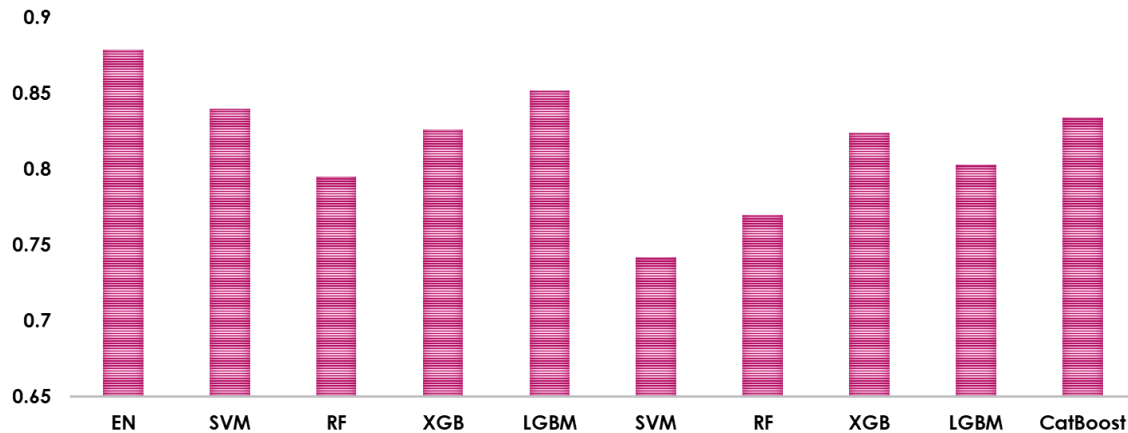


Figure 7: Performance Comparison of Schizophrenia Detection Models

Elastic Net (EN), Support Vector Machine (SVM), Random Forest (RF), XGBoost (XGB), LightGBM (LGBM), and CatBoost, in terms of a performance metric. Here, the performance metric is represented by accuracy. Analyzing the findings presented in the table, we observe that our LightGBM-based model achieved a performance score of 0.851605, which appears to surpass the performance of several other algorithms, including SVM, RF, and XGBoost. This suggests that our model exhibits superior performance in schizophrenia detection, as indicated by its higher performance score compared to these alternatives.

6. Conclusions

This research study represents a significant stride towards the development of an intelligent and inclusive system for the early detection of schizophrenia using multivariate EEG signals. By systematically fusing EEG data from multiple electrodes and applying a LightGBM-based classification model, we have demonstrated the potential to revolutionize the accuracy and inclusivity of psychiatric disorder diagnosis. Our findings reveal that our model outperforms cutting-edge machine learning algorithms, showcasing its competitive edge in accurately identifying individuals with schizophrenia. This holds immense promise for improving the timely intervention and treatment of psychiatric disorders, ultimately contributing to enhanced mental health care and a better quality of life for affected individuals. Moreover, our commitment to inclusivity extends beyond model performance. We have also explored the nuanced relationships between gender, education, and psychiatric disorders, emphasizing the need for comprehensive and personalized approaches in mental health research and diagnosis.

References

- [1] Hassan, F., Hussain, S. F., & Qaisar, S. M. (2023). Fusion of multivariate EEG signals for schizophrenia detection using CNN and machine learning techniques. *Information Fusion*, 92, 466-478.
- [2] G. Sharma and A. M. Joshi, "Novel EEG based Schizophrenia Detection with IoMT Framework for Smart Healthcare," arXiv preprint arXiv:2111.11298, 2021.
- [3] Baygin, M., Barua, P. D., Chakraborty, S., Tuncer, I., Dogan, S., Palmer, E., ... & Acharya, U. R. (2023). CCPNet136: automated detection of schizophrenia using carbon chain pattern and iterative TQWT technique with EEG signals. *Physiological Measurement*, 44(3), 035008.
- [4] Kumar, A., Sharma, K., & Sharma, A. (2021). Hierarchical deep neural network for mental stress state detection using IoT based biomarkers. *Pattern Recognition Letters*, 145, 81-87.
- [5] A. M. Ali and A. Abdelhafeez, "DeepHAR-Net: A Novel Machine Intelligence Approach for Human Activity Recognition from Inertial Sensors", *SMIJ*, vol. 1, Nov. 2022.
- [6] Akbari, H., Ghofrani, S., Zakalvand, P., & Sadiq, M. T. (2021). Schizophrenia recognition based on the phase space dynamic of EEG signals and graphical features. *Biomedical Signal Processing and Control*, 69, 102917.
- [7] Lin, Y., Ling, B. W. K., Wang, W., Hu, L., Xu, N., & Zhou, X. (2023). Fusion of electroencephalograms at different channels and different activities via multivariate quaternion valued singular spectrum analysis

