

An Interpretable Machine Learning Method for the Detection of Schizophrenia Using EEG Signals

Manuel A. Vázquez¹, Arash Maghsoudi^{2*} and Inés P. Mariño^{3,4,5*}

¹ Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Leganés, Spain, ² Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran, ³ Department of Biology and Geology, Physics and Inorganic Chemistry, Universidad Rey Juan Carlos, Móstoles, Spain, ⁴ Research Laboratory Systemic Medicine of Healthy Ageing, Institute of Biology and Medicine, National Research Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia, ⁵ Institute for Women's Health, University College London, London, United Kingdom

In this work we propose a machine learning (ML) method to aid in the diagnosis of schizophrenia using electroencephalograms (EEGs) as input data. The computational algorithm not only yields a proposal of diagnostic but, even more importantly, it provides additional information that admits clinical interpretation. It is based on an ML model called random forest that operates on connectivity metrics extracted from the EEG signals. Specifically, we use measures of generalized partial directed coherence (GPDC) and direct directed transfer function (dDTF) to construct the input features to the ML model. The latter allows the identification of the most performance-wise relevant features which, in turn, provide some insights about EEG signals and frequency bands that are associated with schizophrenia. Our preliminary results on real data show that signals associated with the occipital region seem to play a significant role in the diagnosis of the disease. Moreover, although every frequency band might yield useful information for the diagnosis, the beta and theta (frequency) bands provide features that are ultimately more relevant for the ML classifier that we have implemented.

Keywords: electroencephalography, machine learning, random forest, schizophrenia, connectivity, direct directed transfer function, generalized partial directed coherence

OPEN ACCESS

Edited by:

Shangbin Chen,
Huazhong University of Science and
Technology, China

Reviewed by:

Sergio E. Lew,
University of Buenos Aires, Argentina
Ehren Lee Newman,
Indiana University Bloomington,
United States

*Correspondence:

Inés P. Mariño
ines.perez@urjc.es
Arash Maghsoudi
maghsoudi@srbiu.ac.ir

Received: 12 January 2021

Accepted: 30 April 2021

Published: 28 May 2021

Citation:

Vázquez MA, Maghsoudi A and
Mariño IP (2021) An Interpretable
Machine Learning Method for the
Detection of Schizophrenia Using
EEG Signals.
Front. Syst. Neurosci. 15:652662.
doi: 10.3389/fnsys.2021.652662

1. INTRODUCTION

Schizophrenia is a severe mental disorder that compromises significantly many aspects of the quality of life and affects more than 20 million people worldwide (Insel, 2010). Due to the absence of validated and reliable biological markers, diagnosis of schizophrenia is mostly subjective and mainly based on documented symptoms (such as hallucinations, disorganized speech, etc.), their duration, or apathy at work and/or social activities (see Segal, 2010 for a thorough review). While schizophrenia is known to have an effect on the activity of the brain (Rubinov et al., 2009), other mental disorders such as, e.g., obsessive compulsive disorder, attention deficit hyperactivity disorder, or bipolar disorder produce similar variations in the baseline brain activity (Anier et al., 2012). Moreover, mental diseases such as bipolar disorder or major depressive disorder are often confused with schizophrenia. Thus, finding automatic tools to help clinicians in the diagnosis of the disease is a challenging problem.

In recent years machine learning (ML) techniques have become important tools in addressing classification tasks that involve medical problems. As examples, we can mention the use of

long short-term memory recurrent neural networks (RNNs) to classify diagnoses from pediatric intensive care unit data (Lipton et al., 2015), the use of RNNs and Bayesian models to discriminate patients with ovarian cancer (Mariño et al., 2017; Vázquez et al., 2018), the use of support vector machines (SVMs) for attention deficit hyperactivity disorder prediction (Dai et al., 2012), the application of convolutional neural networks (CNNs) to classifying electroencephalogram (EEG) signals for emotion recognition (Luo et al., 2020), or the combination of multilayer perceptrons and SVMs to diagnose major depressive disorders (Saeedi et al., 2020b).

In this work we contribute to this mainstream of research by proposing an ML method to aid in the diagnosis of schizophrenia using EEGs as input data. This type of signals has been extensively used in the literature for non-invasive studies of the brain electrical activity (Asadzadeh et al., 2020), including classification of several mental disorders as dementia (Durongbhan et al., 2019), depression (Acharya et al., 2018; Saeedi et al., 2020a) and schizophrenia (Sabeti et al., 2011; Thilakvathi et al., 2017; Shalhaf et al., 2020; Chandran et al., 2021).

Current research in the literature for schizophrenia classification is rooted in black-box models that fail to provide transparency for clinicians (Sabeti et al., 2011; Thilakvathi et al., 2017; Shalhaf et al., 2020; Chandran et al., 2021). While such methods may attain good classification performance on validation data sets, they yield no “justification” of their outputs (i.e., these outputs bear no interpretable features). Clinicians should be able to judge how automatic classifications are made, and choose how to use that information in combination with their own examination and training. Some attempts at interpretability in the field of discrimination of patients with schizophrenia have been carried out very recently by using magnetic and structural magnetic resonance imaging (MRI) (de Pierrefeu et al., 2018; Reiter, 2020). Similarly, Acar et al. (2017) have proposed a technique based on tensor decompositions for the identification of event related potentials (ERPs) in functional MRI data and EEG recordings that may be indicative of schizophrenia.

The computational technique proposed in this paper not only yields a proposal of diagnosis but, even more importantly, it provides additional information that admits clinical interpretation. Specifically, the proposed approach is based on an ML model called random forest (Breiman, 2001) that operates on connectivity metrics extracted from EEG signals. Random forests allow assessing which input features are most relevant to the classification task, which is the reason why they are specially appealing for the problem at hand. Input features to this ML model are given by measures of generalized partial directed coherence (GPDC) (Baccalá and Sameshima, 2001; Baccalá et al., 2007) and direct directed transfer function (dDTF) (Kamiński et al., 2001) computed from the EEG signals. The studies we have conducted using the outlined methodology allow to identify EEG signals (and frequency bands) that might play a key role in revealing important information about the physiology of the brain in schizophrenia patients.

