Artificial Intelligence In Medicine 105 (2020) 101852

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

On the use of pairwise distance learning for brain signal classification with limited observations

David Calhas[a,*], Enrique Romero[b], Rui Henriques[a]

[a] INESC-ID and IST, Universidade de Lisboa, Lisbon, Portugal
[b] Universitat Politècnica de Catalunya, Barcelona, Spain

ARTICLE INFO

Keywords:
Pairwise learning
Schizophrenia
Classification
Electroencephalography

ABSTRACT

The increasing access to brain signal data using electroencephalography creates new opportunities to study electrophysiological brain activity and perform ambulatory diagnoses of neurological disorders. This work proposes a pairwise distance learning approach for schizophrenia classification relying on the spectral properties of the signal. To be able to handle clinical trials with a limited number of observations (i.e. case and/or control individuals), we propose a Siamese neural network architecture to learn a discriminative feature space from pairwise combinations of observations per channel. In this way, the multivariate order of the signal is used as a form of data augmentation, further supporting the network generalization ability. Convolutional layers with parameters learned under a cosine contrastive loss are proposed to adequately explore spectral images derived from the brain signal. The proposed approach for schizophrenia diagnostic was tested on reference clinical trial data under resting-state protocol, achieving 0.95 ± 0.05 accuracy, 0.98 ± 0.02 sensitivity and 0.92 ± 0.07 specificity. Results show that the features extracted using the proposed neural network are remarkably superior than baselines to diagnose schizophrenia (+20pp in accuracy and sensitivity), suggesting the existence of non-trivial electrophysiological brain patterns able to capture discriminative neuroplasticity profiles among individuals. The code is available on Github: https://github.com/DCalhas/siamese_schizophrenia_eeg.

1. Introduction

The recording of increasingly affordable and precise electroencephalography (EEG) data is creating unprecedented opportunities to understand brain activity, aid personalized prognostics, and promote health through wearable biofeedback systems [1]. Electroencephalography is non-invasive, safe, inexpensive, and shows rich temporal content; in contrast with other brain imaging modalities, such as magnetic resonances, entailing higher costs and restrictions on the longitudinal periodicity of recordings [2]. EEG monitoring is widely used to assess psychiatric disorders, and has shown to be a valuable source to study schizophrenia, a disorder affecting about 1% of the world population, largely susceptible to misdiagnoses [3]. Since 2017, cases of individuals with schizophrenia able to regulate their brain activity using real-time EEG neurofeedback in therapeutic settings have been reported [4]. Comprehensive reviews of EEG-based studies of schizophrenia from case-control populations reveal general spectral deviations, including predisposition for decreased alpha power and an increase of activity in the lower spectrum [5]. Slow wave abnormality (mainly delta activity) can be primarily localized in frontal lobe regions, and is suggested to be a relevant neurophysiological marker of schizophrenia [5]. Connectionist and information theoretic features to discriminate brain electrophysiology have been additionally proposed [6,7]. Despite the inherent advantages of the spectral markers and proposed scores, their use for schizophrenia diagnosis still results in high false positive and false negative rates due to the extent of individual differences on the electrophysiological activity of the brain, irrespective of clinical condition. In particular, when considering resting-state protocols for EEG recordings – clinically deemed as desirable in psychiatric settings against task-oriented and stimuli-induced settings [8] –, state-of-the-art classifiers based on the aforementioned features generally show diagnostic accuracy rates below 70%.

The difficulty of EEG-based diagnostics of neuronal diseases is mainly driven by two major factors: the limited size of case-control populations [9], and the intrinsic difficulties of mining brain signals. Brain signal data is high-dimensional, multivariate, susceptible to noise/artifacts, rich in temporal-spatial-spectral content, and highly variable between individuals [10].

This work proposes a dedicated class of neural networks to extract discriminative features of schizophrenia from electrophysiological

* Corresponding author at: INESC-ID, Rua Alves Redol 9, 1000-029 Lisboa, Portugal.
E-mail addresses: david.calhas@tecnico.ulisboa.pt (D. Calhas), eromero@cs.upc.edu (E. Romero), rmch@tecnico.ulisboa.pt (R. Henriques).

https://doi.org/10.1016/j.artmed.2020.101852
Received 23 October 2019; Received in revised form 30 March 2020; Accepted 31 March 2020
0933-3657/

brain data previous to the classification step. The proposed approach combines principles from pairwise distance learning and spectral imaging in order to address the aforementioned challenges, enabling superior diagnostics. Accordingly, the proposed approach offers six major contributions:

1. Ability to learn from small datasets by taking advantage of Siamese network layering, inherently prepared to work in augmented data spaces mapped from a limited number of observations. Specifically, our approach is suggested for databases with dozens to hundreds of EEG recordings [11]. The features produced by Siamese networks have shown to be useful to perform classification as they rely on either the homologous or discriminative properties of observation-pairs in a pairwise distance domain [12];
2. Ability to deal with the rich and complex spectral and temporal content of EEG data by processing the signal into spectral images with a fine frequency and temporal resolution per electrode [13,14], and by subsequently reshaping the Siamese network architecture with adequate convolutional operations;
3. Robustness to noise and wave-instability by assessing distances on the spectral content (frequency domain) under a cosine-loss. Gathered evidence shows less susceptibility to artifacts and the inherent variability of electrophysiological potentials associated with continuously changing overlapping electrical fields produced by localized neurons [10];
4. Ability to deal with the multivariate nature of the signal (rich spatial content) by capturing interdependencies between channels as their content is simultaneously used to shape the learned classifiers;
5. Ability to handle the extremely high dimensional nature of the gathered spectral content from brain signals (high-resolution spectral image per electrode) under L1 regularization [15,16];
6. Applicability of the proposed EEG-based diagnostics to alternative populations or diseases, motivated by the: (i) placed Bayesian optimization step [17] for hyperparameter tuning and fixing feature space dimension; (ii) fully automated nature of the approach once signals are recorded; and (iii) generalization ability of the learning process on validation data.

In contrast with the currently established views on neural information processing systems, this manuscript explores whether we can go deep on highly dimensional spatiotemporal data in the presence of a very limited number of data observations. This stance is much-needed in healthcare given the limited size of trials (cohort studies), often driven by disease rarity, capped size of control population, trial eligibility requirements, or the facultative nature of EEG assessments. Experimental results confirm this possibility.

This work is validated on the clinical trial conducted by Gorbachevskaya and Borisov [11], a reference database for the resting-state analysis of schizophrenia. Further, details can be found in Section 3.1. The proposed learning approach achieves 0.95 ± 0.05 accuracy, 0.98 ± 0.02 sensitivity on schizophrenia diagnostics, remarkably attaining an improvement of over 20pp against peer approaches.

The features extracted from the proposed spectral and pairwise distance space further suggest the presence of discriminative electrophysiological patterns linked to neuroplasticity aspects of the individuals. This observation is in accordance with findings from previous studies that established statistically significant relationships between variations in the frequency band spectrum and neuroplasticity conditions [18,19].

The manuscript is organized as follows. After formalizing the problem, Section 2 surveys existing contributions to the diagnosis of individuals from brain signal data. Section 3 describes the proposed solution. Section 4 shows extended evidence of its relevance for diagnosing schizophrenia. Finally, concluding remarks are drawn in Section 5.

