Electrode Importance for EEG-based Schizophrenia Detection

Luís A. Alexandre[1[0000−0002−5133−5025]], **Wlodzisław Duch**[2[0000−0001−7882−4729]]

[1]*Universidade da Beira Interior and NOVA LINCS*
Covilhã, Portugal
luis.alexandre@ubi.pt
[2]*Dept. of Informatics, Institute of Engineering and Technology, Faculty of Physics, Astronomy & Informatics, Nicolaus Copernicus University,*
Toruń, Poland
wduch@umk.pl

Abstract. *EEG is the preferred technique for objective diagnosis of mental disorders. Unfortunately, complex methods of EEG signal processing do not generalize well beyond a single data set and have not yet provided clinically useful biomarkers. Good biomarkers should be simple to interpret, use a minimal number of electrodes, and be based on short resting state EEG recordings. We have used the vector autoregressive EEG signal representation of two schizophrenia data sets, showing that a simple k-NN classifier provides state-of-the-art results with only 5 or 6 EEG channels.*
Keywords: *EEG, Resting State, Schizophrenia Diagnosis, Electrode Selection*

1. Introduction

EEG has been used as a simple means to detect several brain disorders and increase our understanding of the brain. In this paper, we focus on the relative importance of the electrodes used in EEG. We expect that results will strongly depend on the type of disorder. With a large number of electrodes and small data sets (typically less than 100 cases, including the control group), many input combinations may give similar accuracy. Although our method is general, We focus here on the detection of schizophrenia using two popular data sets. The importance of identifying which electrodes have more discriminative power is twofold: first, it

can help as a feature selection technique to use the best possible combination of EEG signal sources to improve the detection; second, it helps to identify which brain regions are involved in schizophrenia pathology.

A review of candidate biomarkers in psychiatric disorders has recently been published in [1]. Despite enormous investment in biomarkers based on neuroimaging, genetics, molecular and peripheral assays for autism, schizophrenia, anxiety disorders, major depression, bipolar disorder, substance use disorders, and PTSD, we do not have reliable biomarkers for objective diagnosis of patients. Many sophisticated methods have been proposed for diagnosis based on EEG, reviewed in [2], but they are too complex and thus difficult to use in clinical practice. Only simple and robust methods have a chance to be useful in practice and provide a real baseline for more refined approaches.

This paper is organized as follows: the next section presents a description of both the data used in this work and the methods used in the evaluation; the following section contains the experiments, followed by a discussion and some conclusions.

2. Methods and Data

To identify the most relevant electrode(s) for schizophrenia diagnosis, we will use a brute force approach that tries all possible electrode combinations and, for each of these, perform a k-fold cross-validation. The brute force method is done only once and ensures that we discover all sets of electrodes that give the best results. The data sets are not completely balanced; therefore, we will select the combinations that show the best F1 score for analysis.

2.1. Data Sets

For tests, we have selected the two most commonly used data sets [2]. The first [3], which we will refer to in the paper as data set A, contains measurements collected from 84 male adolescent subjects, 45 with schizophrenia, and the remaining 39 in the healthy control group. The data was collected using 16 electrodes at the 128Hz sampling rate.

The second data set [4], which we will call data set B, contains data from only 28 subjects, 14 control and 14 with schizophrenia. It was collected using 19 electrodes at the 250Hz sampling rate. To make the data sets similar in terms of the number of electrodes used and facilitate the comparison of results and the ability

to obtain results in an acceptable time frame, we reduce the 19 electrodes in this data set to the same 16 used in data set A. We also subsample the data from 250Hz to 128Hz, and use only the first minute of data recordings for all subjects. This makes both data sets comparable in terms of the position of electrodes, sampling frequency, and signal duration. For both data sets, the only pre-processing done consisted of applying a high-pass filter to remove data below 0.1Hz.

2.2. Data Representation

The data is a set of N $(C \times S)$ matrices, where N is the number of subjects, C the number of channels (electrodes), and S is the number of samples. Each of these matrices is represented using the Vector AutoRegressive (VAR) approach [5], by considering the data at each time step i as a random vector y_i, and modeling each subject's data at a given time step t as a combination of the previous L random vectors:

$$y_t = \nu + u_t + \sum_{i=1}^{L} A_i y_{t-i}$$

where the $C \times C$ matrices A_i and the $C \times 1$ vector ν contain the model's parameters. These will be used to represent the data. The value L is called the lag, and u_t is a zero mean random noise vector.

2.3. Classification

For each combination of electrodes, a k-fold cross-validation is used to evaluate the capability of the combination to distinguish between normal and schizophrenic patients. As a classifier, we used the 3-NN. It is very fast for small data sets, avoids draws for the two class problems, and provides deterministic results. Following the most used approaches in the literature, 5-fold cross-validation on data set A and 10-fold on the smaller data set B was used.

3. Experiments

Calculations were performed on a PC with an AMD Ryzen 7 3700X 8-Core processor, Pop!_OS 22.04, 32GB RAM, 1TB SSD and an NVidia RTX 3080TI GPU. The code was made in Python 3.10, and we used the libraries Scikit-learn for the classifier, Statsmodels for the VAR model, and Matplotlib for the figures. These calculations took around 40 hours to execute for each data set.

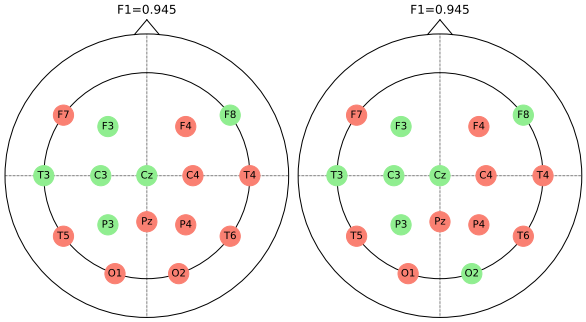

Figure 1. The two best subsets of electrodes for data set A, both with F1 score of 0.945. Note that they only differ in electrode O2.