The paper is organized as follows. In Section 2 we describe the dataset used in this manuscript as well as the necessary

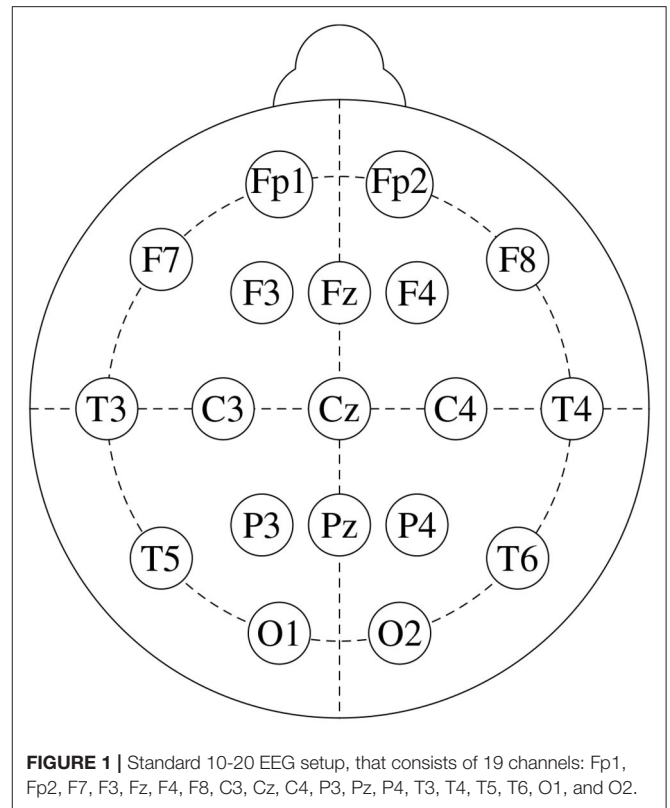


FIGURE 1 | Standard 10-20 EEG setup, that consists of 19 channels: Fp1, Fp2, F7, F3, Fz, F4, F8, C3, Cz, C4, P3, Pz, P4, T3, T4, T5, T6, O1, and O2.

pre-processing. The random forest algorithm is presented in Section 3. In Section 4 we describe the two data analysis carried out in this work, and Section 5 is devoted to a discussion of the results.

2. DATA PRE-PROCESSING

2.1. Raw Data

We base this study on freely available data from a public repository¹. It consists of EEG recordings from 14 patients suffering from paranoid schizophrenia (7 males: 27.9 ± 3.3 years old, and 7 females: 28.3 ± 4.1 years old) and 14 age-matched healthy controls (7 males: 26.8 ± 2.9 years old, and 7 females, 28.7 ± 3.4 years old). Signals were recorded for 12 minutes, with subjects in closed-eyes state, with a sampling frequency of 250 Hz. **Figure 1** shows the standard 10-20 EEG setup that was used to record the data. It yields 19 channels per subject: Fp1, Fp2, F7, F3, Fz, F4, F8, C3, Cz, C4, P3, Pz, P4, T3, T4, T5, T6, O1, and O2. For further details on the dataset, see Olejarczyk and Jernajczyk (2017).

From the latter reference, and regarding the use of data coming from human subjects, it is important to emphasize that the “study protocol was approved by the Ethics Committee of the Institute of Psychiatry and Neurology in Warsaw,” and “all participants received a written description of the

¹<http://dx.doi.org/10.18150/repod.0107441>

protocol and provided written consent to take part” in the corresponding study.

2.2. Effective Connectivity Estimation

Effective connectivity has become a prevalent analysis tool in current neuroimaging since it is able to explore causal effects between different brain areas and determine directionality of neural interactions (Astolfi et al., 2007). The most popular strategy for connectivity estimation is Granger causality (GC) which is a model-based methodology (Granger, 1969). In the event that a signal x can be estimated by previous data from another signal y in a way that is better than from its own data, then the signal y is viewed as the cause for the primary signal x . GC measures can be obtained in the frequency domain, which permits the investigation of EEG recurrence in different bands (Geweke, 1984). In order to accomplish this, a multivariate autoregressive (MVAR) model of each individual signal is required.

Let x_t be a vector representing an m channels signal at time t . The MVAR model is written as

$$x_t = \sum_{k=1}^p A_k x_{t-k} + u_t, \tag{1}$$

where p denotes the model order, A_k is an $m \times m$ matrix, and u_t is an $m \times 1$ (column) vector of white noise with covariance matrix C . By rearranging terms, Equation (1) can be written as

$$u_t = \sum_{k=0}^p \hat{A}_k x_{t-k}, \text{ with } \hat{A}_k = \begin{cases} -A_k, & k = 1, \dots, p \\ I_p, & k = 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where I_p is the identity matrix of order p . The summation on the right-hand-side of Equation (2) is a convolution sum, and hence taking the Fourier transform on both sides of the equation, we have

$$U(f) = A(f)X(f), \tag{3}$$

where $U(f)$ and $X(f)$ are the spectral representations of vectors u_t and x_t , respectively, and

$$A(f) = \sum_{k=0}^p \hat{A}_k e^{-2\pi f k \sqrt{-1}}. \tag{4}$$

Solving for $X(f)$ in Equation (3) results in

$$X(f) = A(f)^{-1}U(f). \tag{5}$$

For the sake of conciseness we define the $m \times m$ matrices

$$\begin{aligned} H(f) &= A(f)^{-1} \\ S(f) &= X(f)X^*(f), \end{aligned} \tag{6}$$

which along with $A(f)$ can be exploited to compute various effective connectivity measures. Two common quantitative

spectral measures are GPDC (Baccalá and Sameshima, 2001; Baccala et al., 2007) and dDTF (Kamiński et al., 2001), which are defined between channels i and j as, respectively,

$$GPDC_{ij}(f) = \frac{\frac{1}{C_{ii}} A_{ij}(f)}{\sqrt{\sum_{k=1}^m \frac{1}{C_{kk}} |A_{kj}(f)|^2}} \tag{7}$$

and

$$dDTF_{ij}(f) = \frac{|H_{ij}(f)|^2}{\sum_f \sum_{k=1}^m |H_{ik}(f)|^2} \frac{S_{ij}^{-1}(f)}{\sqrt{S_{ii}^{-1}(f)S_{jj}^{-1}(f)}}, \tag{8}$$

where, for a matrix A , we denote as A_{ij} the element in the i -th row and the j -th column. Both GPDC and dDTF try to describe the causal relationship between a pair of signals (coming from the EEG, in our particular case), say i and j . GPDC puts the focus on signal i as a source (producing a flow of information), whereas dDTF is concerned with signal j considered as a sink (receiving the flow).