1.1. Problem formulation

1.1.1. Problem

A EEG recording or brain signal observation is a multivariate time series $X = \{x_t^j | j \in \{1, ..., M\}, t \in \{1, ..., T\}\}$, where x_t^j is a measure of the electrophysiological activity in scalp channel j and instant t, T is the number of time points, and M is the multivariate order (number of channels). Given brain signal dataset, $\{(X_i, c_i) \mid i = 1, \ldots, N\}$, where N is the number of EEG recordings and each recording X_i is annotated with a label $c_i \in \Sigma$, our task is to identify a discriminative feature space to classify (unlabeled) observations. Specifically, we are interested in classifying schizophrenia given case-control populations.

1.1.2. Essential background

The electrophysiological signal produced by a specific channel in the cerebral cortex is a univariate time series that can be decomposed into a frequency time series using a discrete Fourier transform. The analysis of the frequency domain of a signal, generally referred as spectral analysis, determines the predominant waves monitored at a certain location. A short-time discrete Fourier transform can be alternatively applied along a sliding window of the raw signal to capture potentially relevant changes on the spectral activity of the brain throughout the EEG recording. The spectral content produced by this time-varying form of spectral analysis is here informally referred as a *spectral image* since it measures brain activity along two contiguous axes: frequency and time.

2. Related work

Recent works on deep learning provide principles to attemptively learn from small datasets [20,21], a critical requirement if we want to guarantee their applicability for most cohort studies available worldwide. The use of surrogate data analysis in the context of regresssion tasks [20], or data augmentation procedures for image recognition [22] are paradigmatic cases. Despite their relevance, they either tackle different tasks or assume a substantial higher amount of data observations than the ones commonly available in clinical trials; leave aside the need to handle the high dimensionality, spectral variability, and rich spatiotemporal content of EEG data.

To form a comprehensive picture of relevant contributions, sections below provide a state-of-the-art views on EEG classification (irrespective of clinical condition), EEG-driven analysis of schizophrenia, and relevant advances on deep learning.

2.1. EEG classification

EEGNet [23], EEGNet-SSVEP [23], DeepConvNet [24] and ShallowConvNet [24] are considered state-of-the-art EEG classification built models that make use of convolutional operations directly on the raw EEG data. These convolutions are placed along time and channels. Approaches like these rely on the properties of its models to extract discriminative features from EEG signals. These models have been primarily validated in the context of stimuli-induced recording sessions. One can see directly that these networks learn event related potentials from the EEG signal, which makes the EEG recording session dependable of a task environment for evoking potentials. In contrast, we aim at extracting neuroplasticity-related features from resting state EEG data, for which effective deep learning methods are still in demand. Section 4 confirms the limited relevance of existing methods to learn from resting state EEG data.

2.2. EEG on schizophrenia

Dvey-Aharon et al. [25] claim mostly changes in functional connectivity are seen in patients with schizophrenia, as well as differences in theta-frequency activity. A classification approach was applied on 1-

min signals recorded by a single electrode. The developed system consists of four stages: performing preprocessing tasks and breaking the raw signals into relevant intervals; transformation of the EEG signal into a time-frequency representation via the Stockwell transformation; feature extraction from the time-frequency representation; and discrimination of stimuli-induced responses between features gathered from healthy and schizophrenia subjects. A K-nearest neighbor classifier under a Euclidean distance is suggested to diagnose individuals from the extracted time-frequency features. Despite its promising results, the approach requires the performance of cognitive tasks by the individuals under assessment throughout the recording. More recently, the authors introduced another way of looking at the EEG signal using connectivity maps derived from the brain activity [6]. In order to build these maps, a similarity function needs to be chosen, so one can check which nodes are more similar to which ones. Results showed that the degradation of connectivity is being accelerated within schizophrenia individuals. And that information relay changes in an abnormal manner primarily in the prefrontal area. This gives a good insight on how connectivity maps can be applied to discriminate schizophrenia. And most important, that one should take into account that a change in a certain region can influence other regions in the brain.

Sabeti et al. [7] introduced another approach to classify schizophrenia based on entropy and complexity measures of the EEG signal. The features extracted from the signal were: Shannon entropy, spectral entropy, approximate entropy, Lempel-Ziv complexity and Higuchi fractal dimension. Genetic programming was used for feature selection. With these features, adaptive boost (AdaBoost) and linear discriminant analysis (LDA) classifiers were validated, showing performance improvements against peer approaches. The recordings were done with eyes open, a setting easily biased by environmental effects. Zhang et al. [5] provided a status overview on EEG abnormalities in individuals with schizophrenia. To this end, they examined the status of development of spectral EEG deviations. In the gathered studies, the meta analysis was limited to those works comparing spectral power between one group of schizophrenia patients and one group of healthy control subjects. The presence of two groups (or populations), one with the pathology and a healthy control group, is essential to identify discriminative features from the gathered signals. The hypothesized differences of schizophrenia individuals were increased delta, increased theta, decreased alpha, and increased beta power. A number of subsequent studies suggested that an increase of activity in the lower spectrum (slow waves) is significantly higher in schizophrenia populations. It is also noted that slow wave abnormality (mainly delta increase) is mainly localized in frontal lobe regions. One of the conclusions, is that the delta excess (and to a lesser extent the theta excess) is a strong biological marker of schizophrenia [5].

The therapeutic utility of EEG based assessments has been recently shown in the context of a neurofeedback training on a female patient, who suffered from schizophrenia for more than 7 years and experienced several psychotic episodes [4]. Along the neurofeedback training, increased amplitude in alpha waves and decreased amplitude in beta waves were observed. The patient was able to successfully identify the most effective mental strategies and learnt how to regulate her brain activity (mental strategies were induced with the help of a psychotherapist). Event-related EEG stems suggest that neural oscillations and their synchronization represent important mechanisms for interneuronal communication and binding of information that is processed in distributed brain regions [26]. Despite their relevance, frequency-based analysis is generally encouraged when performing EEG-based studies of schizophrenia. This can be experimentally shown when comparing the performance of frequency models against models based on event related potentials [27,23,24], i.e. models that purely process the signal in a time domain not taking into account the frequency domain (results in Section 4).

In summary, notable examples of connectionist and spectral approaches were introduced to discriminate and characterize schizophrenia. Nevertheless, there is still a research gap on how to simultaneously explore the rich spectral, temporal and spatial nature of brain signals to perform classification.

2.3. Deep learning from EEG data

In spite of the indisputable role of neural network learning for the analysis of complex spatiotemporal signal data, its role for EEG-based diagnostics of psychiatric disorders remains largely unexplored due to the absence of large clinical trials [28]. A few recent studies counteract previous observations, offering principles on how to apply deep learning over EEG data [29–31].

Ieracitano et al. [29] specifies principles to apply deep learning to discriminate Alzheimer's disease and mild cognitive impairment from cognitively normal individuals. The EEG recording session was setup with 19 channels, and 189 recordings were collected (63 individuals from each class). As stated by [29], standard machine learning methods are unable to adequately deal with the high dimensionality of EEG data (when taking into account channels and frequency bands). To address this problem, and focus on relevant patterns from EEG data, convolutional operations were employed to extract features from the power spectrum density, using rectified linear unit (ReLU) as activation functions. In contrast with this study, our work takes into account the raw frequency time series (instead of the overall magnitude of each frequency) and proposes a pairwise schema to improve the ability to learn from small datasets. Note that the analysis of the raw frequency time series is useful as it has been previously correlated with neuroplasticity properties of the brain [18].

