



HAL
open science

Sparse control-inspired features generation for Schizophrenia diagnosis

Housseem Meghnoudj, Bogdan Robu, Mazen Alamir

► **To cite this version:**

Housseem Meghnoudj, Bogdan Robu, Mazen Alamir. Sparse control-inspired features generation for Schizophrenia diagnosis. IFAC WC 2023 - 22nd IFAC World Congress, IFAC, Jul 2023, Yokohama, Japan. hal-04051822

HAL Id: hal-04051822

<https://hal.science/hal-04051822>

Submitted on 30 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sparse control-inspired features generation for Schizophrenia diagnosis [★]

H. Meghnoudj ^{*} B. Robu ^{*} M. Alamir ^{*}

^{*} Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000
Grenoble, France

Abstract: In this paper, a recent control-oriented features generation method is used to diagnose schizophrenia using electroencephalogram (EEG) signals. The methodology has already been used for the diagnosis of Parkinson’s disease with competitive results. The method is directly inspired by the functioning of the brain and is mainly based on optimal control theory and sparse optimisation. An appealing feature in the proposed solution is that it allows to combine both frequency and temporal-related aspects of the signal which are known to be detrimental in this context. The proposed solution is evaluated on a publicly available dataset that includes 81 subjects, of which 49 suffer from schizophrenia and 32 are healthy. Results show that by mean of only one extracted feature, fed to a linear discriminant analysis (LDA) classifier, high accuracy separation is obtained. Several validity tests were carried out to assess the statistical relevance of the findings.

Keywords: Dynamical systems, Inverse problem, Schizophrenia diagnosis, Electroencephalogram, Sparse features, Machine Learning

1. INTRODUCTION

Schizophrenia (SZ) is a chronic neuropsychiatric disorder that affects approximately 24 million individuals of the population worldwide (GHDx, 2021). Schizophrenia is characterised by the presence of positive (psychotic) symptoms which are not a dysfunction or an extreme version of normal physiological functioning, but are new features which are unique to SZ. Among these symptoms we can find: delusions, hallucinations, disorganized speech etc. Negative symptoms (e.g. limited speech) and cognitive deficit are other symptom categories of the disorder.

The diagnosis of SZ is until now entirely clinical, and is based on the manifestation of psychotic symptoms as well as the evaluation of the patients’ self-report of their own subjective experiences. The clinical diagnosis is supported by various tools and guidelines such as: the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (APA, 2022) and the International Statistical Classification of Disease (ICD-10) (WHO, 1997). Although clinical diagnosis has been improved with the latest versions of DSM-5 and ICD-10, clinical diagnosis is still limited by human subjectivity (Walsh-Messinger et al., 2019).

Early diagnosis of the disease allows for better control of some symptoms before complications arise, thus improving the long-term outlook of the disease (Insel, 2010). Nevertheless, early diagnosis of the disease is difficult since some symptoms are not exclusive to SZ and even healthy individuals report having psychosis experiences (Linscott and van Os, 2012).

The abundance of data and its ease of access helped in the development of various data-driven methods for the diagnosis of SZ. These methods aim to identify, directly from

the data, new biomarkers of the disease. Different diagnostic directions are investigated, the main ones being: functional Magnetic Resonance Imaging (fMRI), Electromyography (MEG) and Electroencephalography (EEG), each one with its advantages and drawbacks.

EEG have stood out for the diagnosis of SZ especially for its ability to record the dynamics of the brain underpinning sensory, cognitive, affective, and motor processes in response to a stimulus as well as for their high temporal resolution, facility of usage and price. These brain and motor processes are often affected by and directly correlated with the illness, and can even be observed before the onset of the latter (Green et al., 2011).

Nevertheless, EEG signals are known to be very noisy with a low signal-to-noise ratio due to the large amplification required to record the very low amplitude of the brain’s electrical activity. In addition, various unwanted artefacts perturbing the overall signal are present in the recordings such as: ocular, muscular, cardiac, etc (Urigüen and Garcia-Zapirain, 2015). Another encountered difficulty analysing EEGs is volume conduction, i.e. the transmission of electric fields from a primary current source through biological tissue towards the recording electrodes (Olejniczak, 2006). Due to the latter, we lose the ability to target precisely a brain source, and unwanted artefacts will spread and contaminate more electrodes.