We have computed both connectivity measures, $GPDC_{ij}(f)$ and $dDTF_{ij}(f)$, for every possible combination of EEG channels ($i, j = 1, \dots, m, i \neq j$), and frequency bands:

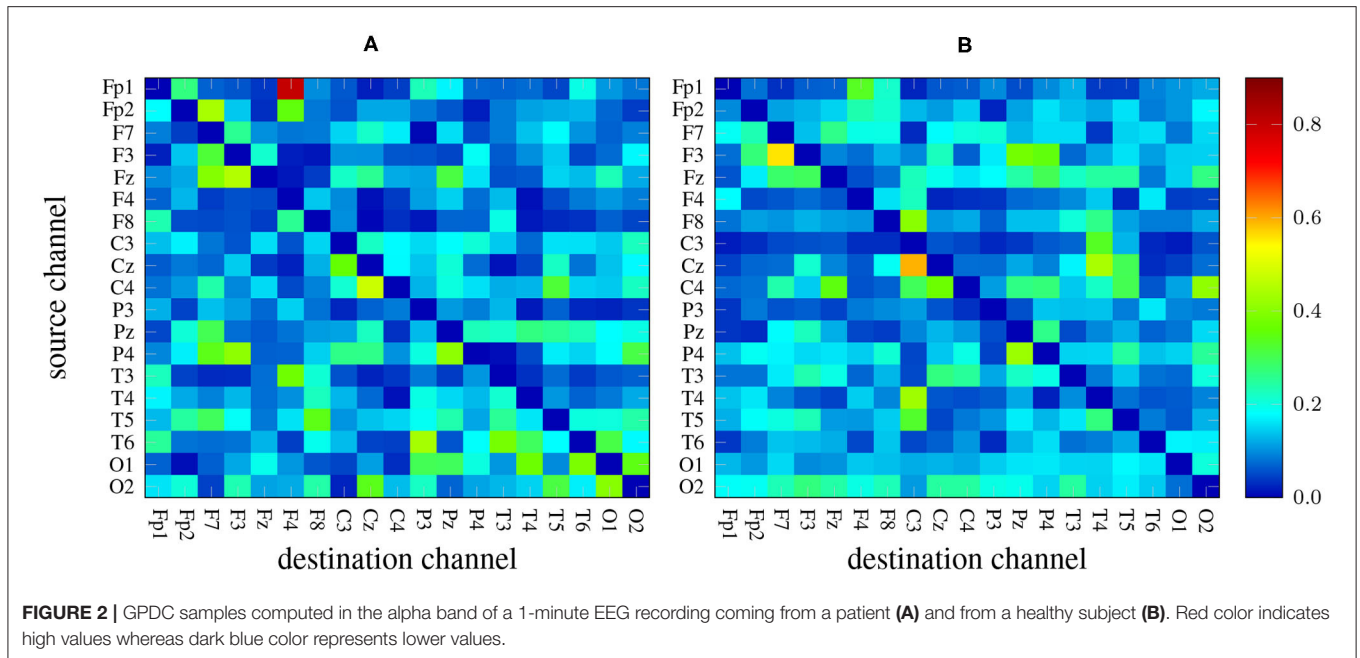
- delta (1-4 Hz),
- theta (4-8 Hz),
- alpha (8-12 Hz),
- beta (12-30 Hz), and
- gamma (30-50 Hz).

There are $19 \times 19 - 19 = 342$ channel combinations² that, along with the 5 frequency bands yield an overall number of 1710 features per measure. As an example, **Figure 2** depicts two heat maps with the values of GPDC (for every pair of channels) computed in the alpha band of a 1-minute EEG segment coming from a patient (left) and a healthy subject (right). Notice the values in the diagonal are, in any case, all zero.

2.3. From Raw EEG to Samples

Raw signals were processed with low- and high-pass Butterworth filters with cut-off frequencies of 0.5 and 50 Hz, respectively. Afterwards, the EEG recording of every subject (encompassing the 19 channels) was split into 1-minute windows (signal segments). An MVAR model with $p = 5$ was then fitted (see Equation (1)) to every individual window. From the latter, Equations (7), (8) allow computing the GPDC and dDTF metrics, respectively, between any given signals, i and j , and for any frequency of interest, f . Getting the value of a metric for a certain frequency band (as opposed to a particular frequency, f) involves evaluating the metric at a sequence of frequencies covering the corresponding interval and computing the average. In our experiments, this sequence is spanned from the lower to the upper bounds of the interval by increments of 1 hertz. The length of the window used to split the signals into samples (1 minute) and the order of MVAR model fitted to the EEG signals

²Notice that the connectivity measures are not *symmetric* and hence $GPDC_{ij}(f) \neq GPDC_{ji}(f)$.



($p = 5$) were selected using the autocorrelation function and portmanteau tests.

Ultimately, all the EEG recordings are split into 644 segments, and this is the overall number of samples. Each one of them encompasses features from both GPDC and dDTF connectivity measures, and hence has size 3420.

3. MATERIALS AND METHODS

A random forest (Breiman, 2001) is simply an *ensemble* of decision trees, each one trained on a different subset of the data and the available features. A decision tree (Magee, 1964), in turn, is a modeling approach based on splitting a collection of data points into mutually exclusive groups by asking a series of binary *yes-or-no* questions. Each group or *leaf* in the resulting tree is then assigned an outcome (a number) in a regression problem, or a label in a classification one. **Figure 3** shows an example of decision tree to tell apart the *Russian blue* and *Korat* cat breeds.

Training a decision tree involves finding, at every step of the algorithm, the best split of the data according to some prescribed metric. The process starts with all the samples at the *root* of the tree, and proceeds recursively splitting the samples at every leaf node (in the beginning just the root) into two children nodes until a certain stopping criterion is met. At every step of the algorithm, in order to find the best split for a node, we need to compare the results that would be obtained using every possible combination of feature and value thereof. This comparison is carried out using the metric of choice, which in our case is the Gini index (Breiman et al., 1984). The latter yields a measure of the *impurity* of a group of samples according to the classes they belong to. Specifically, if in a binary classification problem a certain node has p_A percentage of samples from class *A* and $p_B = 100 - p_A$ percentage from class *B*, then the Gini impurity

of that node is given by

$$G = p_A(1 - p_A) + p_B(1 - p_B). \quad (9)$$

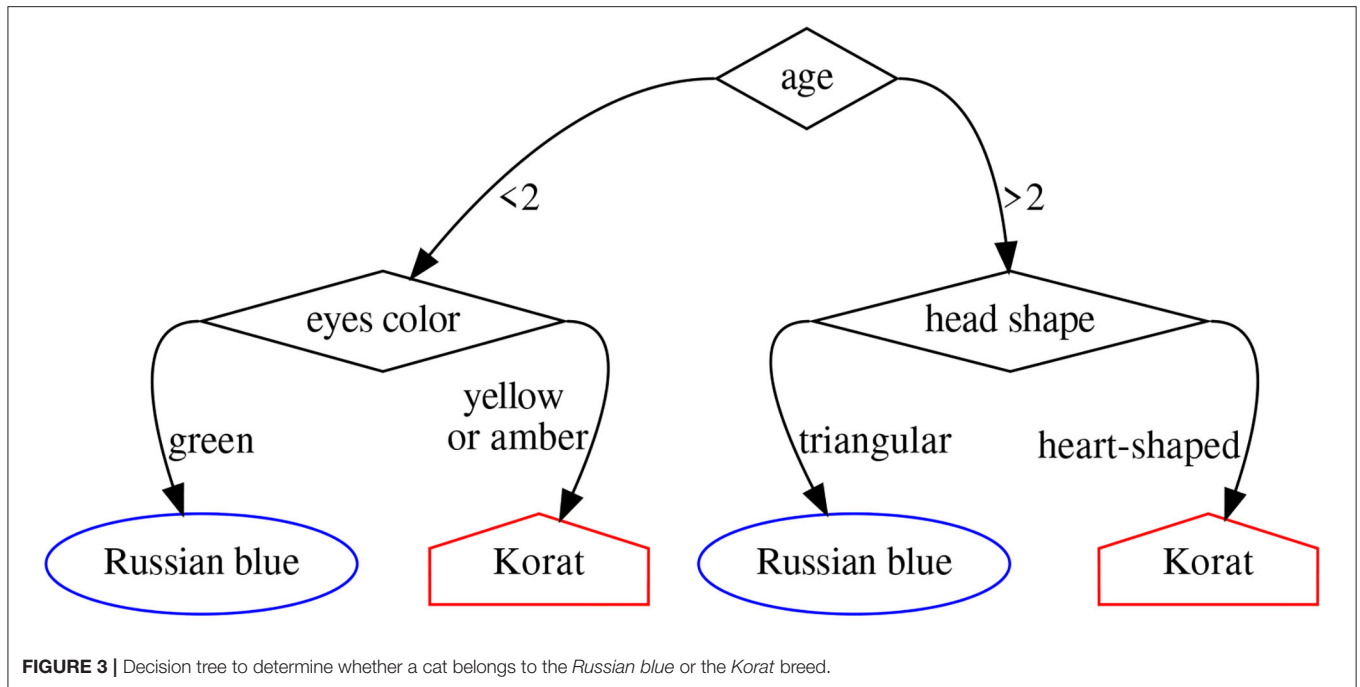
When evaluating a split for a node, each child will have its own Gini index, and a measure of the latter for the parent node can be computed as a weighted average of those from the children, each one multiplied by the percentage of samples from the parent node that it contains after the split.