Oh et al. [30] performs schizophrenia classification using principles of deep learning from EEG recordings. EEG recordings were gathered from 14 healthy controls and 14 schizophrenic individuals. A total of 19 electrodes were used and the signal was sampled at 250Hz. Convolutional operations on the time domain, which is not encouraged when following a resting state protocol. The target EEG dataset had segments with a naming activity (task-oriented), and we hypothesize this is the reason for the observed competitive results (81.26% accuracy).

Previous works [29,30] have an additional downside: the chosen architectures were manually tuned and can bring discussion on the nature of the selected hyperparameters. In contrast, our work does not attempt to manually tune the hyperparameters of the network, leaving hyperparameterization to a Bayesian optimizer. The goal is to easily bridge the research done to a real-world setting, where irrespective of the application domain: two populations of individuals are gathered and the proposed SNN architecture is tuned using a Bayesian optimization algorithm. After this training procedure, the model is able to diagnose patients in a few seconds.

2.4. Siamese neural network

Siamese neural networks (SNNs), first introduced by Bromley et al. [32] to distinguish signature forgeries from real ones, are deep learning architectures with two sub-networks that consist on the same instance, hence being called “Siamese networks”. This architecture receives as input a pair of samples. Subsequently, the outputs of the pairs used as input to these “Siamese networks” are joined in a distance function. The proposed distance function between the output of the SNNs is the cosine similarity (for signatures from the same person the output should be 1, and -1 for forged ones). This model had outstanding results at the time, detecting 80.0% of the forged signatures and 95.5% of the genuine signatures. More recently, Kock et al. [12] successfully used a SNN architecture for one shot learning (meaning the model only sees each class once in an epoch). This approach reached 92.8% accuracy in the test set. These results were achieved through a Siamese convolutional architecture. Once this kind of network is trained, its learned representations via a supervised metric-based approach with SNNs are useful to perform tasks like classification, relying on the discriminative properties of these features.

Medicine has its roots on statistics [33], as it purely compares observations with history archives when making diagnosis. In contrast with Siamese neural networks, traditional CNNs for EEG classification identify features able to discriminate between case and control individuals, yet there is no direct comparison necessarily being done. A SNN learns by comparison, a learning mechanism quite similar to the way medicine diagnoses are operated by humans. This observation, together with the need to learn from small datasets, offers the preliminary motivation for the application of SNNs in the target medical task: schizophrenia diagnosis.

3. Our approach

The proposed architecture is inspired by the architecture formerly introduced by Kock et al. [12]. An advantage of this type of architecture is the ability to augment the original dataset from an instance-based data space to a pair-based one. Our approach has two main steps: (1) feature extraction; and (2) classification. In step 1, the internal representations obtained from the SNN architecture model are extracted after training. In step 2, a classification task is performed using these extracted features. Previous to both steps, we perform hyperparameter optimization for every model using Bayesian optimization (BO) [17].

3.1. Dataset description

Approaches based on induced stimuli or task performance, followed by the analysis of event related potentials, are not considered in this work. Instead, a resting state setup is considered to monitor the underlying brain patterning at the brain cortex, independently of the surrounding environment/undertaken task. Resting-state protocols are clinically deemed as having inherent advantages in psychiatric settings. The resting-state protocol minimizes electrophysiological differences associated with the individuals' subjective perception and elicited response to the applied stimuli or task. Finally, it prevents additional interference on the recorded EEG signal, and lowers the number of visual artifacts in the EEG signal. Howells et al. [8] findings support the use of this setup, claiming that differences on the spectral activity – such as higher delta and a lower alpha synchronization in psychotic disorders – can be optimally detected in resting state protocols with both open and closed eyes.

Table 1 shows the content of EEG datasets containing healthy control individuals and schizophrenic individuals. Dvey-Aharon et al. [6,25] and Sabeti et al. [7] works were introduced and discussed in Section 2. Unfortunately, the considered datasets have a strictly low number of observations, and are not made publicly available. Nonetheless, Gorbachevskaya and Borisov [11] performed a broader resting state recording on a total of 84 individuals (45 regarded as schizophrenic and 39 as healthy controls). This dataset is used to thoroughly assess the proposed contributions. By being publicly available, it allows the reproducibility of the presented results. This population consists of adolescents who had been screened by a psychiatrist and got either a positive or negative diagnostic for the schizophrenia neuropathology. EEG recordings were sampled at 128 Hz with 1 min duration. Individuals were set in a resting state with eyes closed. In accordance with the 10–20 system of electrode placement, the topographical positions of the placed EEG channels are: F7, F3, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2. Given the properties of the recording protocol, as well as the subsequent application of spectral analysis and convolutional operations, the raw signal was not subjected to artifact removal or bandpass filters of frequencies below 1Hz.

Table 1
Schizophrenia EEG datasets.

Dataset reference	Healthy controls	Schizophrenic individuals	Access
Dvey-Aharon et al. [6,25]	20	20	Private
Sabeti et al. [7]	25	25	Private
Gorbachevskaya and Borisov [11]	**39**	**45**	**Public**

3.2. Siamese neural network architecture

The SNN architecture contains two sub networks that correspond to the same instance (twin networks). Both of these twin networks are referred to as the base network (BN). The input and output of the BN are an example and a feature vector, respectively. The output feature vector corresponds to the features extracted in the aforementioned step 1.

In our case, the BN receives as input a discrete short-time Fourier transform (DSTFT) representation of the EEG signal, that is extracted from the 1-min recording of a channel of an individual. The DSTFT is taken with 2 s length windows in order to capture frequencies as low as 0.5 Hz, corresponding to the delta wave frequencies (Howells et al. [8] points out that frequencies lower than 2 Hz are relevant to differentiate schizophrenia) and as high as 50 Hz. This image is processed through two convolutional layers, followed by a fully connected layer. The activation function used in the convolutional layers is the rectified linear function [34], while the fully connected layer uses the softmax activation function, normalizing the domain of the feature representations, $\mathbf{f} \in R^q,\ i \in [1, q] : \mathbf{f}_i \in [0, 1]$.

Once the BN network (Fig. 1) is built, a replication of it is made, producing its twin and sharing their weights. The SNN layout is achieved joining these twins and computing a distance metric between their outputs, as shown in Fig. 2. In our case, the inputs to the SNN are pairs of DSTFT representations and the outputs are the computed distance between the representations obtained by the BN.

The SNN tries to solve what is known as a neighbor separation problem, consisting on the separation of instances in a dataset that contains different classes. In our case we have two classes: schizophrenic and healthy control individuals. In this neighbor separation problem, pairs of individuals of the same class (schizophrenic with schizophrenic or healthy with healthy) are called neighbors and pairs of individuals of different classes (schizophrenic with healthy) are called non-neighbors. The network learns a transformation with the objective of assigning small distance to neighbors and large distance to non-neighbors.

