Schizophrenia Detection using EEG: A Study on Frequency Relevance

Luís A. Alexandre
NOVA LINCS
Universidade da Beira Interior
Covilhã, Portugal
ORCID: 0000-0002-5133-5025

Włodzisław Duch
Dept. of Informatics, Institute of Engineering and Technology
Faculty of Physics, Astronomy & Informatics, Nicolaus Copernicus University
Toruń, Poland
ORCID: 0000-0001-7882-4729

***Abstract*—One important application of machine learning is the analysis of electroencephalographic recordings (EEG). Such oscillatory signals are noisy, non-stationary, full of artifacts, contain transients, and chaotic transitions between meta-stable states. EEG data rarely consists of recordings obtained for more than 100 patients with some brain disorder. This makes objective, reliable clinical diagnosis of many mental disorders, such as schizophrenia, very difficult. The standard approach splits the frequency of EEG oscillations focusing on five classical bands: delta, theta, alpha, beta, and gamma. In this paper, we investigate three main questions: (i) Are there certain frequencies that will allow for a reliable diagnosis of schizophrenia, and (ii) if they are, is there a connection to what is known about their neural basis? (iii) How long segments of EEG recordings are sufficient for reliable classification? We filter EEG signals, obtaining a set of 64 very narrow 1Hz frequency bands. A genetic algorithm, with a simple k-NN classifier, finds an optimal combination of these bands. The method is sufficiently simple to be used in clinical settings. A public schizophrenia EEG data set, containing 60 seconds of EEG recordings in the resting state, with only 16 electrodes, for 45 adolescent patients and 39 healthy controls, was used for testing. Signal duration of about 30 seconds enables to reach an accuracy of over 96% in 5-fold cross-validation tests. We compare our results to much more sophisticated state-of-the-art methods and discuss insights gained by our analysis into the brain basis of schizophrenia.**

***Index Terms*—Genetic Algorithm, EEG, Schizophrenia, Narrow frequency bands.**

I. Introduction

Objective diagnosis of psychiatric disorders is very difficult. Symptoms of mental disorders are subjective, very diverse, and may be misleading. Non-invasive neuroimaging and neurophysiological methods allow for observations of neurodynamics. Interesting results have been achieved using functional magnetic resonance (fMRI) methods [1] for many diseases, such as autism spectrum disorders, major depression, obsessive-compulsive disorder, and schizophrenia. Abnormal connections between brain regions have been identified, and biomarkers based on their strength are used for diagnosis. Other effects, such as insufficient concentration of neurotransmitters, dysfunctional ion channels due to genetic mutations that affect protein shapes, or other cellular mechanisms, also lead to brain states that are responsible for deviant behavior. fMRI, based on oxygen transport, is too slow to show oscillations that may be observed only at high temporal resolution. Biomarkers based on the analysis of EEG are better suited for large-scale use in clinical practice.

Schizophrenia (SZ) has very heterogeneous symptoms, usually involving time periods with hallucinations, delusions, disorganized thinking, perceptual misinterpretation, and sensorimotor disintegration. Diagnosis is based on observed behavior and rarely is done before 17 years of age. Results of treatment are very hard to predict, as there is no specific relation between biological basis and symptoms. About half a percent of the general population suffers from schizophrenia. It is thus very important to search for objective diagnostic methods that can be applied at a young age. Moreover, the diversity of SZ symptoms may require a battery of different tests for correct diagnosis. It would be ideal to discover useful biomarkers in the EEG recordings during the resting state.

The main contributions of this work are:

- An empirical study of the impact of using different narrow (1Hz) frequency bands for classifying SZ EEG signals;
- An optimization approach to select the best combination of bands to attain the highest classification performance;
- Identification of the specific frequencies that distinguish SZ patients from control subjects more precisely than can be done with the 5 classical EEG bands;
- An analysis of the impact of the signal duration on the classification accuracy;
- Comparison with the state-of-the-art results on a real EEG data set of children diagnosed with schizophrenia;
- A discussion linking the insights gained by our analysis with the neural basis of schizophrenia.

In the next section, we review recent work on the detection of SZ using EEG recordings. Next, our approach is described, followed by a section containing the experimental results. Section 5 contains the final discussion and conclusions.

II. Related Work

Attempts to use EEG for the diagnosis of schizophrenia are quite frequent. An overview of 42 papers on this subject was recently published by Khare, Bajaj, and Acharya [2]. The number of patients in the EEG datasets was between 5 and 62, with a similar number of control healthy subjects. 33 papers used resting state data, and 9 papers were based

on experiments with auditory, visual, and motor activities. Accuracies on small datasets with just 14 SZ cases and the same number of health controls (HC) [3] were quite high, between 89% and 100%. Classification methods were based on k-NN, SVM, convolutional neural networks, ResNet-SVM, AdaBoost, random forest, probabilistic neural networks, and LSTM. Methods used to analyze signals included mean spectral amplitude, spectral power, Hjorth parameters, multivariate empirical mode decomposition, Short-Term Fourier Transform, variance-based features, continuous wavelet transform (CWT), partial least squares (PLS), expectation-maximization-based, principal component analysis (EM-PCA), nonlinear regression, entropy, isometric mapping (Isomap), nonlinear features, L1 norm features obtained with optimal wavelet, dynamic functional connectivity (DFC), complexity measures, Higuchi Fractal Dimension (HFD), Lyapunov exponents, and cyclic group of prime order patterns. In [4], the authors studied microstate features to see if they were suitable for SZ classification. They concluded that microstate features outperformed the 31 types of features tested. However, the data with 14 SZ cases is definitely too small to draw any definite conclusions. A larger data [5] (45 SZ, 39 HC) has been used in 10 papers [2], with reported accuracies between 84% to 100%.