The training algorithm is greedy in the sense that, at every step, the best split is selected for each node (that must be further split according to the stopping rule), and no *backtracking* is later performed. The stopping rule is a hyperparameter, and a common choice is to not further split leaf nodes whose number of samples is below a certain threshold (Ranganathan et al., 2018). We abide by this criterion here.

As mentioned above, a random forest is simply a collection of decision trees that are trained on different subsets of the same dataset. They are based on the idea of *bagging predictors* (Breiman, 1996), which consists in constructing different versions of a classifier (a predictor) each one trained on a different *bootstrap replicate*³ of the training set. Since the classifiers are trained on different datasets, the errors they make are (approximately) uncorrelated, and hence their average is 0 (assuming every individual classifier is working properly and its expected error is also 0).

Bagging is a very general procedure in machine learning that can be, in principle, applied to many different kinds of predictors. In random forests, *bootstrap replicates* are obtained from the training set by subsetting both dimensions in the data: the *sample dimension* and the *feature dimension*. In other words, each

³This is simply an independent draw from a distribution, which is here provided by the training dataset.



version of the classifier is trained on a subset of randomly selected samples, using only a subset of randomly selected features.

One appealing feature of random forests is that, after training, they allow computing a measure of *importance* for every feature that indicates how much it contributed to the classification process. This is achieved by exploiting the Gini index in a slightly different way. For a certain feature in a given decision tree, Gini *importance* is computed by adding up the *decrease* in the Gini index that is attained every time a split on that feature takes place. In a random forest, we must account for this metric in all the trees in the ensemble (see Breiman et al. (1984) for details). Although other (equally performing) measures of importance are possible, we rely on the Gini index because it is readily available in most ML software libraries.

4. RESULTS

Random forests often require very little tuning (see e.g., Hastie et al., 2009). In this particular case, each random forest encompasses $T = 200$ individual decision trees in which the minimum number of samples per leaf is $M = 10$. Every decision tree is built on only 85 randomly selected features, which is around $P = 2.5\%$ of the $N = 3420$ overall number. The hyperparameters T , M , and P have been selected empirically after a few preliminary experiments.

Since (after pre-processing) the number of samples is relatively small (only 644), we use K -fold cross-validation to get a more accurate assessment of the classifier's performance. Hence, we split the data into K equal-sized disjoint sets (also known as folds), and separately (in turns) evaluate the performance on each one while training on the rest. We apply this strategy with $K = 7$ in two different ways.