With the previously described architecture, the neighbor separation problem can be posed as a minimization problem of a certain loss function that depends on such distance. In [35], the contrastive loss function is introduced to that end, defined as:

$$L(W, Y, X_1, X_2) = YD_W{}^2 + (1 - Y)\max(0, m - D_W)^2 \tag{1}$$

where (X_1, X_2) is the input pair, $Y = 1$ if X_1 and X_2 are neighbors and 0 otherwise, D_W is the distance between the predicted values of X_1 and X_2, and m is the margin value of separation. Minimizing the contrastive loss function leads to a scenario where neighbors are pulled together and non-neighbors are pushed apart, according to a certain distance metric. The margin value is sensitive. High values of m increase the separation between non-neighbors (pairs of different class), impacting positively the accuracy although making the training slower. In contrast, low values of m may cause the model not to learn the desired behavior.

3.2.1. Loss and regularization

The suggested contrastive loss function to measure the correlation between two feature vectors is the cosine loss. This metric generally shows reasonable performance improvements, suggesting that the coherence of spectral variations between spectral images (cosine loss) is more relevant than the actual absolute differences between images (Euclidean loss), an observation corroborated in other recent studies [21]. This observation also sheds light on how the schizophrenia

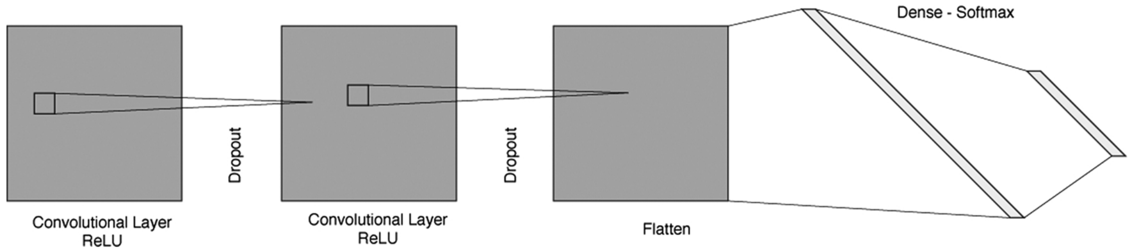


Fig. 1. Base network from the SNN.

pathology is expressed in the EEG.

In addition to the layering of the proposed network and the applied distance metrics, the following transformations are further applied to the target network: $L1$ regularization and dropout layers. The $L1$ regularization is useful as it helps removing features and/or associations that are not useful for the targeted task. Dropout layers are introduced to improve generalization. Regularization is applied at the kernel of all layers. The dropout probability used is 0.5, as suggested by [36], and is applied after each convolutional layer. Adam [37] is used to optimize the network during the training session.

3.2.2. Hyperparameter tuning

The number of layers, as well as their type, are fixed. The rest of hyperparameters (regularization factor, margin value, learning rate, kernel size and output dimension of the BN) are susceptible to optimization. As previously mentioned, we apply BO to that end. BO is set to run with a maximum of 50 acquisitions and starts with 5 iterations to perform an initial exploration. In each iteration and acquisition, a K-fold Cross Validation with $K = 5$ is done with the training set of a leave-one-subject-out cross validation (LOOCV) partition. The combination of hyperparameters that has the best average validation accuracy across the 5-folds is chosen to perform the feature extraction. Each of the hyperparameters are assigned the following value domains to explore: regularization factor $\in[10^{-3}, 10^{-1}]$, margin value $\in[1.0, 2.0]$, learning rate $\in[10^{-6}, 10^{-3}]$, kernel size $t \times f$ with $t = f = \{3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ (the same kernel size is used for both convolutional layers) and final output dimension $\in\{2, 4, 6, 8, 10, 12, 14\}$. The BO surrogate model is a standard Gaussian process. Expected improvement is used as an acquisition function and the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm as the acquisition optimizer.

The DSTFT magnitudes are normalized, under the hypothesis that there exists a threshold from which there is no additional information to identify the schizophrenia pathology. With this, the values are normalized by an upper value, U. Values of f smaller than U are divided by U and magnitudes bigger than U are set to 1.0. This allows every magnitude of the frequencies to be within the interval [0, 1] after the normalization is performed. We take advantage of the BO exploration to obtain U, by introducing it in the same optimization process made for the SNN hyperparameters. The domain assigned to be explored for U is [100.0, 500.0].

3.2.3. Pairwise dataset structure

To guarantee that the target network is able to learn valid transformation for all channels, the pairs are set such that only same channels are paired (Fig. 4). Pairs of different channels are not considered, since different channels are seen as correlated spaces with different properties. Fortunately, the SNN is capable of learning different spaces/classes, as shown in [12], where the proposed system is able to learn a similar setup. The pairwise schema brings a new optimization space to the classifier and consequently more observations versus a traditional classifier, which is one of the strongest motivations to use the SNN architecture. The main difference is that, by bringing the problem from an instance-based problem to a relation-based one (class agreement),

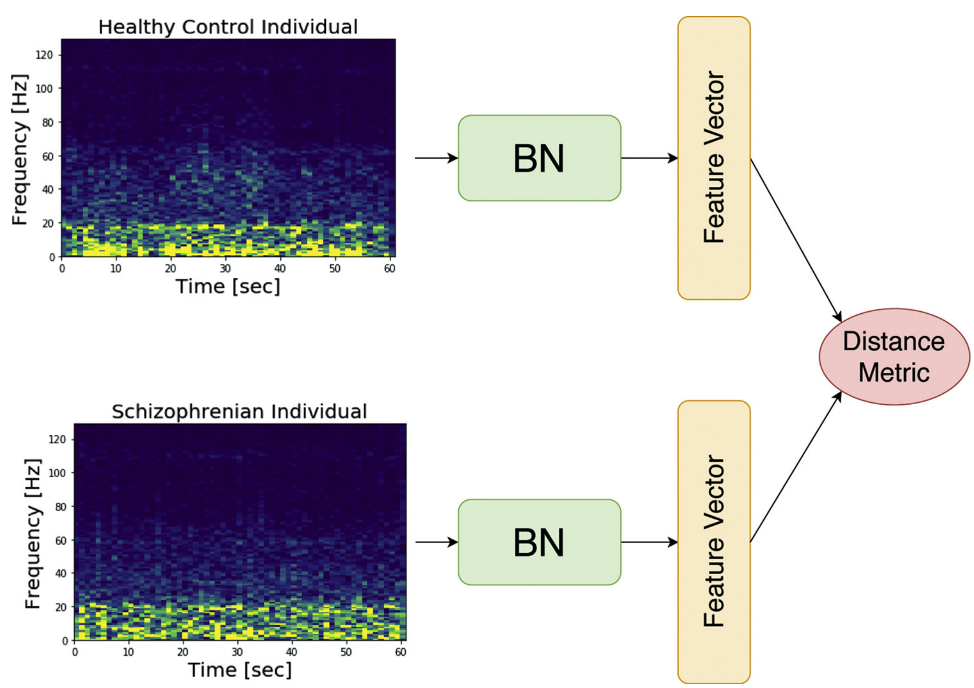


Fig. 2. SNN architecture.