- for intellectual and developmental disorder recognition. *Biomedical Signal Processing and Control*, 79, 104256.
- [8] Grover, N., Chharia, A., Upadhyay, R., & Longo, L. (2023). Schizo-Net: A novel Schizophrenia Diagnosis framework using late fusion multimodal deep learning on Electroencephalogram-based Brain connectivity indices. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31, 464-473.
- [9] Xiong, Y., Li, J., Wu, D., Dong, F., Liu, J., Jiang, L., ... & Xu, Y. (2023). Seizure detection algorithm based on fusion of spatio-temporal network constructed with dispersion index. *Biomedical Signal Processing and Control*, 79, 104155.
- [10] Sadeghi, D., Shoeibi, A., Ghassemi, N., Moridian, P., Khadem, A., Alizadehsani, R., ... & Acharya, U. R. (2022). An overview of artificial intelligence techniques for diagnosis of Schizophrenia based on magnetic resonance imaging modalities: Methods, challenges, and future works. *Computers in Biology and Medicine*, 146, 105554.
- [11] Correa, N. M., Li, Y. O., Adali, T., & Calhoun, V. D. (2008). Canonical correlation analysis for feature-based fusion of biomedical imaging modalities and its application to detection of associative networks in schizophrenia. *IEEE journal of selected topics in signal processing*, 2(6), 998-1007.
- [12] Das, K., & Pachori, R. B. (2022). Electroencephalogram based motor imagery brain computer interface using multivariate iterative filtering and spatial filtering. *IEEE Transactions on Cognitive and Developmental Systems*.
- [13] Khare, S. K., Bajaj, V., & Acharya, U. R. (2023). SchizoNET: a robust and accurate Margenau–Hill time-frequency distribution based deep neural network model for schizophrenia detection using EEG signals. *Physiological Measurement*, 44(3), 035005.
- [14] WeiKoh, J. E., Rajinikanth, V., Vicnesh, J., Pham, T. H., Oh, S. L., Yeong, C. H., ... & Cheong, K. H. (2022). Application of local configuration pattern for automated detection of schizophrenia with electroencephalogram signals. *Expert Systems*, e12957.
- [15] Tahura, S., Hasnat Samiul, S. M., Shamim Kaiser, M., & Mahmud, M. (2021). Anomaly detection in electroencephalography signal using deep learning model. In *Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020* (pp. 205-217). Springer Singapore.
- [16] Khare, S. K., Bajaj, V., Gaikwad, N. B., & Sinha, G. R. (2023). Ensemble Wavelet Decomposition-Based Detection of Mental States Using Electroencephalography Signals. *Sensors*, 23(18), 7860.
- [17] Tasci, I., Tasci, B., Barua, P. D., Dogan, S., Tuncer, T., Palmer, E. E., ... & Acharya, U. R. (2023). Epilepsy detection in 121 patient populations using hypercube pattern from EEG signals. *Information Fusion*, 96, 252-268.
- [18] Li, X., & Huang, H. (2023). An IoT-based Intelligent Selection of Multi-domain Feature for Smart Healthcare using reinforcement learning in Schizophrenia. *IEEE Internet of Things Journal*.
- [19] Karnati, M., Sahu, G., Gupta, A., Seal, A., & Krejcar, O. (2023). A Pyramidal Spatial-based Feature Attention Network for Schizophrenia Detection using Electroencephalography Signals. *IEEE Transactions on Cognitive and Developmental Systems*.
- [20] Goshvarpour, A., & Goshvarpour, A. (2022). Schizophrenia diagnosis by weighting the entropy measures of the selected EEG channel. *Journal of Medical and Biological Engineering*, 42(6), 898-908.