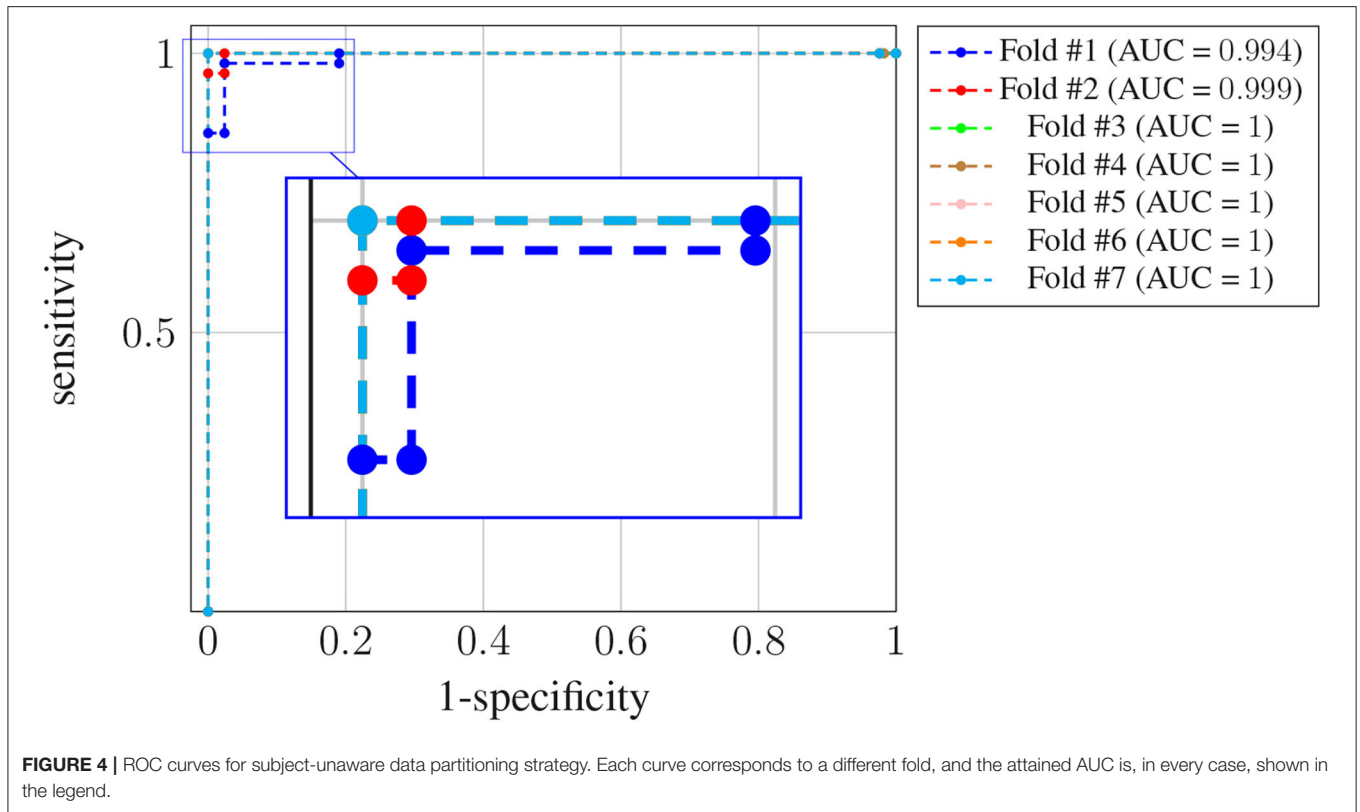
4.1. Subject-Unaware Partitioning

We first split the samples into training and test sets ignoring the subject from which each sample originates. Therefore, this becomes a regular binary classification problem in which each sample is labeled as “coming from a healthy subject” or “coming from a patient”. In order to implement this strategy, we carry out the training-test split within each subject (according to the ratio determined by the number of folds) and then the training sets from all the individual subjects are stacked together to yield the (overall) training set, and an analogous procedure is used to construct the (overall) test set⁴.

Figure 4 shows the receiver operating characteristic (ROC) curve for each one of the $K = 7$ folds. It can be seen that, in every case, a sensitivity very close to 1 can be attained even for large values (also close to 1) of the specificity (notice the horizontal axis is not specificity but its complementary). Also, the difference in the ROC curves of the different folds is negligible, and confined to very large values of the specificity. The zoom box in the figure attempts to make this difference noticeable. The legend next to the plot also shows the area under the curve (AUC) attained for every fold. On average it is above 0.99.

Since we have different results of *feature importance* for different folds, we are going to use the minimum and the average as summary statistics. Specifically, the minimum importance is used to assess whether a certain feature was consistently important across all the folds, whereas the average importance is used to aggregate the results from all the folds. If we let I_i^j denote the importance of the i -th feature in the j -th fold, then

⁴Notice that, since we have 14 healthy subjects and 14 patients, this scheme yields perfectly balanced (same number of cases and controls) training and test sets.



the minimum importance of the i -th feature is

$$I_i^{\min} = \min \{I_i^j\}_{j=1}^K \quad (10)$$

while the average importance is

$$I_i^{\text{avg}} = \frac{1}{K} \sum_{j=1}^K I_i^j. \quad (11)$$

Notice that the importance of every feature is normalized so that they all add up to 1, and hence the importance of a given feature provides information about its *relative* importance as compared to others.

Figure 5 shows the top 10 features when these are ranked (in descending order) according to their corresponding I_i^{\min} , in the top panel, and I_i^{avg} in the bottom one. Every feature name is colored differently but consistently across panels for easier comparison.

It can be seen that the exact same 10 features are present in both panels, meaning not only they have a large importance on average, but they are consistently important across folds (i.e., it is not the case that they are very important in a certain fold and not at all in the rest). Moreover, some features are ranked at exactly the same place in both panels of **Figure 5**.

4.2. Leave-p-Subjects-Out

A more realistic assessment of the model performance entails building the training and test sets while accounting for the subject

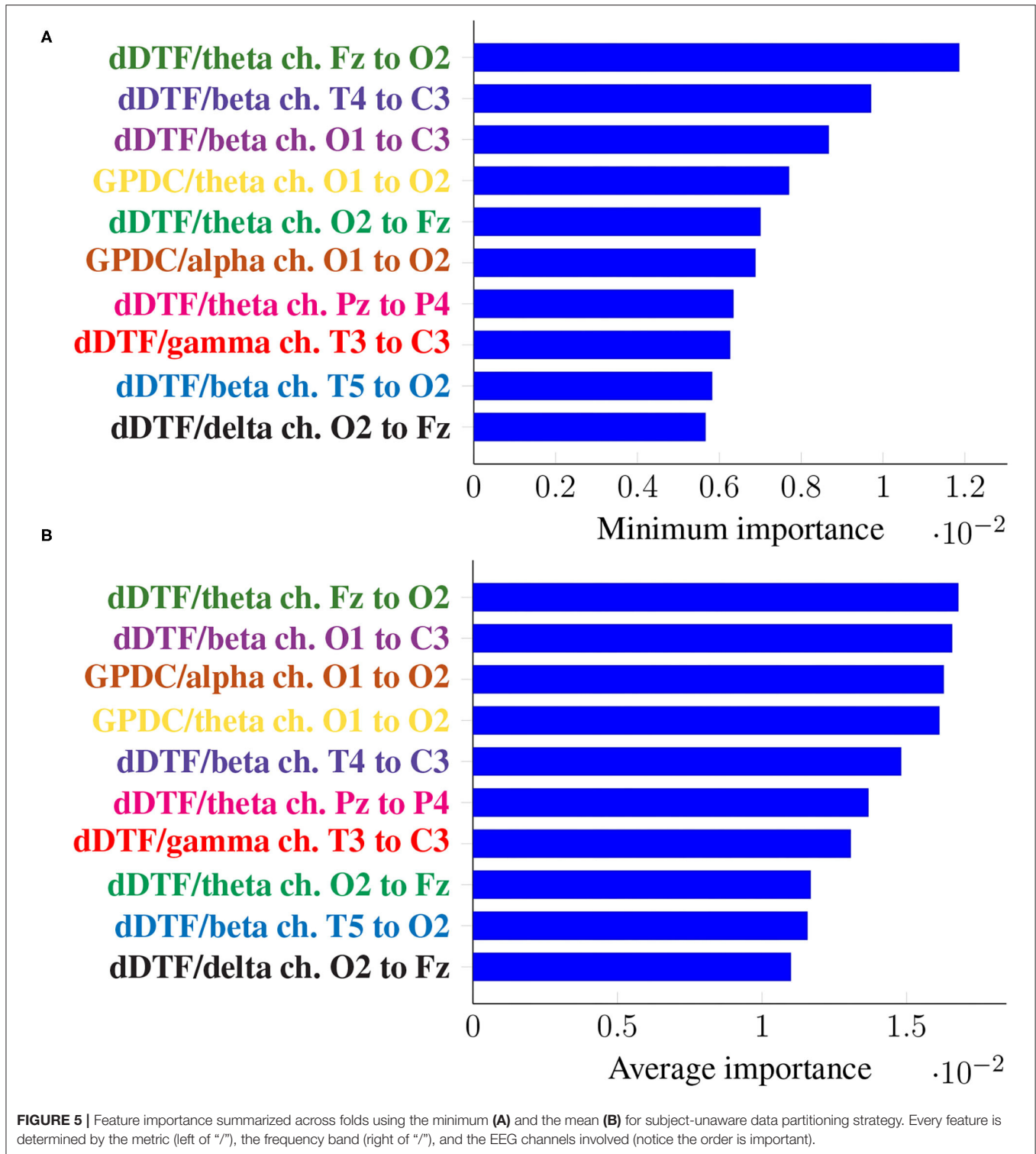
originating every sample. The motivation behind this is that, in a real-world scenario, we usually want to exploit the classifier in deciding whether or not a *new* subject (never seen before) is or not affected by the disease. In order to emulate this scenario in our evaluation of the model we can use a leave-p-subjects-out strategy, which dictates that the classifier must be validated in subjects that are not part of the training set. This is again implemented with a 7-fold cross-validation, but in this case at the subject level: every fold comprises the samples of 2 healthy subjects and 2 patients. Thus, the model is trained each time on 24 subjects (12 healthy ones and 12 patients), and evaluated on 4 *different* ones.