We have found several additional papers that have not been covered in this comparison. In addition to the methods mentioned above, other sophisticated methods have been used: vector-autoregression model-based directed connectivity, graph-theoretical complex network, ϵ-complexity of continuous vector functions of original EEG signals and their finite differences. Recurrence quantification analysis (RQA) combined with the Short-Term Fourier Transform (STFT) to generate power spectra time series from sliding, overlapping windows was used [6]. Feature selection and linear SVM classification allowed for the discovery of fewer than 10 electrodes/bands, giving a 95.9% accuracy. In a recent paper on the same data [7], parameters of up to 20 microstates were used, achieving 86–100% accuracy with linear SVM. Dimitriadis [8] presented an approach based on the dynamic functional connectivity graph estimated using the imaginary part of phase lag value and correlation of the signal envelope. Data was cleaned using wavelet transform and ICA, but no details on this process were given. Multi-kernel SVM classifier gave 75-100% accuracy on this data set, depending on the features extracted. This is an interesting but quite complex method.

The authors of [9] used an auditory oddball task for detecting SZ in a data set with 68 subjects, 34 healthy control (HC), and 34 SZ. They evaluated both sensor-level (124) and source-level (314) features and concluded that the best results of 88.2% accuracy were obtained when using both types of features. Phang et al. [10] used both hand-crafted and learned features focused on connectivity and other network measures, together with a fusion of three convolutional neural networks (CNNs). They achieved 91.7% accuracy on the data set presented in [5] (more details on this data set will be given below since it is the one we will be using in this work). Using the same data, [11] achieved 95% accuracy using pairwise distance learning and a Siamese neural network to produce features that are then classified by an XGBoost classifier. In [12], the authors used a random forest and focused their work on Event-Related Potentials (ERP). They used the data set produced by the National Institute of Mental Health (49 SZ and 32 HC) and obtained a 96.4% balanced accuracy using a random forest classifier and 10-fold cross-validation. In [13], a hybrid deep neural network to detect SZ in EEG is proposed. The authors evaluated fuzzy entropy and the fast Fourier transform as features, and concluded that fuzzy entropy is a better approach for feature extraction, achieving 99.22% accuracy in a data set that contained 109 patients (54 SZ and 55 HC). The work in [14] focused on the detection of SZ through a passive listening task. The data used consisted of 63 HC and 65 SZ subjects, and both CNNs and random forests were evaluated. The results include 78% accuracy using only five EEG channels and 80% after an ensemble of five CNNs.

Models developed on a small dataset may be useful if they discover features that can serve as biomarkers. In all cases described above, no search for an optimal combination of frequencies that could improve the diagnosis was conducted. Peak EEG frequencies may help to distinguish various brain disorders. The work of [15], focusing on epilepsy detection in EEG and not SZ, did consider 0.5Hz bands, but no optimization process was done to choose the best band combination; the usefulness of each band was evaluated, and all bands that were above a certain threshold were considered in the final setup. The same authors, in [16], looked into 0.5Hz EEG bands up to 100Hz in a study to detect SZ in a data set with 14 SZ and 14 HC subjects, where they achieved a balanced accuracy of 96.8% using a random forest classifier. The authors used all bands, whereas we consider this as an optimization problem and use a genetic algorithm (GA) to find an optimal band combination.

Therefore, it is worth checking if a combination of several characteristic frequencies may provide robust EEG biomarker.

III. Proposed Approach

A. The Data Set and its Representation

To test our approach, we have selected one of the largest EEG data sets that have already been analyzed in more than 10 publications, described in [5]. It consists of 60-second recordings of 84 adolescent boys, 39 healthy control subjects (HC), and 45 boys with diagnosed schizophrenia (SZ), recorded in the resting state, with closed eyes. It was acquired with 16 electrodes placed according to the international 10-20 system, at a 128Hz sampling rate. According to the Nyquist–Shannon sampling theorem, with this resolution, we can analyze frequencies up to 64Hz. This kind of data is relatively easy to obtain in clinical settings, with inexpensive equipment, and short measurement procedures. Methods that can provide reliable diagnosis based on such data may have real clinical value.