The diagnosis of SZ based on EEG has been studied in several works. Khare and Bajaj (2022) propose a robust variational mode decomposition (RVMD) to model and decompose the EEG signal into different modes, and using six features computed on the decomposed signal they classify the SZ from the control group using a optimised extreme learning machine (OELM) to achieve a presumable classification accuracy of 92.9 %. Prabhu and Martis (2020) extract two features corresponding to Kolmogorov

^{*} This work has been partially supported by MIAI@Grenoble Alpes (ANR-19-P3IA-0003).

Complexity and Sample Entropy from the Event Related Potential (ERP) computed at four electrodes locations. The features are separated using a Neural Network (NN) to achieve an accuracy of 91.2 %. Zhang (2019) proposed SZ classification based on the amplitude and latency of N1 and P2 component over the Cz channel combined with non-ERP related information (demographic) during a button-press task to generate a tone. The accuracy was in the order of 81.1 %. Devia et al. (2019) used the mean ERP values computed between 400—600 ms post stimulus during a visual task, these values were then averaged across four regions of interest (occipital, central, frontal, and parietal), then were fed to Linear Discriminant Analysis (LDA) classifier to achieve an accuracy of 71 %. Santos-Mayo et al. (2017) used 16 temporal and 4 frequency features extracted from ERP computed during a 3-oddball auditory task. Combining these features computer over 8 electrodes and by mean of a Multi Layer Perceptron (MLP) they achieved a presumable accuracy of 93.4 %. Other studies exist, however we have not cited them due to space limitation, but the following criticisms may apply to the majority of them.

We strongly believe that the majority of the work we have seen suffers from various problems well known in the machine learning community. The first and most critical issue is data leakage, which is defined as the use of information during the learning phase that is not available during the testing phase. In a real-life scenario, a new, unlabeled sample of data is received for the purpose of categorizing it, thus, the use of any information about that new sample during the learning phase is impossible. This data leakage will bias the evaluation and therefore the claimed performances, the developed model will have excellent performances on the training set but it will have bad performances on new unseen data. Two types of leakage have been noticed: (1) Group leakage, where correlated information from the same patient is found in the training and test set. This bias affects complex architectures (e.g. neural networks) much more, as they can sometimes even detect the signature of a given patient. (2) Consist of optimising hyper-parameters and performing feature selection directly on the test set (absence of a validation set). The other observed problem is the fact of not taking into consideration the data unbalance (numerical majority of a category). This induces a bias on the developed model as it will tend to favour the dominant category, thus biasing the evaluation and increasing false positive or false negative rate.

This article addresses the detection of SZ from EEG signals using a feature generator inspired by the brain’s functioning. The method used is not novel and has already been tested on Parkinson’s disease yielding excellent results (Meghnoudj et al., 2022). It combines the temporal, frequency and dynamic aspects of the signal in order to generate Virtual Modal Stimuli (VMS), which consist of virtual rhythms parsimoniously calculated to fit the signal of interest. Basic features are then extracted from these VMS and fed to an LDA classifier to assess subjects’ condition.

This paper is structured as follows: Section 2 describes the dataset we use as well as the pre-processing we applied to the latter, then presents our brain inspired feature generator and gives details about how to have the least

unbiased evaluation of our approach. In section 3 the results obtained are presented and section 4 concludes the paper.

2. MATERIAL AND METHODS

2.1 Dataset

The data we use in the present article are publicly accessible (Roach, 2021). The study includes 81 subjects, of which 49 suffer from schizophrenia and 32 individuals serve as a control group (CTL). The task that the participants were subjected to was a sensory task involving button pressing and/or an auditory tone. The subjects either: (a) pressed a button to immediately generate a tone, (b) passively listened to the same tone, or (c) pressed a button without generating a tone.