Figure 6 shows the ROC curve for each one of the $K = 7$ folds.

In this case we have more variance across the different folds, and the AUC for two of them (#3 and #5) is significantly worse⁵. Nevertheless, the AUC is still above 0.95 for the rest of the folds, and the average is around 0.87. The decrease in the performance obtained when using this last strategy suggests there is significant subject-to-subject variation. This issue might be addressed by way of a user-specific calibration of the classifier, as it is sometimes done in brain-computer interface systems [see, for instance, Wilson et al. (2009)]. However, such approach falls out of the scope of this work.

As before, it is of interest to identify features that are consistently relevant for classification. **Figure 7** summarizes

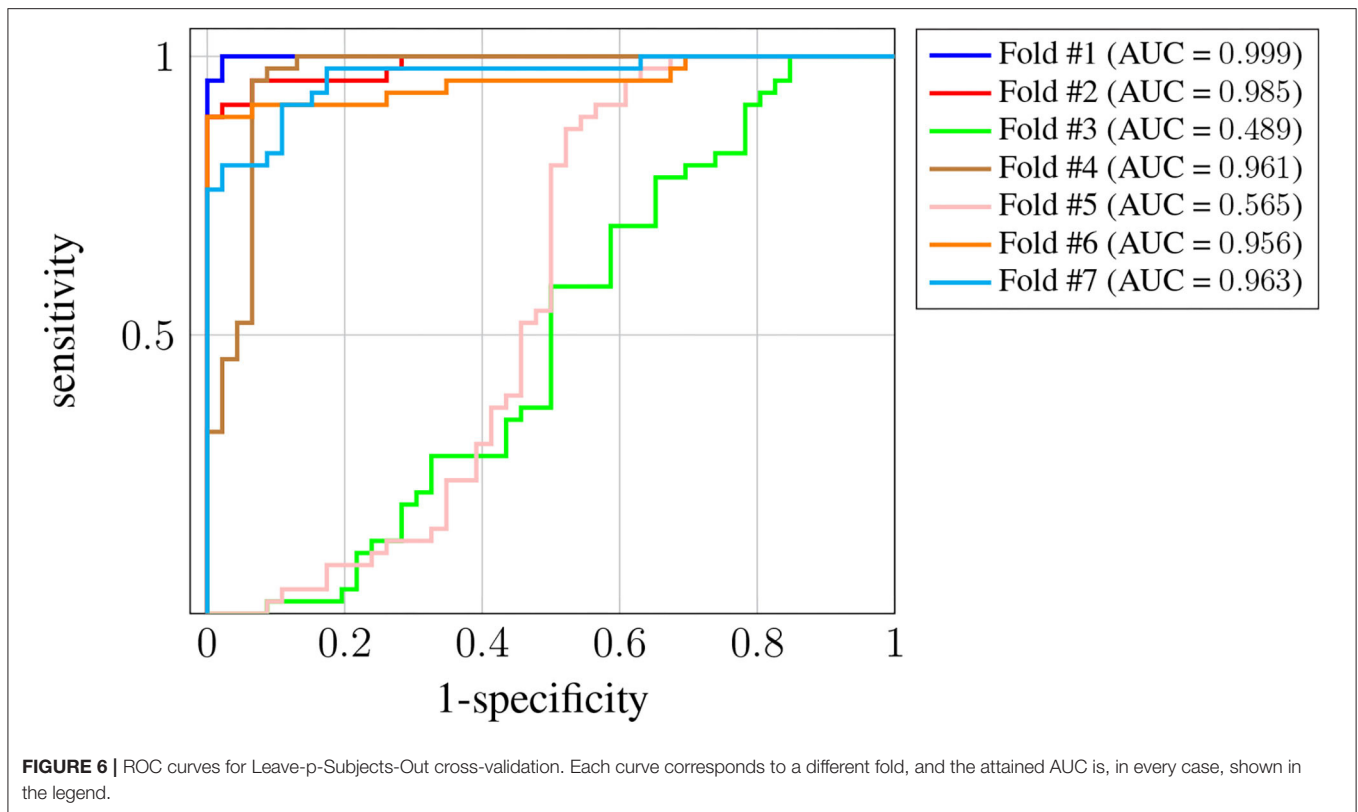
⁵ There is now a trade-off between specificity and sensitivity, but it is still possible to get a useful classifier.



feature importance in the same way we did in the previous section. In this case, the two rankings fail to exhibit the exact same features, but many features with large minimum importance across folds also attain a large average importance. Furthermore, many of the features that were

relevant in the previous section are still relevant here (for easy matching, colors are consistent in both Figures 5, 7). Features that are unique to a single panel are grayed out.

Notice that in both this section and Section 4.1, evaluation is carried out using a 7-fold cross-validation strategy, entailing a *test*



set that encompasses $100/7 \approx 14\%$ of the samples, which seems a sensible choice given the relatively small size of the dataset.

5. DISCUSSION

In this work we have tackled the problem of assessing whether a subject suffers from schizophrenia or not by analyzing their recorded EEG signals. We compute effective connectivity measures on the latter that become the input features of a random forest. This ML technique allows to interpret the results of the classifier by identifying those features that are most relevant for performance. Thus, we have selected seven features (highlighted in color in **Figure 7**) that in our analysis seem to play a role in telling apart patients from healthy subjects.