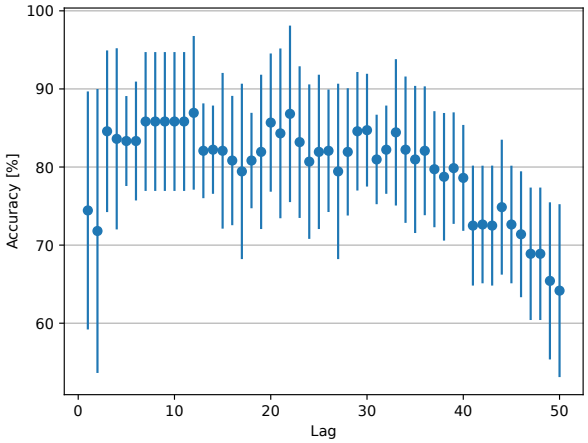

Fig. 1. Accuracy (and std) using all 1Hz bands with different VAR lags, obtained with 10-fold CV.

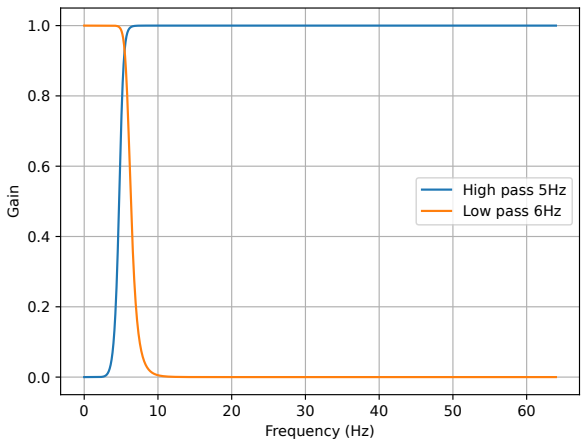

Fig. 2. Filter response for 1Hz band, in this case, between 5 and 6Hz.

EEG time series for each case is represented by a $n \times m$ matrix, for $n = 16$ channels and $m = 128 \times 60 = 7680$ samples. We chose not to do any pre-processing other than fitting a Vector AutoRegressive (VAR) model [17] to each subject's data. Consider that at each time step i, the data from the EEG consists of a random vector y_i, of size n. We then adjust a model to estimate the vector at time t considering a linear combination of the previous L (lag) vectors:

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

where u_t is a random noise vector with zero mean. Vector ν and the $n \times n$ matrices A_i contain parameters that are obtained through a least-mean squares approach. To determine the value of the lag parameter, L, we tested values from 1 to 50, using all 64 one Hz bands simultaneously, and 10-fold cross-validation. The results are in Figure 1. We noticed a region with L between 7 and 12 with high VAR accuracy, and selected lag 10, as it is in the middle of this region.

Since the data set was acquired using $n = 16$ channels and the samples were recorded for 60 seconds at a sampling rate of 128Hz, each recording was originally represented by a 16×7680 matrix. The data representation of each subject after the VAR model estimation is a vector with 2576 features, that contains ten 16×16 A_i matrices, one for each lag, along with the 16 values of vector ν.

Note that the VAR model representation is created *after* the selection of the appropriate (combination of) bands (see below).

B. Single Band Classification

We wish to study three main questions: (i) are there preferred frequencies that facilitate the detection of schizophrenia, and (ii) if they exist, what is their relation to the known physiological characteristics of this disease; finally, (iii) what is the impact of the signal duration on the detection accuracy.

To answer the first question, we split the frequency range into 1Hz bands and investigated the classification accuracy using each band separately. The number of 1Hz bands considered is 64, and the range starts at 0Hz (in fact, 0.1Hz), so the full range of analyzed frequencies varies from 0 to 64Hz. The width of Fourier peaks in the EEG time series is about 2-3Hz, so using 1Hz bands can represent the power of the signal peaks with sufficient precision. The 1Hz band filtering was made using an order 10 Butterworth digital filter for a high pass at frequency i and a low pass at frequency $i + 1$. An example filter response is shown in Figure 2, for a filter between 5 and 6 Hz. A more costly alternative is to perform Fourier transform shifting the time window on one sample.

Current deep-learning approaches usually involve complex architectures with many design decisions and many parameters that typically require large data sets for training. In the case of small EEG datasets, when we are limited to a total of only 84 cases, it is not a good approach. Hence, we chose to use a simple k-nearest neighbors (k-NN) classifier and fixed $k = 3$. The reason for choosing this classifier has to do with the fact that it does not need to be trained, and for small data sets it is very fast. A fast classifier is necessary for the optimization of subsets of narrow bands. The value of k was chosen by considering that $k = 1$ might be too sensitive to outliers, $k = 2$ would allow for draws because this is a two-class problem, so the next value would be a good compromise between the size of the neighborhood to search and the size of the data set. Other odd small values might be as good or better than 3, but we did not focus on optimizing this parameter, as this smallest acceptable value proved to be sufficient.