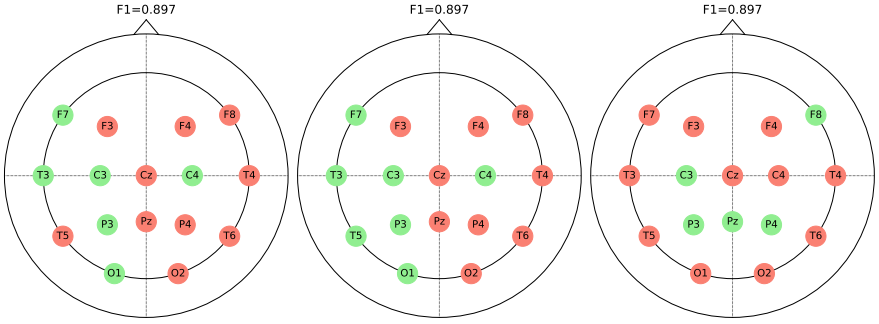

Figure 2. The three best subsets of electrodes for data set B, all with an F1 score of 0.897. Note that the two on the left and center only differ in electrode T5.

Searching for minimal sets of electrodes that reliably distinguish schizophrenic patients, we check all possible combinations of electrodes. For m electrodes, we conduct $2^m - 1$ k-fold cross-validation runs to obtain an estimate of the classification F1 score for each subset. For both data sets, we have 16 electrodes, so the total number of possible combinations is $2^{16} - 1 = 65535$ (all combinations with the exception of the case where no electrodes are used).

In data set A, out of the 65535 combinations, the best two yielded an excellent F1 score of 0.940. Figure 1 shows the electrodes used in these two subsets. For data set B, the three best subsets with F1 score of 0.897 were discovered. Classification using all electrodes (the original data set) gave an F1 score of 0.854 in data set A and 0.667 in data set B.

4. Discussion

Our goal was to establish a simple reference for more advanced machine learning approaches to the diagnosis of mental disorders. We have found that reducing 19 or 16 EEG channels to just five or six, and using the simplest classification method leads to excellent results. Comparison of results achieved with deep learning methods [2] shows that our simple approach is as good as any other method. Fusion of a vector autoregressive model, partial directed coherence, complex network measures of network topology to generate hand-crafted and learned features, followed by three convolutional neural networks (CNNs) in 2730 dimensional input space, reaches 91.7±4.6% accuracy, or F1=0.93 [6]. Dozens of papers have reached significantly worse results than our simple benchmark calculations.

Positions of the electrodes, presented in Figures 1 and 2 for both data sets, may be justified based on the interpretation of differences in EEG power distribution between healthy controls and schizophrenic adolescents. We have already done a preliminary analysis of 5 classical frequency bands (delta-gamma), finding that in θ band F8, T3, T4, P4 electrodes are most important, in α bend F8, O2, Pz, and in β band P3 [7]. Unfortunately, we do not have space here to describe a detailed relationship of these findings with the vast literature on schizophrenia.

5. Conclusions

The main contribution of this paper is to show that a simple baseline approach should be performed before sophisticated deep-learning methods are applied to complex data. They frequently obfuscate the interpretation and are not more accurate than simple approaches. Therefore, it should be mandatory to compare results with simpler methods. Second, we have performed a selection of EEG channels showing that simple equipment, with just 5 electrodes, may provide accurate data. Here, we have analyzed two data sets commonly used to test methods for schizophrenia diagnosis. They are rather small, but in this field, it is impossible to find large data sets.

We will perform a similar analysis on other schizophrenia data sets, describe the relations of our findings to known brain processes that characterize schizophrenia, analyze power distributions in relevant brain regions, and use other EEG data sets to check if a simple reference model can also be used for other psychiatric conditions.

Acknowledgment

This work was supported by NOVA LINCS (UIDB/04516/2020) with the financial support of FCT.IP, Portugal, and the Polish National Science Center grant UMO-2016/20/W/NZ4/00354.

References

[1] Abi-Dargham, A., Moeller, S. J., Ali, F., and et. al. Candidate biomarkers in psychiatric disorders: state of the field. *World Psychiatry*, 22:236–262, 2023.

[2] Khare, S. K., Bajaj, V., and Acharya, U. R. 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, 2023.

[3] Borisov, S., Kaplan, A., Gorbachevskaya, N., and Kozlova, I. Analysis of EEG structural synchrony in adolescents with schizophrenic disorders. *Human Physiology*, 31(3), 2005.

[4] Olejarczyk, E. and Jernajczyk, W. EEG in schizophrenia, 2017. URL `https://doi.org/10.18150/repod.0107441`.

[5] Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*. New York, Springer, 2005.

[6] Phang, C.-R., Noman, F., Hussain, H., and et. al. A multi-domain connectome convolutional neural network for identifying schizophrenia from EEG connectivity patterns. *IEEE J. of Biomedical and Health Informatics*, 24:1333, 2020.

[7] Duch, W., Tołpa, K., Ratajczak, E., Hajnowski, M., Furman, L., and Alexandre, L. A. Asymptotic spatiotemporal averaging of the power of EEG signals for schizophrenia diagnostics. Communications in Computer and Information Science, page 428–439. Springer Nature, Singapore, 2024.