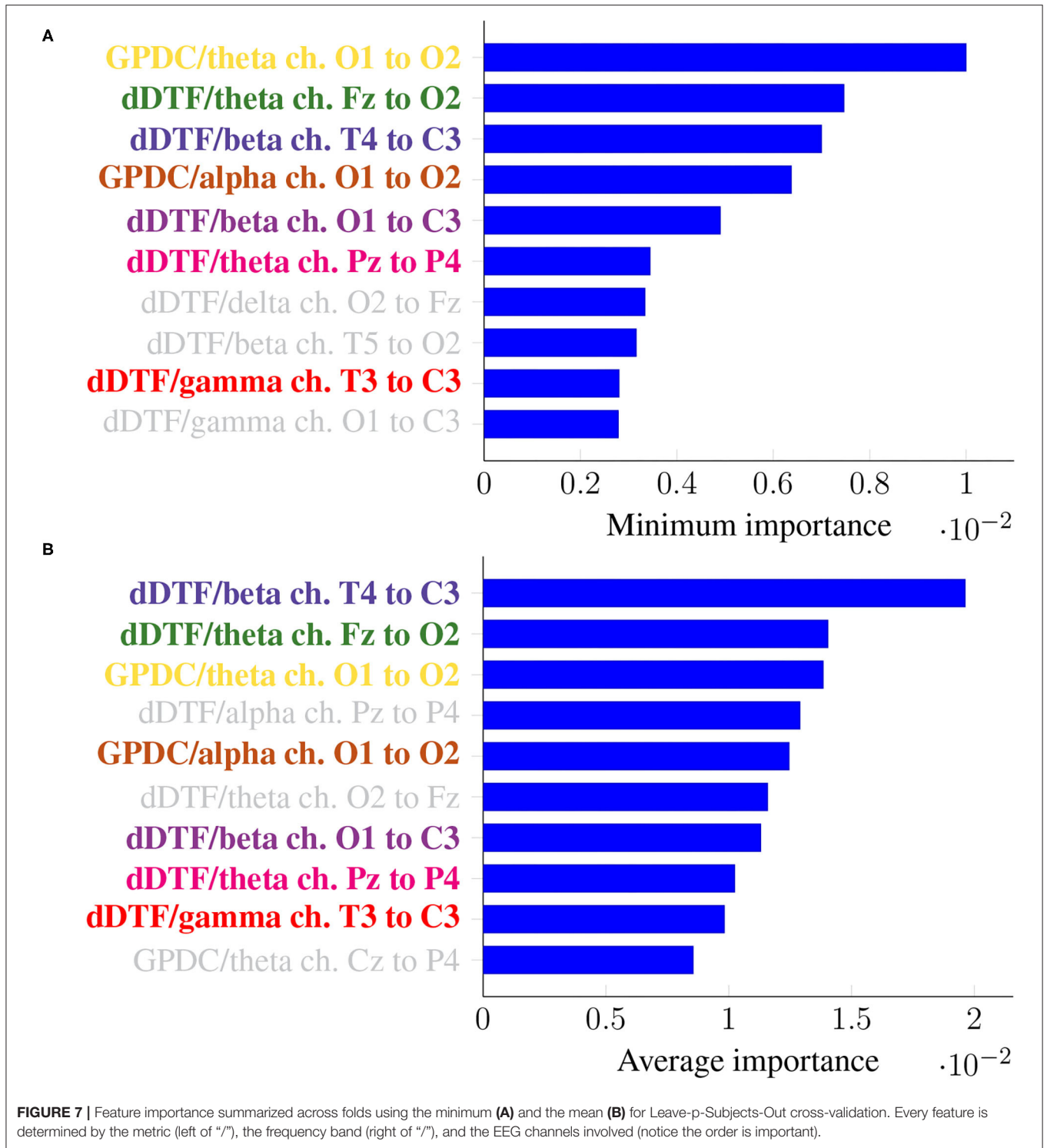
Attached to every selected feature is a connectivity measure (either GPDC or dDTF), a frequency band, and a pair of channels that are causally related. Hence, a method like the one proposed in this work can help clinicians locate key areas and/or connections in the brain that are related to schizophrenia. For instance, our results suggest that signals O1 and O2 are important, specially in the theta and alpha bands. These two signals are associated with the occipital lobe region, and a link between the latter and schizophrenia has already been established before in the literature (see e.g., Onitsuka et al., 2007; Tohid et al., 2015). At the sight of **Figure 7**, and given that signal C3 shows as the *sink* in three different features, the central lobe of the brain also seems to play a prominent role in the disease. Some other conclusions can be drawn from the same figure. However, and due to the relatively limited sample size, we reckon this is a

pilot study and further research, with a larger sample size, would be needed to validate and, afterwards, properly interpret the preliminary results reported here. When comparing our results with those in Olejarczyk and Jernajczyk (2017) (where the dataset was originally studied) there are some similarities. Although the work in the aforementioned paper is based on an entirely different method (relying on graph analysis), it also evinces, for instance, the importance of the occipital area in the alpha band.

Regarding the frequency bands, there are many studies supporting the influence of the alpha band, as well as delta and theta. A thorough review of many of them is carried out in Newson and Thiagarajan (2019). The authors of the latter claim that findings regarding schizophrenia are mostly coherent, though there are some inconsistencies. Therefore, this is still an open problem that should be further pursued.

With respect to the evaluation of the method, we remark that performance can be very different depending on whether or not the classifier is evaluated on samples originating from subjects that have been seen during training. If we guarantee that a few samples from each subject are included in the training set, then the average AUC is 0.99, whereas if subjects in the training and test sets are different, that number decreases down to 0.87. Nevertheless, in this work the focus is on the interpretability of the decisions, and we have found that, in any case, some common conclusions can be drawn.

As noted in Section 1, there are other mental disorders that affect EEG in a similar way as schizophrenia does, and hence can be easily confused with it. A relevant line of



future research is to address the question of whether the proposed ML approach is useful in separating schizophrenia from other mental diseases producing similar variations in the EEG.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://doi.org/10.18150/repod.0107441>.

AUTHOR CONTRIBUTIONS

MV and IM conceived the idea. AM performed data pre-processing. MV designed and performed the data analysis. All authors contributed to writing the manuscript.

FUNDING

We acknowledge support by the Agencia Estatal de Investigación of Spain (CAIMAN, reference TEC2017-86921-C2-1-R