Data used here were recorded over 64 electrodes with a sampling rate of 1024 Hz. All the available data underwent already the following preprocessing before being made available: (1) Re-reference to averaged ear lobes. (2) High-pass filtered at 0.1 Hz. (3) Outlier channels interpolation. (4) Epoching of the continuous signal and creation of 3s time windows centred around the stimulus arrival. (5) Canonical correlation analysis to remove muscle and high frequency white noise. (6) Outlier trial rejection. (7) Independent Component Analysis removing unwanted artefacts. (8) Outlier channels interpolation. The reader can refer to the complete details about the experience here: (Roach, 2021; Ford et al., 2013).

As for the pre-processing we applied, the data were first down-sampled to 128 Hz, then a low-pass filter at 50 Hz was applied and finally the data were re-sliced to form windows starting from -300 ms pre-stimulus to 600 ms post-stimulus.

The ERP of each subject was calculated separately for each condition and channel by vertically averaging all segments corresponding to the same stimulus type and channel (Luck, 2005). The purpose of this step is to filter the signal by summing up the events occurring at the same instant to make them stand out from the ambient white noise.

2.2 Main idea of the method

To process, encode, retrieve and transmit information, biological neuronal networks oscillate (Ward, 2003). The frequency and timing of these oscillations is very important as it is the basis of cognitive processes and it is what allows the synchronization of the oscillation of a group of neurons (Buzsaki and Draguhn, 2004; Ward, 2003). It should be noted that the EEG records mainly the electrical activity of a group of neurons oscillating in synchrony. The synchronization and desynchronization of a group of neurons suggests that the rhythms that contribute to the EEG occur in a pulsatory fashion (Olejniczak, 2006). The ERP associated with a given stimulus can be decomposed into two types of rhythms:

- (1) An ongoing activity: it refers to the oscillatory state of the brain when no stimulus is perceived.
- (2) An evoked activity: it refers to the brain rhythms in response to a stimulus.

2.3 Dynamical feature generation

In order to get closer to the simplified model of brain function and to capture this pulsating nature, we model the ERP through a battery of pendulums, each with its own angular frequency. These pendulums are switched on and off at specific times to best match the ERP signal $y(k)$.

Consider m harmonic oscillators with distinct angular frequency $\{\omega_1, \omega_2, \dots, \omega_m\}$. The system that combines these m decoupled oscillators can be described with the following discrete state-space representation:

$$\Sigma: \begin{cases} x_{k+1} = Ax_k + Bu_k; \\ \hat{y}_k = Cx_k \end{cases}; x_k \in \mathbb{R}^{2m}, u_k \in \mathbb{R}^m, \hat{y}_k \in \mathbb{R} \quad (1)$$

where:

$$A = \text{diag}(a_1, \dots, a_m), B = \text{diag}(b_1, \dots, b_m), C = (c_1 \dots c_m)$$

and the matrices a_i , b_i , and c_i are the matrices that describe the dynamics of a single harmonic oscillator. For a sampling period of $T_s = 1/f_s$ they are defined by:

$$a_i = \begin{pmatrix} 1 & T_s \\ -T_s \omega_i^2 & 1 \end{pmatrix}, b_i = \begin{pmatrix} 0 \\ T_s \end{pmatrix}, c_i = (f_s \omega_i \ 0) \quad (2)$$

With this form, the m oscillators remain decoupled but they are merged through matrices A , B , and C such that their individual contribution is summed to form the signal $\hat{y}(k)$.

Let us note $u_i(k)$ as the excitation force acting on the i^{th} pendulum, so the control input $u(k)$ which is defined by $u(k) = (u_1(k) \ u_2(k) \ \dots \ u_m(k))^T$ contains the excitation forces of all the m pendulums at the instant k . The ω_i term present in the c_i matrix is a scaling term and its role is to make sure that a resting pendulum excited by a force $u_i(k_1) = h$, will start to oscillate at the instant $k = k_1$ with a magnitude h . This gives a physical meaning to the $u(k)$ values since they are directly proportional to the ERP amplitude. Moreover, this allows for a fair comparison of $u_i(k)$ values between them since they are now on the same scale. This opens up the possibility of comparing the modes in terms of energy, signal contribution magnitude, activation time, switching frequency, etc. It should be noted that with this representation we can study the excitation forces of a single mode or a group of modes, over the entire time duration or a defined period.