The first experiment looked at the results of the classification for single 1Hz frequency bands. We run this for all 64 bands, and the results are presented in Figure 3. The dashed line represents the accuracy using all bands (82.0%). Interestingly, using only the band that contains frequencies between 13 and 14Hz (low beta) outperforms this baseline

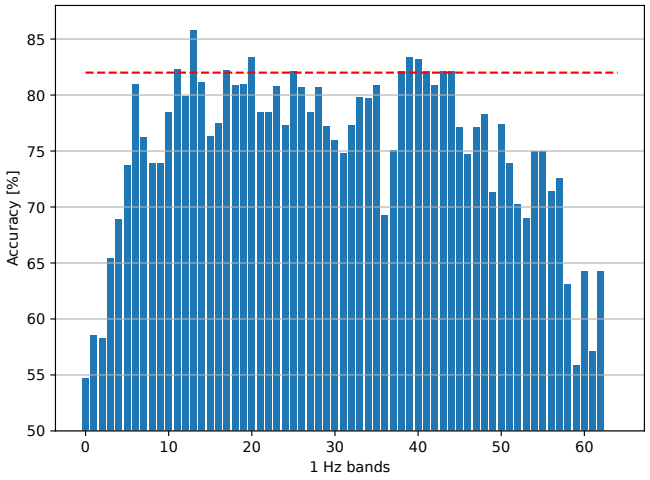

Fig. 3. Accuracy when using one single 1Hz band at a time, for all 64 studied bands. The dashed red line shows the accuracy when using all bands simultaneously (82.0%).

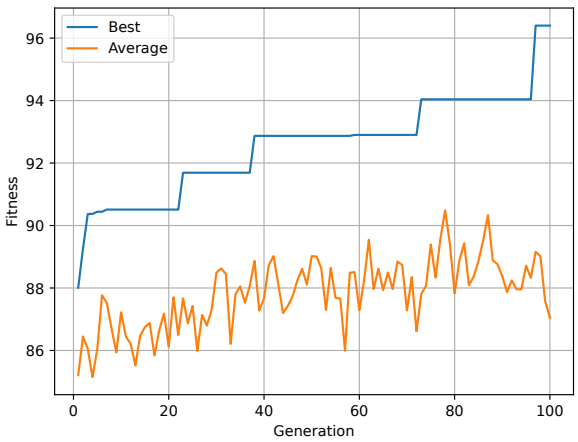

Fig. 4. The fitness of the best chromosome, as well as the average fitness of each generation's population, during the optimization process, for one of the 5 repetitions.

value (85.8%). This is the simplest diagnostic method with quite high accuracy. Some other frequencies also show performance above the baseline. For the 16Hz band accuracy drops by almost 10%. This shows that some frequencies should be excluded, and others added to the optimal subset. The worst classification results were obtained with the lowest and the highest frequency bands. The best single-band results are on the border of alpha and beta bands (14Hz), around 20Hz, and between 38-44Hz.

C. Multiple Band Classification

To find a subset of several 1Hz bands that could be more useful for the detection of schizophrenia, we need to solve an optimization problem, perform a search in the space of all possible combinations of 1Hz bands, and identify a combination that enables a high classification accuracy. To explore the possible frequency band combinations, we have used a Genetic Algorithm (GA). We considered the 64 1Hz bands and represented a solution to this problem as a binary chromosome with a length 64. If a particular position is 1, then that band will be used to represent the data. The size of the search space is then 2^{64}, which makes it unfeasible for exploration with a brute-force exhaustive search.

The fitness function that was evaluated during the GA search was the classification accuracy on a 5-fold cross-validation using the 3-NN classifier. The mutation rate was 0.05, the selection rate (percentage of the population selected for mating) was 0.5, the selection strategy was the roulette wheel, and the number of crossover points was 1. We used a population size of 10 chromosomes and ran the optimization for 100 generations. We repeated the optimization 5 times and presented the average and standard deviation of the accuracy of the best chromosome. Figure 4 shows the evolution of the fitness of the best chromosome and of the average population during the optimization process, for one repetition of the process.

The result of the optimization produced a combination of 1Hz bands that enabled an average 95.2% classification accuracy (and standard deviation of 0.8%) over 5 repetitions of the search process, which is much higher than the result obtained when using all frequencies (82.0% accuracy). The best accuracy on an individual repetition of the experiments yielded 96.4% with 2.9% standard deviation (from the 5-fold CV), with around 30 frequency bands.

D. Relevance of Segment Duration

To study how the size of samples impacts the ability to detect the disease, we run the optimization process with samples of smaller duration. Instead of the full 60 seconds (7680 samples) for each subject, we have used 30 sec (3840 samples), 15 sec (1920 samples), 7.5 sec (960), and 3.75 sec (480 samples). We could have increased the number of data points per subject; for instance, when using 3840 samples, we could have had two such segments per subject. This would require care not to include samples from the same subject in both training and test sets, and we would also have more data points per subject than in the original representation. Instead, we decided to retain only one shorter signal sequence per each subject, with the new sample size. We can define how to choose the used samples from the original ones in two ways: first, simply use samples from the beginning of the recording and discard the remaining data. The second approach was to randomly choose, at each cross-validation fold, between the existing samples, which sample to use. Figure 5 shows the results of both procedures. Deterministic choice (in this case, starting from the initial sample) produces better results than random choice at every cross-validation fold. For small sample durations, differences in accuracy are large, but already for samples with half of the maximum duration (30 sec),