REFERENCES

- Acar, E., Levin-Schwartz, Y., Calhoun, V. D., and Adali, T. (2017). "Tensor-based fusion of EEG and fMRI to understand neurological changes in schizophrenia," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)* (IEEE), 1–4.
- Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adeli, H., and Subha, D. P. (2018). Automated EEG-based screening of depression using deep convolutional neural network. *Comput. Methods Progr. Biomed.* 161, 103–113. doi: 10.1016/j.cmpb.2018.04.012
- Anier, A., Lipping, T., Ferenets, R., Puumala, P., Sonkajarvi, E., Ratsep, I., and V. J. (2012). Relationship between approximate entropy and visual inspection of irregularity in the EEG signal, a comparison with spectral entropy. *Br. J. Anaesthesia* 109, 928–934. doi: 10.1093/bja/aes312
- Asadzadeh, S., Rezaei, T., Beheshti, S., Delpak, A., and S. M. (2020). A systematic review of EEG source localization techniques and their applications on diagnosis of brain abnormalities. *J. Neurosci. Methods* 339, 1–21. doi: 10.1016/j.jneumeth.2020.108740
- Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Baccala, L. A., de Vico Fallani, F., et al. (2007). Comparison of different cortical connectivity estimators for high-resolution EEG recordings. *Hum. Brain Mapp.* 28, 143–157. doi: 10.1002/hbm.20263
- Baccalá, L. A., and Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.* 84, 463–474. doi: 10.1007/PL00007990
- Baccala, L. A., Sameshima, K., and Takahashi, D. (2007). "Generalized partial directed coherence," in *2007 15th International Conference on Digital Signal Processing (IEEE)* 163–166.
- Breiman, L. (1996). Bagging predictors. *Mach. Learn.* 24, 123–140. doi: 10.1007/BF00058655
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. Boca Raton: Chapman & Hall/CRC.
- Chandran, A. N., Sreekumar, K., and Subha, D. (2021). "EEG-based automated detection of schizophrenia using long short-term memory (LSTM) network," in *Advances in Machine Learning and Computational Intelligence* (Singapore: Springer), 229–236.
- Dai, D., Wang, J., Hua, J., and He, H. (2012). Classification of ADHD children through multimodal magnetic resonance imaging. *Front. Syst. Neurosci.* 6:63. doi: 10.3389/fnsys.2012.00063
- de Pierrefeu, A., Löfstedt, T., Laidi, C., Hadj-Seleem, F., Leboyer, M., Ciuciu, P., et al. (2018). "Interpretable and stable prediction of schizophrenia on a large multisite dataset using machine learning with structured sparsity," in *2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI)* (Singapore: IEEE), 1–4.
- Durongbhan, P., Zhao, Y., Chen, L., Zis, P., De Marco, M., Unwin, Z., et al. (2019). A dementia classification framework using frequency and time-frequency features based on EEG signals. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 826–835. doi: 10.1109/TNSRE.2019.2909100
- Geweke, J. F. (1984). Measures of conditional linear dependence and feedback between time series. *J. Am. Stat. Assoc.* 79, 907–915. doi: 10.1080/01621459.1984.10477110

and CLARA, reference RTI2018-099655-B-I00) and by the grant of the Ministry of Education and Science of the Russian Federation Agreement No. 074-02-2018-330.

ACKNOWLEDGMENTS

We would like to thank Elena Rodriguez-Vieitez for insightful comments that contributed to improve the overall quality of the manuscript.

- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438. doi: 10.2307/1912791
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.
- Insel, T. (2010). Rethinking schizophrenia. *Nature* 468, 187–193. doi: 10.1038/nature09552
- Kamiński, M., Ding, M., Truccolo, W. A., and Bressler, S. L. (2001). Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol. Cybern.* 85, 145–157. doi: 10.1007/s004220000235
- Lipton, Z., Kale, D., Elkan, C., and Wetzell, R. (2015). Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677*.
- Luo, Y., Wu, G., Qiu, S., Yang, S., Li, W., and Bi, Y. (2020). EEG-based emotion classification using deep neural network and sparse autoencoder. *Front. Syst. Neurosci.* 14:43. doi: 10.3389/fnsys.2020.00043
- Magee, J. F. (1964). Decision trees for decision making. *Harvard Bus. Rev.* 20, 35–48.
- Mariño, I. P., Blyuss, O., Ryan, A., Gentry-Maharaj, A., Timms, J. F., Kalsi, J., et al. (2017). Change-point of multiple biomarkers in women with ovarian cancer. *Biomed. Sig. Proce. Control* 33, 169–177. doi: 10.1016/j.bspc.2016.11.015
- Newson, J. J., and Thiagarajan, T. C. (2019). EEG frequency bands in psychiatric disorders: a review of resting state studies. *Front. Hum. Neurosci.* 12:521. doi: 10.3389/fnhum.2018.00521
- Olejarczyk, E., and Jernajczyk, W. (2017). Graph-based analysis of brain connectivity in schizophrenia. *PLoS ONE* 12:e0188629. doi: 10.1371/journal.pone.0188629
- Onitsuka, T., McCarley, R. W., Kuroki, N., Dickey, C. C., Kubicki, M., Demeo, S. S., et al. (2007). Occipital lobe gray matter volume in male patients with chronic schizophrenia: a quantitative MRI study. *Schizophr. Res.* 92, 197–206. doi: 10.1016/j.schres.2007.01.027
- Ranganathan, S., Nakai, K., and Schonbach, C. (2018). *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*. Elsevier.
- Reiter, J. (2020). "Developing an interpretable schizophrenia deep learning classifier on fMRI and sMRI using a patient-centered DeepSHAP," in *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)* (Montreal: NeurIPS), 1–11.
- Rubinov, M., Knock, S. A., Stam, C. J., Micheloyannis, S., Harris, A. W., and Williams, L. M. (2009). Small-world properties of nonlinear brain activity in schizophrenia. *Hum. Brain Mapp.* 30, 403–416. doi: 10.1002/hbm.20517
- Sabeti, M., Katebi, S. D., Boostani, R., and Price, G. W. (2011). A new approach for EEG signal classification of schizophrenic and control participants. *Exp. Syst. Appl.* 38, 2063–2071. doi: 10.1016/j.eswa.2010.07.145
- Saeedi, A., Saeedi, M., Maghsoudi, A., and Shalhaf, A. (2020a). Major depressive disorder diagnosis based on effective connectivity in EEG signals: a convolutional neural network and long short-term memory approach. *Cogn. Neurodyn.* 15, 1–14. doi: 10.1007/s11571-020-09619-0
- Saeedi, M., Saeedi, A., and Maghsoudi, A. (2020b). Major depressive disorder assessment via enhanced k-nearest neighbor method and EEG signals. *Phys. Eng. Sci. Med.* 43, 1007–1018. doi: 10.1007/s13246-020-00897-w
- Segal, D. L. (2010). Diagnostic and statistical manual of mental disorders (DSM-IV-TR). *Corsini Encycl. Psychol.* 1–3. doi: 10.1002/9780470479216.corpsy0271

- Shalhaf, A., Bagherzadeh, S., and Maghsoudi, A. (2020). Transfer learning with deep convolutional neural network for automated detection of schizophrenia from EEG signals. *Phys. Eng. Sci. Med.* 43, 1229–1239. doi: 10.1007/s13246-020-00925-9
- Thilakvathi, B., Shenbaga Devi, S., Bhanu, K., and Malaippan, M. (2017). EEG signal complexity analysis for schizophrenia during rest and mental activity. *Biomed. Res.* 28, 1–9.
- Tohid, H., Faizan, M., and Faizan, U. (2015). Alterations of the occipital lobe in schizophrenia. *Neurosciences* 20, 213. doi: 10.17712/nsj.2015.3.20140757
- Vázquez, M. A., Mariño, I. P., Blyuss, O., Ryan, A., Gentry-Maharaj, A., Kalsi, J., et al. (2018). A quantitative performance study of two automatic methods for the diagnosis of ovarian cancer. *Biomed. Sig. Proc. Control* 46, 86–93. doi: 10.1016/j.bspc.2018.07.001
- Wilson, J. A., Schalk, G., Walton, L. M., and Williams, J. C. (2009). Using an EEG-based brain-computer interface for virtual cursor movement with BCI2000. *J. Vis. Exp.* e1319. doi: 10.3791/1319

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Vázquez, Maghsoudi and Mariño. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.