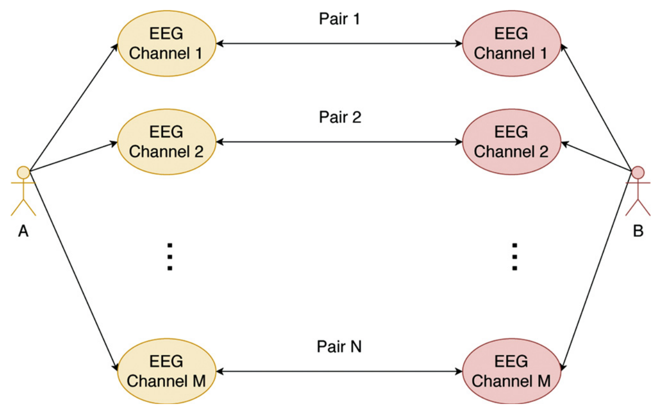


Fig. 4. Pair structure between two individuals and the corresponding EEG channels.

much more data is available to learn from. In other words, the feature variance is the same, but the optimization space has more information available. This pairwise scheme is not a classic data augmentation scheme, such as the ones achieved by oversampling, image data transformations (scaling, rotations), and noise addition. Nevertheless, as the number of available instances increases, the pairwise data space can be informally seen as an augmented data space.

From our original EEG dataset, $X_1, \ldots, X_N$ spectral images are derived with $N = 84$ examples, and a pairwise dataset P is built. Formally, $P = P_1, \ldots, P_O$ with $O = c\ C_2^N = M\ C_2^{84} = 55,776$, where $M = 16$ is the number of EEG channels. The space complexity of the pair dataset is $O(c\ C_2^N)$. The SNN training session is done with a batch size multiple of the number of channels. In particular, we use $B = 16 * c$. Therefore, there are 16 pairs of individuals in each batch and each pair of individuals has $c = 16$ channel pairs. This scheme can only be applied in small datasets, since the model does not scale well in terms of space complexity, but our goal is precisely to tackle small datasets with the creation of a whole new optimization space, where the variability contained in the data can be exploited in a different way.

3.3. Validation

Once the SNN has been tuned and trained (in a 20 epochs session), the outputs of the BN for every example were the result of our feature extraction process. With these features, the following classifiers were trained to identify schizophrenia: support vector machines (SVM), random forest (RF), XGBoost (XGB), naive Bayes (NB) and k-nearest neighbors (kNN). This process, illustrated in Fig. 3, was performed with a LOOCV, where each fold consists on one subject (16 channels/instances). For each of these classifiers, BO hyperparameter tuning is also performed, setup with a maximum of 10 acquisitions and 5 iterations for initial exploration. Algorithm 1 describes the validation schema. The hyperparameter domains for each classifier were:

- SVM: type of kernel (linear or radial-basis function kernel), cost $C \in [0.5, 5]$, and gamma coefficient $\gamma \in [0.00001, 1.0]$
- RF: number of estimators $N_e \in \{5, 10, 15, 20, 25\}$
- XGB: maximum depth $d \in \{3, 4, 5, 6, 7\}$, learning rate $\lambda \in [0.001, 0.1]$, and number of estimators $N_e \in \{10, 50, 100, 200\}$
- NB has no hyperparameters
- kNN: number of neighbors $k \in \{2, 3, 4, 5, 6, 7, 8\}$

Algorithm 1. Leave one subject out cross validation.

predictions $\longleftarrow$ {}
for each $X_i \in X$ **do**
  train $\longleftarrow X \setminus \{X_i\}$
  paired_train $\longleftarrow$ pair_structure(train)
  SNN.hyperparameters $\longleftarrow$ SNN.BO(paired_train)
  snn $\longleftarrow$ SNN.fit(paired_train)
  extracted_features_train $\longleftarrow$ snn.BN.predict(train)
  extracted_features_X_i $\longleftarrow$ snn.BN.predict(X_i)
  classifier.hyperparameters $\longleftarrow$ classifier.BO(extracted_features_train)
  clf $\longleftarrow$ classifier.fit(extracted_features_train)
  test_prediction $\longleftarrow$ clf.predict(extracted_features_X_i)
  predictions $\longleftarrow$ predictions $\cup$ mean(test_prediction)
end for
return predictions

The hyperparameter tuning optimization for the classifiers is also performed in a K-fold cross validation setup ($K = 5$), but instead of using the whole dataset (as was the case for the SNN) only the training set of the LOOCV partition was used. Similar to the BO for the SNN, the combination of hyperparameters with the best average validation accuracy is chosen for each classifier.

4. Results

Given the recording setting introduced in Section 3.1 consider the two following sets of paired individuals:

- $hc\_v\ s\_s\ cz$ – set of all pairs of non-neighbor individuals (healthy controls paired with schizophrenic);
- $hc\_a\ nd\_s\ cz$ – set of all pairs of neighbor individuals (healthy controls paired with healthy controls and schizophrenic paired with schizophrenic).

Fig. 5 shows the spectral differences using FFT between concordant pairs of individuals ($hc\_a\ nd\_s\ cz$) and discordant pairs of individuals ($hc\_v\ s\_s\ cz$). Delineate differences would indicate the possibility to correctly group individuals. However, the gathered differences are remarkably low – less than 1% for every channel –, confirming the difficulty of discriminating true pairs of individuals. Despite the nearly absent differences, cosine distance achieves higher percentage differences than the Euclidean distance, motivating its choice for the contrastive loss.

To assess the proposed contributions, classification results were collected using the extracted features from the developed SNN, and compared with state-of-the-art classifiers developed by Schirrmeister [24], Charles [27] and Lawhern et al. [23]. We further compare our approach against classifiers able to learn directly from spectral/FFT features extracted from each channel [39]. The EEG classifiers proposed in previous works are referred to as: (vi) EEGNet, (vii) EEGNet-SSVEP, (viii) Riemann, (ix) DeepConvNet, (x) ShallowConvNet. The FFT features classifiers are referred to as: (i) FFT-kNN, (ii) FFT-NB, (iii) FFT-RF, (iv) FFT-SVM, (v) FFT-XGB. The proposed classifiers based on the SNN extracted features are referred to as: (xi) DSTFT-SNN-kNN, (xii) DSTFT-SNN-NB, (xiii) DSTFT-SNN-RF, (ix) DSTFT-SNN-SVM, (x) DSTFT-SNN-XGB. Fig. 2 provides the accuracy, sensitivity, specificity and the Matthews correlation coefficient (MCC) collected for each approach under a 10-fold cross-validation scheme. The Matthews correlation coefficient (MCC) [38] varies between $[-1, 1]$ and offers a combined view of sensitivity and specificity scores, measuring the significance of the results in face of the available number of data instances and the different sizes of the case and control populations. In addition, all results were statistically compared under a t-test in order to assess the statistical significance of the improvements on performance. According to Table 2, the SNN features outperform the baselines considered by an average of 20pp both in accuracy, specificity and sensitivity. The distinctive performance of SNNs in terms of sensitivity and specificity, respectively indicates that the proposed approach is able to simultaneously minimize false negatives (schizophrenia individuals diagnosed as healthy) and false positives (healthy individuals diagnosed as schizophrenic). All the collected differences are statistically significant under significance thresholds below $1E-5$.