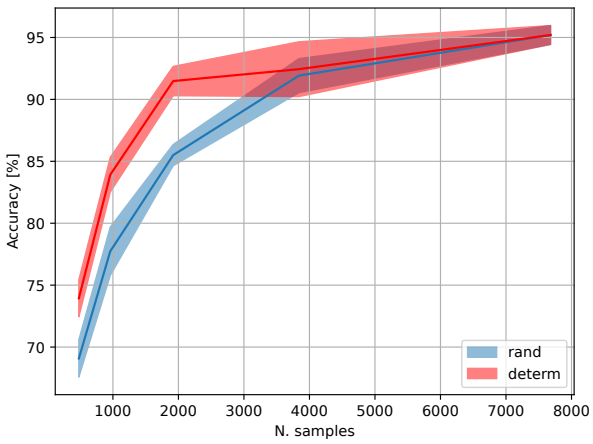

Fig. 5. Accuracy and standard deviation of 5-fold CV using sample sizes 480, 960, 1920, 3840 and 7680. In red are the results when choosing the segments deterministically and in blue when the choice was random, at each CV fold.

average values of accuracy are almost identical between the two approaches.

E. Experimental Details

The experiments were run 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 optimization process took around 9 hours to execute for each run. The code was made in Python 3.10, and we used the Scikit-learn [18] libraries for the classifier, Statsmodels [19] for the VAR model, and Matplotlib [20] for the figures. We adapted the code from the GeneAl Genetic Algorithm library [21] to our needs.

IV. Discussion

A. The Physiology of Schizophrenia

According to the DSM-5 [22], the diagnosis of schizophrenia is based on subjectively reported experiences, symptoms, and observed behavior. It can be misdiagnosed with many other mental disorders which have similar clinical manifestations. Symptoms may be positive (similar to psychosis), negative (deficits of normal emotional responses), and cognitive (social cognition, perception, memory). We should expect that schizophrenic boys exhibit symptoms that may be reflected in the neurodynamics of their brains. However, symptoms are often transient, so EEG data sets may not show clear signs of unusual processes. Complex genetic, neurophysiological, and environmental interactions are implicated in schizophrenia etiology, with dysfunctions in the regulation of dopamine, glutamate, serotonin, and nicotinic/acetylcholine systems, inflammation, and hormonal abnormalities. Review of structural and functional magnetic resonance imaging results by Mubarik and Tohid [23] was focused on the frontal lobe, temporal lobe, and limbic system, noting changes in ”the left and right inferior temporal, right supramarginal/superior temporal, right and left inferiorfrontal, left frontopolar, right and left dorsolateral/ventrolateral prefrontal cortices”.

Changes have also been found in the spatial distribution of power [24]: schizophrenic patients with negative symptoms had smaller activity in the left temporal area. Doege et al. [25] mention a link between P300 reduction in schizophrenia and reductions in delta and theta activity in the auditory oddball task. Such effects can also be noticed in the asymptotic power distributions presented in [7], where decreased left hemisphere activity predominantly in alpha and theta bands is shown. Ellis et al. [26] used a convolutional neural network and perturbation-based approach to find the most important electrodes and bands. They have found that T8 and C3 electrodes with the delta and gamma bands provide the most distinctive information.

It is quite likely that schizophrenia and other mental disorders are the result of problems with synchrony between different brain areas. Temporo-spatial processing disorders (TSPD) are responsible for various cognitive and affective disabilities [27], resulting from abnormal functional connectivity and neuronal synchronization between multiple brain areas. Review of brain oscillations in various brain disorders [28] showed various changes that can be observed in EEG of schizophrenics: decrease of gamma and delta activity, and frequency shifts in alpha and lower frequencies. However, there are also surprising observations, such as an increase, greater than in the control subjects, of gamma activity during cognitive loading. The index of structural synchrony, calculated by counting the coincidences of boundaries of quasi-stationary alpha band segments, is 2.03 for the healthy group, and only 1.67 for the schizophrenia group [5]. Changes in the oscillations of local cortical neural ensembles and weaker interdependence between different regions of the brain of schizophrenics were also confirmed using another dataset [29]. The largest connectivity group differences were in the alpha band, reduced connectivity strength of SZ patients, and changes in graph connectivity measures. Such changes in the connectivity should also result in differences in the power of oscillations in different brain areas.

B. Connection between Physiology and Selected 1Hz Bands

To understand if there is a connection between the bands identified as more useful in the GA search results and what is known about schizophrenia physiology, we first relate the 1Hz bands to the 5 classical EEG bands. We calculate the percentage of usage (PU) of each classical band by a particular chromosome, defined as the ratio between the number of 1Hz bands used by a chromosome and the total number of 1Hz bands in that EEG band. For instance, if a chromosome is using 3 of the 4 available 1Hz bands inside the delta band, then the PU of this band by this chromosome is 75%.

