Neuropsychiatric Disorders Classification using EEG Signal and Deep Neural Networks

Mateus Balda Mota[1], **Alessandro Bof de Oliveira**[1], **Patricia Bof**[1], **Dante Augusto Couto Barone**[2]

[1]Universidade Federal do Pampa (UNIPAMPA) – Campus Alegrete
LAPIA – 97.541-300 – Alegrete – RS – Brazil

[2]Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

{mateusmota.aluno,alessandrooliveira}@unipampa.edu.br

pati.bof@gmail.com, barone@inf.ufrgs.br

Abstract. *Electroencephalogram (EEG) is widely used to measure brain activity and serves as a promising tool in the diagnosis of neuropsychiatric disorders such as anxiety, schizophrenia, and epilepsy, among others. We proposed a deep neural network model to classify neuropsychiatric conditions using EEG signals. The data was processed with PSD and FC features using SMOTE method to handle with unbalanced classes and dimensionality reduction using PCA. The model achieved an accuracy of 71.43% and a mean AUC of 0.9383, with better performance in the obsessive compulsive and anxiety disorders classification. At the same time, mood disorder showed to be more challenging in the classification.*

1. Introduction

Neuropsychiatric disorders impact directly social life and productivity. In 2019, it was estimated that 1 in 8 people worldwide lived with a mental disorder, totaling 970 million [World Health Organization 2022]. This number increased significantly with the COVID-19 pandemic, with a 26% rise in anxiety cases and a 28% rise in depression cases within just one year [World Health Organization 2022]. Considering this issue, methods for evaluating and diagnosing neuropsychiatric disorders are becoming increasingly important. Among the tools available, the electroencephalogram (EEG) stands out for its ability to provide data on brain activity across different regions of the brain [Siuly et al. 2016], making it valuable for identifying mental disorders [Ay et al. 2019]. When combined with machine learning approaches, the use of EEG has demonstrated relevant accuracy, particularly in disorders such as depression, schizophrenia, and anxiety [Park et al. 2021], [Ahmed et al. 2024], [Müller-Putz 2020], [Parsa et al. 2023], and [Beniczky and Schomer 2020]. According to [Ahmed et al. 2024] and [Júnior 2023], deep neural networks have demonstrated good performance in diagnostic classification, outperforming other machine learning methods. For example, [Wang et al. 2022] has used a MUCHf-Net neural network to classify EEGs from individuals with depression (DP), schizophrenia (SCZ), and healthy controls. The results showed that the network achieved an accuracy of 0.9112 in distinguishing between control and pathological EEGs, with the greatest contribution from the low-frequency bands and the frontal and parietal regions of the brain. The analysis also indicated that the model had more difficulty differentiating between DP and SCZ. [Shah et al. 2023] categorized individuals with schizophrenia (SZ), biological relatives (REL), and healthy controls (HC) using resting EEG signals from 78 cortical regions. The proposed deep neural network ETSNet achieved an accuracy of 0.9957 for classifying SZ, REL, and HC with open eyes (EO), and 0.9315 for closed eyes (EC). The authors, [Ahmed et al. 2024], used EEG and deep learning models (ANN, KNN, LSTM, Bi-LSTM, CNN-LSTM). The ANN achieved an accuracy of 0.9683 in identifying obsessive-compulsive disorder using all frequency bands. The CNN-LSTM model achieved the same accuracy rate for adjustment disorder. The KNN and LSTM models reached an accuracy of 0.9894 for acute stress disorder with specific features, while KNN and Bi-LSTM achieved an accuracy of 0.9788 for obsessive-compulsive disorder. In this work, we propose a method based on a deep neural network to classify EEGs using principal component analysis (PCA) to reduce the number of features. Another important contribution of this work is handling the unbalanced numbers of data samples in each class. Our goal is to develop an artificial intelligence model capable of classifying different types of neuropsychiatric disorders to assist specialists in diagnosing patients. In section 2, we describe the database, methods, and the proposed neural network architecture; in section 3, we show the preliminary results, and the conclusion and future works are in section 4.

2. Methodology

In this section we describe the methods and the data processing used in this work. The section 2.1 describes the EEG dataset, section 2.2 describes the preprocessing steps to prepare the dataset. The metric used to evaluate the experiments are show in section 2.3 and finally the proposed model is described in section 2.4.