The results observed when considering FFT features underline the difficult nature of the problem at hands, showing that the use of spectral

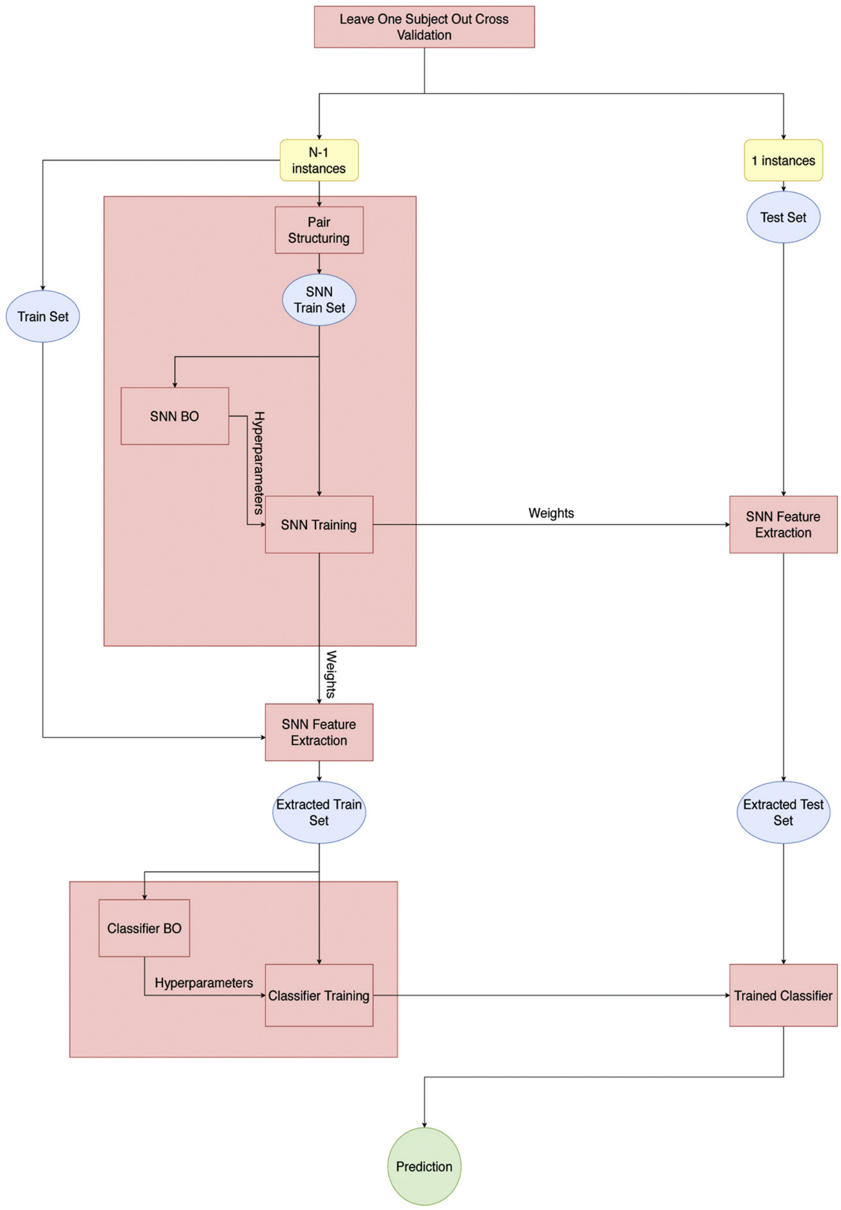
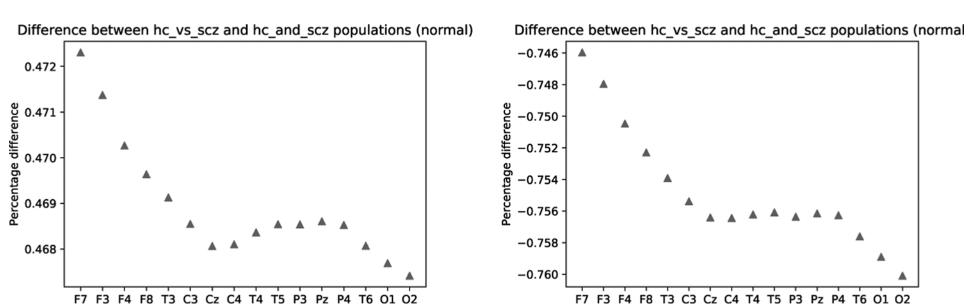


Fig. 3. Schematic representation of the proposed validation procedure: SNN feature extraction (Section 3.2.2) and classification (Section 3.3).

features is not sufficient to capture discriminative electrophysiological brain patterns.

As previously mentioned in Section 2, the previous work on EEG data classification – referred in Table 2 as (vi), (vii), (viii), (ix) and (x) – is unable to capture neuroplasticity differences between healthy and schizophrenia individuals from resting state data. These approaches are mainly prepared to detect evoked potentials in response to specific stimuli, thus generally neglecting subtle, spontaneous



(a) Euclidean Distance. (b) Cosine Distance.

Fig. 5. Distance type comparison on hc_and_scz and hc_vs_scz sets.

Table 2
Comparison of classifiers based on discriminative spectral features, state-of-the-art EEG data classifiers, and the proposed SNN-based classifiers. Sensitivity (specificity) refers to the proportion of actual schizophrenic (healthy control) individuals correctly classified. Accuracy refers to the proportion of schizophrenic and healthy controls correctly classified. MCC refers to the Matthews correlation coefficient [38], which allows us to analyze the significance of our results based on the number of available data instances and the different sizes of the case and control populations. Bold results are the best performers.

Classifier	Accuracy	Sensitivity	Specificity	MCC
(i) FFT-kNN	0.60 ± 0.31	0.56 ± 0.33	0.64 ± 0.30	0.17
(ii) FFT-NB	0.57 ± 0.32	0.33 ± 0.38	**0.85 ± 0.14**	0.18
(iii) FFT-RF	0.58 ± 0.32	0.58 ± 0.32	0.64 ± 0.29	0.19
(iv) FFT-SVM	0.66 ± 0.28	0.69 ± 0.26	0.63 ± 0.29	0.30
(v) FFT-XGB	0.65 ± 0.28	0.68 ± 0.26	0.61 ± 0.30	0.26
(vi) EEGNet [23]	0.58 ± 0.32	0.58 ± 0.31	0.59 ± 0.32	0.17
(vii) EEGNet-SSVEP [23]	0.54 ± 0.34	0.60 ± 0.31	0.46 ± 0.37	0.04
(viii) Riemann [27]	0.41 ± 0.50	0.47 ± 0.54	0.44 ± 0.50	−0.10
(ix) DeepConvNet [24]	0.54 ± 0.12	0.64 ± 0.08	0.41 ± 0.14	0.01
(x) ShallowConvNet [24]	0.57 ± 0.32	0.58 ± 0.31	0.56 ± 0.32	0.12
(xi) DSTFT-SNN-kNN	0.88 ± 0.12	0.90 ± 0.09	0.85 ± 0.14	0.74
(xii) DSTFT-SNN-NB	0.83 ± 0.16	0.82 ± 0.16	0.83 ± 0.15	0.62
(xiii) DSTFT-SNN-RF	0.88 ± 0.11	0.93 ± 0.07	0.82 ± 0.16	0.71
(ix) DSTFT-SNN-SVM	0.87 ± 0.12	0.96 ± 0.04	0.78 ± 0.20	0.74
(x) **DSTFT-SNN-XGB**	**0.95 ± 0.05**	**0.98 ± 0.02**	**0.92 ± 0.07**	0.88

electrophysiological variations in the brain of individuals.