We calculated the average PU for the 5 EEG bands, by the best 10, 20, 30, and 40 chromosomes taken from all generations and repetitions and also by the worst 10, 20, 30, and 40 chromosomes, again from all generations and repetitions. These values are presented in Fig. 6. The worst

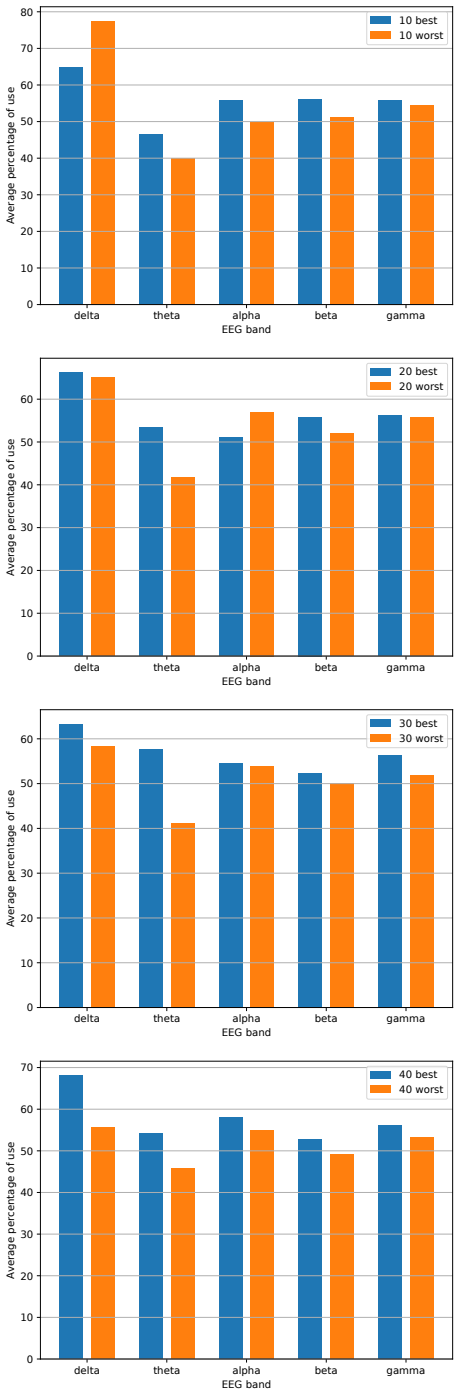

Fig. 6. Average percentage of usage of the five classical EEG bands by the 10, 20, 30, and 40 best and 10, 20, 30, and 40 worst chromosomes.

chromosomes use frequencies from each band almost as often as the best ones. Frequencies that belong to all classical bands are important, with the low-frequency delta band used more often than the remaining ones. For 10 chromosomes the delta frequencies in the best chromosomes are not used as much as in the worst ones.

Figure 7 shows the 40 best and 40 worst chromosomes

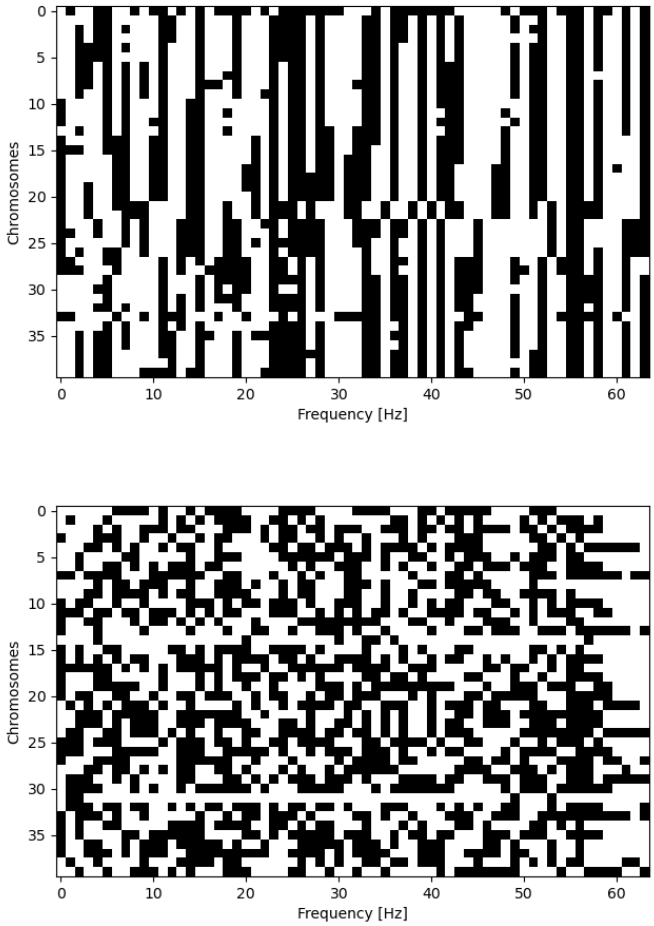

Fig. 7. Best 40 (top) and worst 40 (bottom) chromosomes. Each chromosome is a row. White represents the use of a particular 1Hz band. Note that the 64 columns represent the 64 1Hz bands encoded in the chromosomes.