The system (1) can be rewritten in its explicit form as follows:

$$\hat{y}_k = CA^k x_0 + \sum_{i=0}^{k-1} CA^i B u_{(k-1-i)}, \quad k = 1, 2, \dots \quad (3)$$

where x_0 represents the initial oscillation state of our model (initial position and velocity of each pendulum). If no force is applied to these pendulums as in the case of a non-forced regime ($u(k) = 0$ for $\forall k$), the pendulums will keep swinging in the same manner. Therefore, the information of the ongoing rhythm will be carried by x_0 . For a given ERP signal Y with length L we define:

$$Y = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{L-1} \end{pmatrix}, \hat{Y} = \begin{pmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \vdots \\ \hat{y}_{L-1} \end{pmatrix}, U = \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_{L-2} \end{pmatrix} \in \mathbb{R}^{m(L-1)} \quad (4)$$

where \hat{Y} is the predicted output signal of the system Σ and U is the control sequence that corresponds to the way

our pendulums are swung and excited in time. By forming the vector U in this manner, i.e. by concatenating the elements u_k , the information of which mode is activated is embedded in the u_k element whereas the temporal information about when a mode is activated is indicated by the subscript k .

Given (4), the equation (3) can be written in a matrix form as:

$$(\phi_1 \ \phi_2) \cdot \begin{pmatrix} x_0 \\ U \end{pmatrix} = \phi \cdot \beta = \hat{Y} \quad (5)$$

where:

$$\phi_1 = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{pmatrix}, \phi_2 = \begin{pmatrix} 0 & 0 & \dots & 0 \\ CB & 0 & \dots & 0 \\ CAB & CB & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2}B & CA^{L-3}B & \dots & CB \end{pmatrix}$$

To replicate the pulsating nature of brain rhythms described in section 2.2, we have induced sparsity in the solution of the equation (5). Effectively, with a sparse beta, each mode will be switched-on only a few times which is close to the pulsating nature of the brain rhythms. The Lasso-LARS algorithm was chosen to induce sparsity in the β solution for reasons that will be explained later. The Lasso-LARS minimizes a classical prediction error $\epsilon = \|Y - \phi\beta\|_2^2$ whereas parsimony is induced by penalising non-zero β entries through a weighted l_1 norm:

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \frac{1}{2} \|Y - \beta_0 - \phi\beta\|_2^2 + \alpha \|\beta\|_1 \quad (6)$$

where β_0 is the intercept value¹, α is a weighting constant, and $\hat{\beta}$ is an estimate of the optimal solution β^* . Complete details about Lasso-LARS can be found in (Hastie et al., 2009, chap. 3). The Lasso-LARS was chosen over other subset selection and shrinkage methods because it is very efficient to find the solution of the lasso problem when $\dim(\beta) \gg L$ which corresponds to our case. Most importantly, it allows to give different β solutions as alpha varies, resulting in multiple solutions with a varying parsimony level.

The way the Lasso-LARS algorithm is designed makes that the final result β is not fully fitted, i.e., to minimize the equation (6) β must be small in magnitude so that $\|\beta\|_1$ is small. Lasso-LARS was used mainly to induce sparsity and find the few most important entries of β to fit Y . In the end, we took the non-zero entries of β computed by Lasso-LARS and then completed the fit with a Least Square method, thus reducing the prediction error ϵ without increasing the number or the frequency of use of the modes.

The weighting value α is used to adjust the sparsity level of the solutions of β . Therefore, the value of alpha will be used to adjust the degree of fit of our model. For a high value of α , the model will have to use fewer modes and less frequently to fit Y whereas for a low value the model will excite more modes and more frequently. The best solution that will give the best result does not have to fit perfectly the Y signal as there is a trade-off between fitting Y perfectly by capturing all the noise and not fitting Y completely but at the expense of losing information (see Fig.1). The Lasso-LARS intrinsically calculates the solutions for a decreasing level of parsimony, starting with

¹ Mean value of Y .