In contrast, the combined use of DSTFT representations with the proposed SNNs are better prepared to detect neuroplasticity characteristics on the EEG signal as motivated by the rich spectral content inputted to the SNN, the properties of the entailed transformations, and the discriminative power of the features outputted from the SNN. These observations are experimentally demonstrated by the results presented in Table 2, with a significant difference between our approach and the previous work on EEG.

Among the classifiers applied to the SNN features, XGBoost yields the better performance, followed by RFs, SVMs with sparse kernel and kNNs. We hypothesize that this observation is primarily driven by the compositional value of the extracted features and the heterogeneity of individual profiles. Understandably, since only a part of the overall features have discriminative value for a given subject due to profile heterogeneity, NB and kNN have an understandable lower performance due to their inherent inability to discard non-relevant features. Similarly, when we compare performance of the classifiers from FFT features, FFT-kNN and FFT-NB have a slightly inferior performance against FFT-XGB and FFT-SVM. Among the five classifiers, all were slightly better at discriminating schizophrenic individuals (sensitivity) than discriminating healthy controls (specificity) due to an inherent ability to avoid false negatives.

Fig. 6a and b complements this analysis by offering a view on the discriminative power of the features produced by the SNN right before the remaining classification step. To this end, under a stratified 80/20 training-testing split, we trained a SNN and then used the learned SNN to assess the discriminative power the extracted features for both the individuals used to train the SNN and for the individuals in the testing set. We further show the distribution of values for healthy controls and schizophrenic individuals, as well as the statistical significance of the differences for each feature.

The provided results in Fig. 6a and b confirm the ability of the proposed SNNs to learn a discriminative feature space. 10 out 12 features hold statistically significant differences between the case and control individuals used to train the neural network. The same SNN further shows a distinctive ability to diagnose unseen individuals from a testing set, as 8 out of the 12 features are able to discriminate schizophrenic from healthy individuals with statistical significance ($\alpha < 0.05$).

Despite the highlighted discriminative power of features, note that the subsequent classification step is an essential part of the proposed approach since it is during this step that cross-channel dependencies are detected to capture information transmission among different cortical regions. The analysis of the of propagation of values from the input nodes of the SNN can be used to inquiry neuroplasticity aspects. Illustrating, the analysis of input neurons corresponding to areas of the spectral image that encode high-frequency beta activity can be further used to assess whether deep learning is sensitive to desynchronized beta activity, a condition associated to individuals with genetic predispositions to schizophrenia [40].

In addition, the learned classification models from the SNN features produced for each channel can be analyzed to inquiry aspects of information transmission between different cortical regions. The relevance of multi-channel dependencies in the context of schizophrenia has been previously highlighted [41]. In line with previous findings [41], our classifiers generally attribute higher importance to features collected from the T5 and C3 electrodes. The analysis of the classification models can be further considered testing whether left hemispheric hypotemporality and the increased interhemispheric information transmission in temporal lobe is present, as it underlines a neuroplasticity deficit found in schizophrenic patients [41,42].

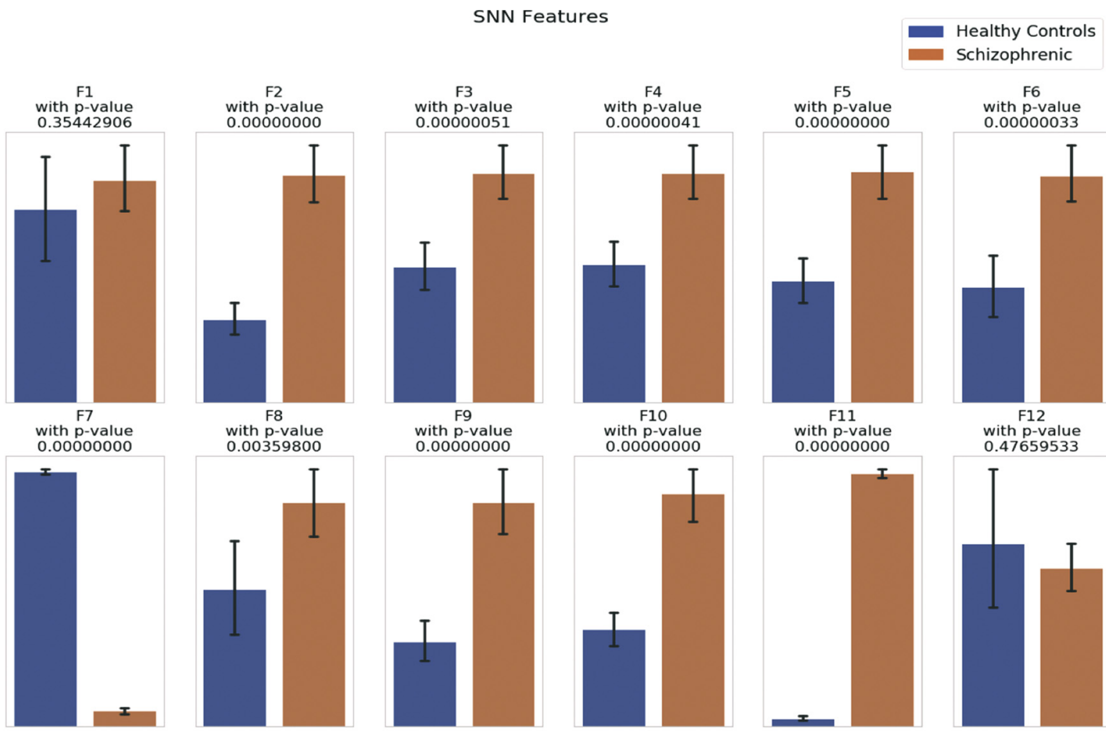
Considering an i7-8550U CPU @ 1.88GHz processor with 8GB RAM, the computational time for performing a diagnostic (testing a new individual) was found to be below 0.01 s. Once hyperparameters are fixed, training the top-performing DSTFT-SNN-XGB classifier from scratch on the target population is achieved in less than 60 s.

In summary, the gathered results confirm the relevance of working in a pairwise distance space to guarantee a good generalization ability. In addition, the applied convolution transformations guarantee a sensitivity to the inherently rich spatial, temporal and spectral nature of the EEG signal. We hypothesize that these aspects, together with the use of regularization and the cosine loss function (able to favor variations over absolute differences in the spectral content), explain the ability to learn extremely discriminative features.