found (each chromosome is a row in this figure). Some frequencies have not been selected by any chromosomes, while others were quite popular among the best chromosomes that have evolved. Vertical lines show how similar to each other the chromosomes are. The worst chromosomes are randomly initialized and do not show much similarity, whereas the best chromosomes have evolved to achieve good performance and show the tendency to use the same frequency bands. With a large input space and comparable contributions from single bands, as seen in Fig. 3, many combinations of bands may give similar accuracies. Fig. 8 shows the most used 1Hz bands in the top 40 chromosomes. Note that 35 out of the 40 best ones are using 8, 16, 30, 35, 38, 40, 46, 50, 54, 59 and 60 Hz bands. The region between 23 and 28Hz is probably the least used. Also, none of these chromosomes used either the 55 or 56Hz bands. This shows that information useful for distinguishing between the schizophrenic and the healthy controls can be obtained from almost the entire 64Hz spectrum considered.

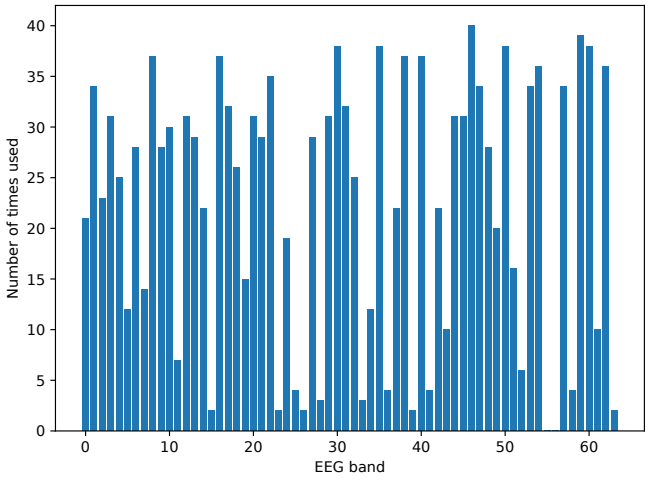

Fig. 8. Number of times a given 1Hz band was used in the best 40 chromosomes.

V. Conclusions

In this paper we introduced an optimization approach to study the classification accuracy for detecting schizophrenia in EEG signals, using combinations of selected 1Hz bands. One goal was to find which frequencies are sufficient to predict schizophrenia. Using all frequencies, we achieved 82% accuracy, whereas the best mix of about 1Hz bands achieved 96.4% accuracy. Another goal was to find what relation these 1Hz bands might have with the physiological knowledge of the disease. The bands used by the best chromosomes are a mix from all over the 64Hz spectrum considered and hence make it hard to see a relation with already identified specific bands that can be used for schizophrenia detection. Nonetheless, we did observe that certain groups of frequencies were not being used and still we could attain a high classification accuracy, pointing to a marginal importance of these bands in the detection of this pathology. We also wished to investigate the influence of the sample size in the ability to detect the disease using our method: the results show that the accuracy drops quickly for samples under 30 seconds duration, when using our approach, making this an important guideline for future research based on EEG schizophrenia detection. Although the optimization step is computationally demanding the whole method is extremely simple and after selecting a subset of the most useful frequencies can be used in practice.

A similar optimization approach can be used to select combinations of specific frequencies and the positions of electrodes. An analysis of recordings made with higher-density EEG equipment to make more precise predictions would be needed. This could be the way to discover specific EEG biomarkers that could be measured using simpler equipment with judiciously positioned electrodes and a small number of frequency filters. In future work, we plan on applying this process to other schizophrenia EEG data sets and also to other brain disorders that can be diagnosed using EEG.

References

[1] N. Yahata, K. Kasai, and M. Kawato, "Computational neuroscience approach to biomarkers and treatments for mental disorders," *Psychiatry and Clinical Neurosciences*, vol. 71, no. 4, pp. 215–237, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/pcn.12502

[2] S. K. Khare, V. Bajaj, and U. R. Acharya, "Schizonet: a robust and accurate margenau–hill time-frequency distribution based deep neural network model for schizophrenia detection using eeg signals," *Physiological Measurement*, vol. 44, no. 3, p. 035005, Mar. 2023.

[3] E. Olejarczyk and W. Jernajczyk, "Graph-based analysis of brain connectivity in schizophrenia," *PLoS ONE*, vol. 12, no. 11, 2017.

[4] K. Kim, N. T. Duc, M. Choi, and B. Lee, "Eeg microstate features for schizophrenia classification," *PloS one*, vol. 16, no. 5, p. e0251842, 2021. [Online]. Available: https://europepmc.org/articles/PMC8121321

[5] S. Borisov, A. Kaplan, N. Gorbachevskaya, and I. Kozlova, "Analysis of EEG structural synchrony in adolescents with schizophrenic disorders," *Human Physiology*, vol. 31, no. 3, 2005.

[6] Ł. Furman, W. Duch, L. Minati, and K. Tołpa, "Short-time fourier transform and embedding method for recurrence quantification analysis of eeg time series," *The European Physical Journal Special Topics*, vol. 232, no. 1, p. 135–149, Feb. 2023.