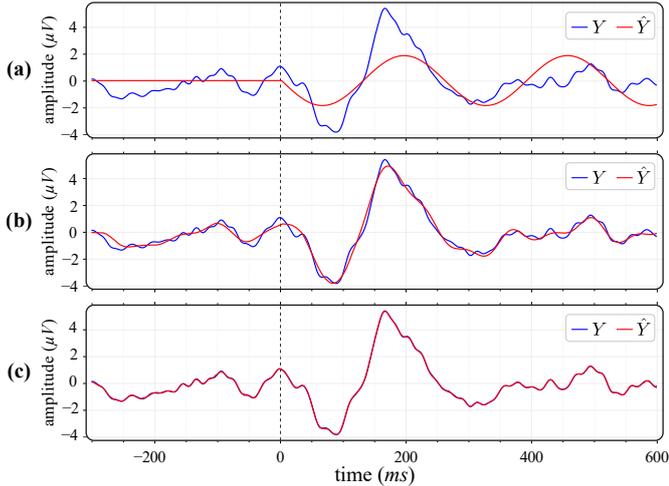


Fig. 1. Effect of the α value on the predicted signal \hat{Y} . (a) Activation of a single mode ($\|\beta\|_0 = 1$). (b) The use of more modes ($\|\beta\|_0 = 17$). (c) Fit the Y signal perfectly with ($\|\beta\|_0 = 84$).

a value² α_{init} where the solution is very sparse and corresponds to a sparsity level of 100% until reaching the final value indicated to the algorithm α_f where the solution is not sparse and corresponds to a sparsity level of 0%. Not knowing in advance which level of parsimony is the best, we divided the interval (100%—0%) into 1% linearly spaced intervals. For the rest of the paper we consider only the U part of β that we call VMS (for Virtual Modal Stimuli), the latter is responsible for capturing the evoked activity. Complete details about the algorithm can be found in (Meghnooudj et al., 2022).

The VMS of the ERP of each channel and each stimulus were generated, they constitute the input of the next step which is the feature extraction.

2.4 Feature selection

For validity reasons due to data limitation, we had to extract features from the calculated VMS. The classifier will rely on these extracted features to classify the subjects into SZ or CTL. The feature extraction has multiple benefits:

- (1) Reduce over-fitting.
- (2) Improves the stability of the model and increase the trustfulness placed on the model as the accuracy should not change much for a new unseen data.
- (3) Extracts only the important information that the model should use.
- (4) Reduces the classifier’s required complexity.
- (5) Provide a physical meaning to our model.

It is very important to note that the more features we have, the more data we will need to converge to the true distribution of the data. Indeed, when the dimension of the features increases, the volume of the space grows exponentially (curse of dimensionality), so that the available data will be scattered in that space. In order to have reliable results, the number of data must also grow exponentially, which is the opposite of what is currently happening in the medical field (Bellman, 1957).

² Calculated by the algorithm on the basis of correlation, see (Hastie et al., 2009) for more details.

Before proceeding to the next step which is the evaluation, it is important to remember that it is the features extracted from each ERP that are evaluated individually for each channel and for each stimulus.

2.5 Evaluation

In order to avoid data leakage, it is necessary that correlated information of the same patient is not present at the same time in both the training and test sets. This corresponds to our case since each patient is represented by a single data instance that can only be in one set at a time. Moreover, tuning the hyper-parameters directly on the test-set or indirectly by choosing the best hyper-parameter that worked best over multiple repeated training-test-validation splits is also considered as data leakage.