2.1. EEG Dataset

We have used the dataset presented by [Park 2021], where the data were collected from SMG-SNU Boramae Medical Center, and confirmation of the diagnosis was established by two psychiatrists and two psychologists between March 2019 and August 2019. The dataset has samples of 945 patients aged between 18 and 70 years, with a mean age of 30.59 ± 11.78, an average education level of 13.43 ± 2.55 years, and an average IQ of 101.58 ± 17.02. The primary disorder categories consist of 7 classes, with a number of samples: mood disorder $= 266$, addictive disorder $= 186$, trauma and stress-related disorder $= 128$, schizophrenia $= 117$, anxiety disorder $= 107$, healthy control $= 95$, and obsessive-compulsive disorder $= 46$. Data collection used 19 EEG channels based on the international 10-20 system. The data were pre-processed by [Park et al. 2021] using the Fast Fourier Transform (FFT) to convert the signals into the frequency domain. The Power Spectral Density (PSD) was calculated for the following frequency bands: delta 1 to 4 Hz, theta 4 to 8 Hz, alpha 8 to 12 Hz, beta 12 to 25 Hz, high beta 25 to 30 Hz), and gamma 30 to 40 Hz, providing the signal power in each band. Functional Coherence (FC) was used to evaluate the synchronization between different brain regions. The detailed method is described in [Park et al. 2021]. After that, we have 1140 features (114 PSD and 1026 FC) and three quantitative variables (age, education, and IQ). The data was randomly divided into training, validation, and test sets with proportions of 70%, 15%, and 15%, respectively.

2.2. Preprocessing data

Samples with missing values (*NaN*) were identified (28 samples) for the features *education* and *IQ*. To impute the *NaN* values, we use the method *KNNImputer*[1] with the number of neighbors set to 5 that demonstrated the best results. Due to the unbalanced number of samples in each class, we have used the method SMOTE (Synthetic Minority Oversampling Technique)[2] to generate synthetic samples and standardize the number of samples to 266 (the number of samples in the largest class). The SMOTE method was used because it achieves good results with low computational cost, mitigating biased data caused by class imbalance. After that, we normalized the features using the *z-score* method and applied PCA (Principal Component Analysis) to reduce the number of features. We will use the first 223 principal components that explained 99% of the data variance. The Smote method was used only in training stage.

2.3. Metrics for performance evaluation

The metrics used in this work to evaluate the performance of the classification model were *Precision* (equation 1), *Recall* (equation 2), and *F1-Score* (equation 3), as defined by [Brownlee 2021] and detailed below:

$$\textit{Precision} = \frac{TP}{TP + FP}, \tag{1}$$

$$\textit{Recall} = \frac{TP}{TP + FN}, \tag{2}$$

And,

$$F1\textit{-score} = 2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}}. \tag{3}$$

2.4. Proposed deep neural network model

The neural network architecture [Haykin 2011] and [Russell and Norvig 2022] consists of an input layer followed by four hidden layers, with a number of the neurons defined by the list [1024, 512, 256, 128, 64], and an output layer of dimension 7. The model and its layers are described in table 1. Each hidden layer is followed by *Batch Normalization - BatchNorm1d* and the GELU (*Gaussian Error Linear Unit*) activation function. A *dropout* of 10% is applied after each hidden layer to prevent *overfitting*. The loss function chosen was Cross Entropy Loss with optimizer sets up to Adam. The learning rate scheduler used was *Reduce Plateau*, which reduces the learning rate by 10% if the loss does not improve after 3 epochs, with a minimum learning rate of (1×10^{-6}). The model was developed using the programming language *Python*[3] and the libraries *PyTorch*[4] and *Scikit-Learn*[5].

[1] https://scikit-learn.org/1.5/modules/generated/sklearn.impute.KNNImputer.html

[2] https://imbalanced-learn.org/stable/references/generated/imblearn.over$_sampling.SMOTE.html$