[7] W. Duch, K. Tołpa, E. Ratajczak, M. Hajnowski, Ł. Furman, and L. A. Alexandre, "Asymptotic spatiotemporal averaging of the power of eeg signals for schizophrenia diagnostics," in *Neural Information Processing*, ser. Communications in Computer and Information Science, B. Luo, L. Cheng, Z.-G. Wu, H. Li, and C. Li, Eds. Singapore: Springer Nature, 2024, p. 428–439.

[8] S. I. Dimitriadis, "Reconfiguration of amplitude driven dominant coupling modes (DoCM) mediated by alpha-band in adolescents with schizophrenia spectrum disorders," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 108, p. 110073, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278584620303894

[9] M. Shim, H.-J. Hwang, D.-W. Kim, S.-H. Lee, and C.-H. Im, "Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features," *Schizophrenia Research*, vol. 176, no. 2, pp. 314–319, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0920996416302274

[10] C.-R. Phang, F. Noman, H. Hussain, C.-M. Ting, and H. Ombao, "A multi-domain connectome convolutional neural network for identifying schizophrenia from EEG connectivity patterns," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, pp. 1333–1343, 2020.

[11] D. Calhas, E. Romero, and R. Henriques, "On the use of pairwise distance learning for brain signal classification with limited observations," *Artificial Intelligence in Medicine*, vol. 105, p. 101852, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365719311261

[12] F. Janek, T. Rieg, and R. Buettner, "Detection of schizophrenia: A machine learning algorithm for potential early detection and prevention based on event-related potentials," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2021.

[13] J. Sun, R. Cao, M. Zhou, W. Hussain, B. Wang, J. Xue, and J. Xiang, "A hybrid deep neural network for classification of schizophrenia using EEG data," *Scientific Reports*, vol. 11, no. 1, p. 4706, Feb 2021. [Online]. Available: https://doi.org/10.1038/s41598-021-83350-6

[14] C. Barros, B. Roach, J. M. Ford, A. P. Pinheiro, and C. A. Silva, "From sound perception to automatic detection of schizophrenia: An eeg-based deep learning approach," *Frontiers in Psychiatry*, vol. 12, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.813460

[15] R. Buettner, J. Frick, and T. Rieg, "High-performance detection of epilepsy in seizure-free EEG recordings: A novel machine learning approach using very specific epileptic eeg sub-bands," in *International Conference on Information Systems*, 2019.

[16] R. Buettner, D. Beil, S. Scholtz, and A. Djemai, "Development of a machine learning based algorithm to accurately detect schizophrenia based on one-minute eeg recordings," in *Proceedings of the 53th Hawaii International Conference on System Sciences*, 2020.

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

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duch-

esnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[19] S. Seabold and J. Perktold, "statsmodels: Econometric and statistical modeling with python," in *9th Python in Science Conference*, 2010.
[20] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
[21] D. M. Chaves, "Geneal – github." [Online]. Available: https://github.com/diogomatoschaves/geneal
[22] A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, 5th ed., 2013.
[23] A. Mubarik and H. Tohid, "Frontal lobe alterations in schizophrenia: a review," *Trends in Psychiatry and Psychotherapy*, vol. 38, no. 4, p. 198–206, 2016.
[24] D.-W. Kim, M. Shim, J.-I. Kim, C.-H. Im, and S.-H. Lee, "Source activation of p300 correlates with negative symptom severity in patients with schizophrenia," *Brain Topography*, vol. 27, no. 2, pp. 307–317, Mar 2014. [Online]. Available: https://doi.org/10.1007/s10548-013-0306-x
[25] K. Doege, A. T. Bates, T. P. White, D. Das, M. P. Boks, and P. F. Liddle, "Reduced event-related low frequency EEG activity in schizophrenia during an auditory oddball task," *Psychophysiology*, vol. 46, no. 3, pp. 566–577, 2009.
[26] C. A. Ellis, A. Sattiraju, R. Miller, and V. Calhoun, "Examining effects of schizophrenia on eeg with explainable deep learning models," in *2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE)*, Nov. 2022, p. 301–304.
[27] B. Gepner and F. Féron, "Autism: a world changing too fast for a mis-wired brain?" *Neuroscience and Biobehavioral Reviews*, vol. 33, no. 8, p. 1227–1242, 2009.
[28] E. Başar, C. Schmiedt-Fehr, B. Mathes, B. Femir, D. D. Emek-Savaş, E. Tülay, D. Tan, A. Düzgün, B. Güntekin, A. Özerdem, G. Yener, and C. Başar-Eroğlu, "What does the broken brain say to the neuroscientist? oscillations and connectivity in schizophrenia, alzheimer's disease, and bipolar disorder," *International Journal of Psychophysiology*, vol. 103, p. 135–148, 2016.
[29] E. Olejarczyk and W. Jernajczyk, "Eeg in schizophrenia," 2017. [Online]. Available: https://repod.icm.edu.pl/dataset.xhtml?persistentId=