Ideally, to prevent data leakage, we should keep a hold-out set aside before the beginning of the experiment on which we will evaluate the method once finalized and once the hyper-parameters have been well chosen using the learning set. However, for a limited number of data as in our case, leaving for example 20% of the data aside can be troublesome as we lose in data richness and this is even more accentuated by the curse of dimensionality. Furthermore, this will result in only one evaluation and given the limited data available it is highly likely that we get a bad or good draw of the hold-out test data. This phenomenon is known as selection bias and has been studied in more detail by Varma and Simon (2006), and Krstajic et al. (2014).

The nested Leave-One-Out CV (LOO), to our knowledge, is the less biased evaluation procedure that prevents data leakage, reduce selection bias, while preserving the richness of the learning data. The procedure is presented in more details in Varma and Simon (2006), and Meghnooudj et al. (2022).

To assess the generalization ability of our approach, we conducted further evaluations under more constrained settings by reducing the amount of available learning data. Specifically, we employed nested 5-fold and 2-fold procedures, where we had access to only 80% and 50% of the learning data, respectively. These tests allowed us to examine how well our approach could perform on new, unseen data and provided valuable insights into its robustness and reliability in real-world applications.

For reproducibility purposes, a step-by-step guide has been prepared detailing how to download the data and how to run our source code. An explanatory animation is also provided at the following link: <https://github.com/HoussemMEG/Schizophrenia-Detection>

3. RESULTS

The system Σ was created using $m = 40$ pendulums that have a linearly spaced frequency taken from (0.1—50) Hz. The weighting constant α_f has been set to $\alpha_f = 0.001$ by visual inspection in order to guarantee a progressive fit at each level of sparsity and that at the end a complete fit is obtained as in Figure 1.c.

The VMSs for each channel and condition were first generated, and then several simple features were extracted from them. Then, each feature was tested individually.

The features used are typical ones and include: mean, max, energy, peak-to-peak value, variance, argmin, etc. We also tried to test the pairwise combinations of the 16 features we picked as well as the pairwise combinations of the channels without getting much improvement in the results.

3.1 Results with different evaluation procedures

A first result was obtained using a LOO evaluation procedure with the feature *mean*(VMS) over the EEG channel C3 and for the condition (b) which corresponds to passively listened to the same tone. The other channels and features did not show significant results. For informative purposes, the generation of the VMS and the LOO evaluation takes about 2 minutes overall.

Table.1 shows the results obtained for the nested LOO procedure as well as for the nested 5-fold and 2-fold CV procedures. For both 5-fold and 2-fold CV the partitioning were repeated 20 000 times.

Table 1. Accuracy obtained by our model using a single feature for different evaluation procedures.

Evaluation scheme	LOO	5-fold	2-fold
Learning (\pm <i>std</i> %)	85.1 % (—)	84.2 % (0.6)	84.5 % (1.6)
Validation (\pm <i>std</i> %)	85.0 % (—)	83.9 % (0.6)	84.3 % (1.5)
Test (\pm <i>std</i> %)	85.2 % (—)	80.0 % (2.6)	77.2 % (3.5)

Figure 2 represents the confusion matrix for the case of the LOO evaluation procedure. The sensitivity is 90 % and the specificity is 78 %. We can observe that the results obtained are not perfectly balanced.

		Predicted label	
		SZ	CTL
True label	SZ	44 (90 %)	5 (10 %)
	CTL	7 (22 %)	25 (78 %)

Fig. 2. Result confusion matrix for LOO procedure.

3.2 Significance and trustfulness

To assess the statistical validity and consistency of our results for the C3 channel we performed two tests:

- (1) Permutation test (Nichols and Holmes, 2003): The probability that the results obtained are simply due to luck is tested by as we have a small amount of data and have performed several runs (each for each type of stimuli, channel and feature). We define the following null hypothesis H_0 : taking the mean of the VMS does not allow to differentiate the SZ group from the CTL group. Under this hypothesis, we randomly shuffled the labels $n = 1000$ times, then

we looked at the maximum accuracy obtained. We obtained a p -value of ($p < 0.005$) indicating that the results obtained are significant. As a reminder, for a statistical significance, the p -value must be lower than 0.05.