[3] https://www.python.org/

[4] https://pytorch.org/

[5] https://scikit-learn.org/stable/

Layer (type)	*Output Shape*	*Param #*
Linear-1	[-1, 1024]	232,448
BatchNorm1d-2	[-1, 1024]	2,048
GELU-3	[-1, 1024]	0
Dropout-4	[-1, 1024]	0
Linear-5	[-1, 512]	524,800
BatchNorm1d-6	[-1, 512]	1,024
GELU-7	[-1, 512]	0
Dropout-8	[-1, 512]	0
Linear-9	[-1, 256]	131,328
BatchNorm1d-10	[-1, 256]	512
GELU-11	[-1, 256]	0
Dropout-12	[-1, 256]	0
Linear-13	[-1, 128]	32,896
BatchNorm1d-14	[-1, 128]	256
GELU-15	[-1, 128]	0
Dropout-16	[-1, 128]	0
Linear-17	[-1, 64]	8,256
BatchNorm1d-18	[-1, 64]	128
GELU-19	[-1, 64]	0
Linear-20	[-1, 7]	455

Table 1. Model architecture of the proposed deep neural network. The value -1 in the output shape indicates that this dimension depends on the batch size.

3. Results

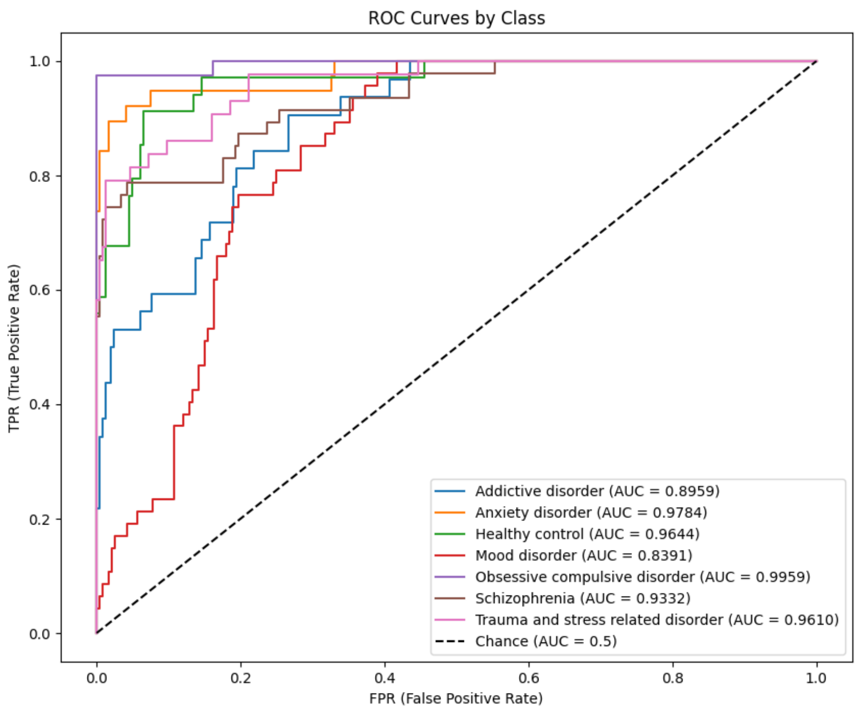
The code of the proposed model, the detailed parameters, and the scripts for training and testing can be accessed by public repository [6]. For the training step, the batch size is defined as 32, while regularization is applied using a dropout of 0.1. The learning rate is set to learning rate $= 1 \times 10^{-5}$, allowing for gradual adjustments of the weights. The model has an input dimension of 226 and an output dimension of 7. Table 2 presents the precision, recall, and f1-score for each class. We can observe in Table 2 that Mood disorder and Addictive disorder present the lowest F1-Score while Obsessive compulsive disorder and Anxiety disorder perform better. In the figure 1, the Receiver operating characteristic (ROC) and area under the curve (AUC) of the 7 classes indicates that the model has good discrimination, with some classes presenting good results, such as Anxiety disorder and Obsessive compulsive disorder, with AUCs of 0.9784 and 0.9959, respectively. The average AUC is 0.9383, which reflects good performance, although Mood disorder has a lower AUC of 0.8391. The preliminary results show that the model achieved a promising mean accuracy of 71.43%, and an average AUC of 93.83%, indicating good discrimination between classes. Disorders such as *Obsessive Compulsive* and *Anxiety* performed well, with metric values and AUCs above 0.97. On the other hand, the *Mood Disorder* class was the most challenging, with *precision* of 0.4667 and a *recall* of 0.2979. The misclassification can be originated in the reduction of features that not capture the details of the EEG behaviour. The significant oversampling of the class (*Obsessive Compulsive Dis-*