- (2) Parametric consistency: For this part we have evaluated the evolution of the accuracy as a function of the parsimony level. Therefore, unlike the other parts where we have a nested cross validation, during this part we just have the outer loop where we used a LOO CV for each parsimony level. The obtained results are shown in Figure.3. It should be noted that the area of interest is the area that has been selected 100 % of the time by the inner loop when choosing the alpha hyperparameter. This indicates that the results obtained have a parametric consistency and that it is not due to chance that one result spikes on a given parsimony level.

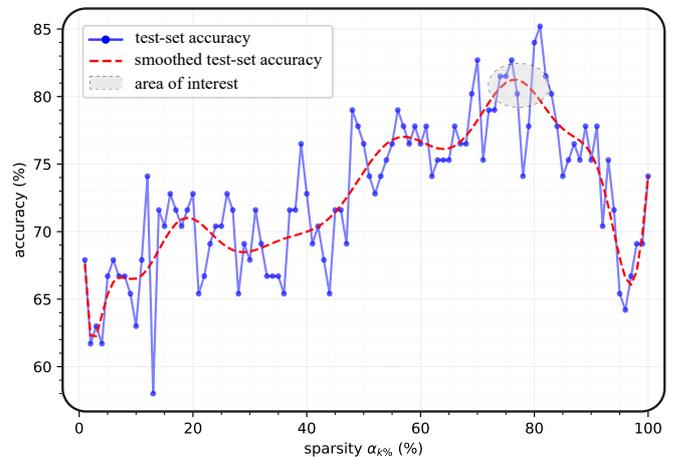


Fig. 3. Model's test accuracy for a varying level of parsimony.

4. CONCLUSION

In the present paper we evaluated an approach which is inspired by brain functioning for the diagnosis of Schizophrenia using EEG signals. The method was validated on a publicly accessible dataset containing 81 subjects, 49 of whom have schizophrenia and 32 of whom serve as controls. The method used is inspired by the functioning of the brain and combines the temporal, frequency and dynamic aspect of the signal to generate Virtual Modal Stimuli. They are virtual stimuli or rhythms that have been computed in a parsimonious way to fit a signal of interest, in our case the ERP. The evaluation of our approach was carefully conducted to induce as little bias as possible so as not to have a model that is over-optimistic and that only works on the present data. A single feature which is the mean of the virtual modal stimuli calculated on the C3 channel was sufficient to separate with a straight line (LDA) the healthy from the SZ with an accuracy of 85.2 % ($p < 0.005$), a sensitivity of 90 %, and a specificity of 78 %. Several tests have been performed to verify the validity of our approach. The results obtained are still preliminary, different perspectives can be considered including the use of a voting strategy between different channels allowing the aggregation of different information from different brain regions.

ACKNOWLEDGEMENTS

The author would like to thank Brain Roach and Dr. Ford for sharing their EEG dataset. The initial data collection was supported by the National Institute of Mental Health (R01MH058262).

REFERENCES

- APA (2022). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association Publishing. doi:10.1176/appi.books.9780890425787.
- Bellman, R.E. (1957). *Dynamic programming*. Princeton University Press, Princeton, NJ.
- Buzsaki, G. and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679), 1926–1929. doi:10.1126/science.1099745.
- Devia, C., Mayol-Troncoso, R., Parrini, J., Orellana, G., Ruiz, A., Maldonado, P.E., and Egana, J.I. (2019). EEG classification during scene free-viewing for schizophrenia detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(6), 1193–1199. doi:10.1109/tnsre.2019.2913799.
- Ford, J.M., Palzes, V.A., Roach, B.J., and Mathalon, D.H. (2013). Did i do that? abnormal predictive processes in schizophrenia when button pressing to deliver a tone. *Schizophrenia Bulletin*, 40(4), 804–812. doi:10.1093/schbul/sbt072.
- GHDx (2021). Global Health Data Exchange, Institute of health Metrics and Evaluation (IHME) [Online]. URL <http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/27a7644e8ad28e739382d31e77589dd7>. Accessed 25 September 2021.
- Green, M.F., Bearden, C.E., Cannon, T.D., Fiske, A.P., Helleman, G.S., Horan, W.P., Kee, K., Kern, R.S., Lee, J., Sergi, M.J., Subotnik, K.L., Sugar, C.A., Ventura, J., Yee, C.M., and Nuechterlein, K.H. (2011). Social cognition in schizophrenia, part 1: Performance across phase of illness. *Schizophrenia Bulletin*, 38(4), 854–864. doi:10.1093/schbul/sbq171.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning*. Springer Series in Statistics. Springer, New York, NY.
- Insel, T.R. (2010). Rethinking schizophrenia. *Nature*, 468(7321), 187–193. doi:10.1038/nature09552.
- Khare, S.K. and Bajaj, V. (2022). A hybrid decision support system for automatic detection of schizophrenia using EEG signals. *Computers in Biology and Medicine*, 141, 105028. doi:10.1016/j.combiomed.2021.105028.
- Krstajic, D., Buturovic, L.J., Leahy, D.E., and Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics*, 6(1). doi:10.1186/1758-2946-6-10. URL <https://doi.org/10.1186/1758-2946-6-10>.
- Linscott, R.J. and van Os, J. (2012). An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychological Medicine*, 43(6), 1133–1149. doi:10.1017/s0033291712001626.
- Luck, S.J. (2005). *An introduction to the event-related potential technique*. Cognitive neuroscience. Bradford Books, Cambridge, MA.
- Meghnoudj, H., Robu, B., and Alamir, M. (2022). Sparse dynamical features generation, application to parkinson’s disease diagnosis [unpublished]. doi:10.48550/ARXIV.2210.11624. URL <https://arxiv.org/abs/2210.11624>.
- Nichols, T.E. and Holmes, A.P. (2003). Nonparametric Permutation Tests for Functional Neuroimaging. *Human Brain Function: Second Edition*, 25(July), 887–910. doi:10.1016/B978-012264841-0/50048-2.
- Olejniczak, P. (2006). Neurophysiologic basis of EEG. *Journal of Clinical Neurophysiology*, 23(3), 186–189. doi:10.1097/01.wnp.0000220079.61973.6c.
- Prabhu, S. and Martis, R.J. (2020). Diagnosis of schizophrenia using kolmogorov complexity and sample entropy. In *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. IEEE. doi:10.1109/conecct50063.2020.9198472.
- Roach, B. (2021). Eeg data from basic sensory task in schizophrenia — button press and auditory tone event related potentials from 81 human subjects. [online]. URL <https://www.kaggle.com/datasets/broach/button-tone-sz>. Accessed 06 November 2022.
- Santos-Mayo, L., San-Jose-Revuelta, L.M., and Arribas, J.I. (2017). A computer-aided diagnosis system with EEG based on the p3b wave during an auditory odd-ball task in schizophrenia. *IEEE Transactions on Biomedical Engineering*, 64(2), 395–407. doi:10.1109/tbme.2016.2558824.
- Urigüen, J.A. and Garcia-Zapirain, B. (2015). EEG artifact removal - State-of-the-art and guidelines. *Journal of Neural Engineering*, 12(3). doi:10.1088/1741-2560/12/3/031001.
- Varma, S. and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7(1). doi:10.1186/1471-2105-7-91.
- Walsh-Messinger, J., Jiang, H., Lee, H., Rothman, K., Ahn, H., and Malaspina, D. (2019). Relative importance of symptoms, cognition, and other multilevel variables for psychiatric disease classifications by machine learning. *Psychiatry Research*, 278(November 2018), 27–34. doi:10.1016/j.psychres.2019.03.048. URL <https://doi.org/10.1016/j.psychres.2019.03.048>.
- Ward, L.M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, 7(12), 553–559. doi:10.1016/j.tics.2003.10.012.
- WHO, W.H.O. (1997). Implementation of the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10). *Epidemiological bulletin*, 18(1), 1–4.
- Zhang, L. (2019). EEG signals classification using machine learning for the identification and diagnosis of schizophrenia. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. doi:10.1109/embc.2019.8857946.