[6]https://github.com/matt-balda/eeg-signals-neuropsychiatric/

Class	***Precision***	***Recall***	**F1-*Score***
Addictive disorder	0.5667	0.5312	0.5484
Anxiety disorder	0.8095	0.8947	0.8500
Healthy control	0.6829	0.8235	0.7467
Mood disorder	0.4667	0.2979	0.3636
Obsessive compulsive disorder	0.8636	0.9744	0.9157
Schizophrenia	0.7000	0.7447	0.7216
Trauma and stress related disorder	0.7907	0.7907	0.7907
Mean Accuracy		0.7143	

Table 2. The table shows the metrics Precision, Recall and F1-score by classes and the *Mean Accuracy* for the model.

Figure 1. Receiver operating characteristic (ROC) curves and area under the curve (AUC) for each class.



order) indicate redundancies in the artificial data. The *Mood Disorder* class obtained poor results, indicating that its representation in the latent space was not sufficiently effective.

4. Conclusions and future works

The proposed method using a neural network model was able to classify the mental disorders using EEG data. This is an important result for the development of solutions that assist healthcare professionals in diagnosing mental illnesses. The results obtained in the experiments indicate that the proposed model can be improved to achieve good classification results for all classes. In future works, we intend to apply Generative Adversarial

Networks (GANs) to avoid redundant artificial data and explore other neural network architectures.

References

Ahmed, Z., Wali, A., Shahid, S., Zikria, S., Rasheed, J., and Asuroglu, T. (2024). Psychiatric disorders from eeg signals through deep learning models. *IBRO Neuroscience Reports*.

Ay, B., Yildirim, O., Talo, M., Baloglu, U. B., Aydin, G., Puthankattil, S. D., and Acharya, U. R. (2019). Automated depression detection using deep representation and sequence learning with eeg signals. *Journal of Medical Systems*, 43:1–12.

Beniczky, S. and Schomer, D. L. (2020). Electroencephalography: basic biophysical and technological aspects important for clinical applications. *Epileptic Disorders*, 22:697–715.

Brownlee, J. (2021). Tour of evaluation metrics for imbalanced classification. Acessado em: 27 nov. 2024.

Haykin, S. (2011). Neural networks: principles and practice. *Bookman*, 11(900).

Júnior, S. B. (2023). Algoritmos genéticos e aprendizado profundo baseado em redes neurais recorrentes do tipo lstm para auxílio ao diagnóstico médico.

Müller-Putz, G. R. (2020). Electroencephalography. *Handbook of Clinical Neurology*, 168:249–262.

Park, S. M. (2021). Eeg machine learning. Retrieved August 16, 2021, from `https://osf.io/8bsvr`.

Park, S. M., Jeong, B., Oh, D. Y., Choi, C. H., Jung, H. Y., Lee, J. Y., Lee, D., and Choi, J. S. (2021). Identification of major psychiatric disorders from resting-state electroencephalography using a machine learning approach. *Frontiers in Psychiatry*, 12:707581.

Parsa, M., Rad, H. Y., Vaezi, H., Hossein-Zadeh, G.-A., Setarehdan, S. K., Rostami, R., Rostami, H., and Vahabie, A.-H. (2023). Eeg-based classification of individuals with neuropsychiatric disorders using deep neural networks: A systematic review of current status and future directions. *Computer Methods and Programs in Biomedicine*, 240:107683.

Russell, S. J. and Norvig, P. (2022). *Artificial intelligence: a modern approach*. Pearson.

Shah, S. J. H., Albishri, A., Kang, S. S., Lee, Y., Sponheim, S. R., and Shim, M. (2023). Etsnet: A deep neural network for eeg-based temporal–spatial pattern recognition in psychiatric disorder and emotional distress classification. *Computers in Biology and Medicine*, 158:106857.

Siuly, S., Li, Y., and Zhang, Y. (2016). Eeg signal analysis and classification. *IEEE Trans Neural Syst Rehabilit Eng*, 11:141–144.

Wang, Z., Feng, J., Jiang, R., Shi, Y., Li, X., Xue, R., Du, X., Ji, M., Zhong, F., Meng, Y., et al. (2022). Automated rest eeg-based diagnosis of depression and schizophrenia using a deep convolutional neural network. *IEEE Access*, 10:104472–104485.

World Health Organization (2022). Mental disorders. Accessed: 2024-10-10.