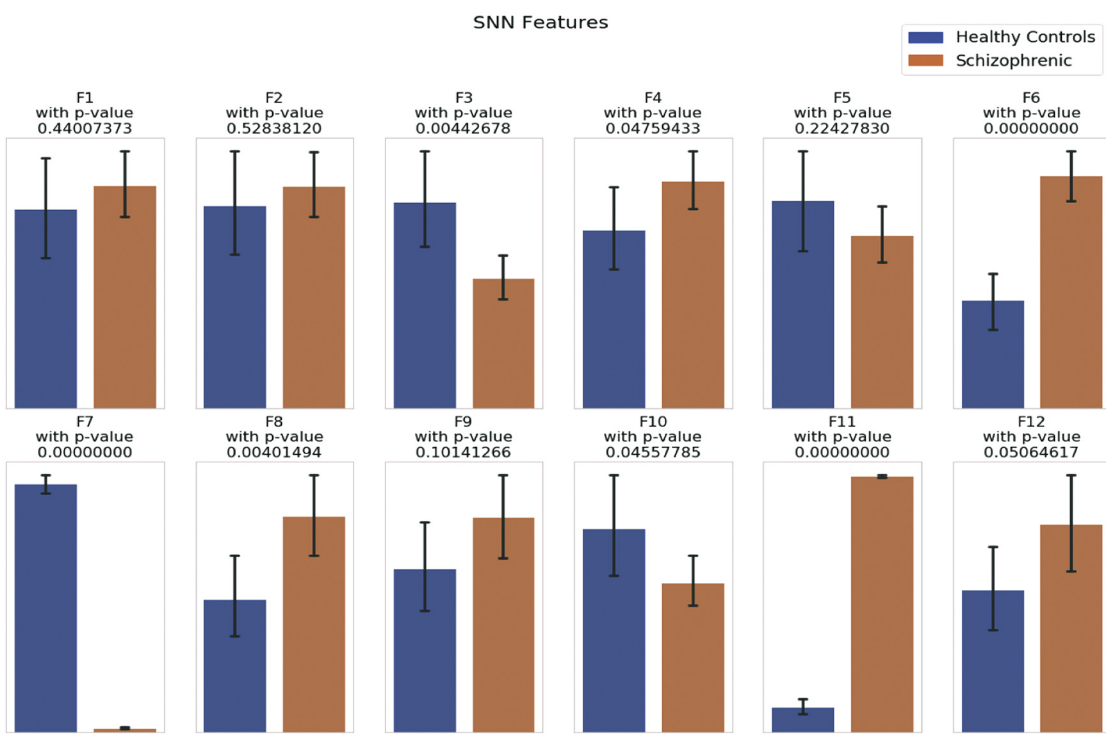
5. Conclusion

The rich nature of the electrophysiological data measured at the cerebral cortex makes deep learning a natural candidate to study disorders disrupting the normal brain activity. Nevertheless, the limited size of case-control populations, together with the inherent variability of the spectral content within and among individuals, has left the value of neural network approaches largely unexplored. This manuscript stresses the relevance of revisiting this problem, showing that adequately reshaped neural networks with proper loss and regularization criteria can increase the accuracy of schizophrenia diagnostics by 15–20 percentage points against peer alternatives (without hampering sensitivity or specificity).

Two master principles underlie these results: (1) the mapping of the



(a) Mean and standard deviation of each SNN feature for healthy controls and schizophrenic individuals in a training set.



(b) Mean and standard deviation of each SNN feature for healthy controls and schizophrenic individuals in a test set.

Fig. 6. Statistical analysis of the SNN features in a 80/20 train and test setting. The hyperparameters were obtained from a random fold of the LOOCV.

original data space into a pairwise distance space to support data augmentation while enhancing the discriminative power of the output features; and (2) the exploration of the rich nature of brain patterning through convolution operations on the spectral imaging of the signal, with weights learned under a cosine loss to improve robustness against the inherent noisy nature of electrophysiologic data. Results suggest that the proposed Siamese neural networks decode structural neuroplasticity differences between healthy and schizophrenic individuals, an observation corroborated by previous studies on how neuroplasticity properties are encoded in the frequency domain of EEG signals from individuals with schizophrenia [43,44]. Furthermore, this property opens up a new door for the guided application of neurofeedback therapies, similarly to the therapeutic settings discussed in [1]. In therapeutic settings, incentives can be given to counteract the differences from the gathered SNN features against control feature expectations. To this end, the neurofeedback system should represent the EEG signal in the frequency domain, by means of the DSTFT; process its spectral content using the proposed deep learning architecture; and output the values for the most critical SNN features against expectations gathered from control individuals. As future work, we aim to extend the experimental analysis toward alternative disorders, and different EEG instrumentation or protocols; contrast the performance of the proposed EEG-based learners against state-of-the-art MRI- and PET-based learners on a population of individuals with (and without) neurodegenerative conditions being currently monitored at Instituto de Medicina Molecular; and to establish a method that is capable of performing neurofeedback techniques to tackle schizophrenia symptoms, similarly to what has been previously proposed by Nan et al. [1].

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] Nan W, Rodrigues JP, Ma J, Qu X, Wan F, Mak P-I, et al. Individual alpha neurofeedback training effect on short term memory. Int J Psychophysiol 2012;86(1):83–7. https://doi.org/10.1016/j.ijpsycho.2012.07.182.

[2] Ataei A, Amini A, Ghazizadeh A. Brain activity estimation using EEG-only recordings calibrated with joint EEG-FMRI recordings using compressive sensing. 2019 13th international conference on sampling theory and applications (SampTA) 2019:1–3.

[3] Owen MJ, Sawa A, Mortensen PB. Schizophrenia. Lancet 2016;388(10039):86–97. https://doi.org/10.1016/S0140-6736(15)01121-6.

[4] Nan W, Wan F, Chang L, Pun SH, Vai MI, Rosa A. An exploratory study of intensive neurofeedback training for schizophrenia. Behav Neurol 2017;2017.

[5] Zhang Y, Quiñones GM, Ferrarelli F. Sleep spindle and slow wave abnormalities in schizophrenia and other psychotic disorders: recent findings and future directions. Schizophr Res 2019. https://doi.org/10.1016/j.schres.2019.11.002.

[6] Dvey-Aharon Z, Fogelson N, Peled A, Intrator N. Connectivity maps based analysis of EEG for the advanced diagnosis of schizophrenia attributes. PLoS One 2017;12(10):e0185852.

[7] Sabeti M, Katebi S, Boostani R. Entropy and complexity measures for EEG signal classification of schizophrenic and control participants. Artif Intell Med 2009;47(3):263–74. https://doi.org/10.1016/j.artmed.2009.03.003.

[8] Howells FM, Temmingh HS, Hsieh JH, van Dijen AV, Baldwin DS, Stein DJ. Electroencephalographic delta/alpha frequency activity differentiates psychotic disorders: a study of schizophrenia, bipolar disorder and methamphetamine-induced psychotic disorder. Transl Psychiatry 2018;8(1):75.

[9] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.

[10] Lopes da Silva F. EEG and meg: relevance to neuroscience. Neuron 2013;80(5):1112–28. https://doi.org/10.1016/j.neuron.2013.10.017.

[11] Gorbachevskaya K, Borisov S. EEG of healthy adolescents and adolescents with symptoms of schizophrenia. 2002 [accessed 01.02.19]. http://brain.bio.msu.ru/eeg_schizophrenia.htm.

[12] Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. ICML deep learning workshop, vol. 2 2015.

[13] Riaz F, Hassan A, Rehman S, Niazi IK, Dremstrup K. EMD-based temporal and spectral features for the classification of EEG signals using supervised learning. IEEE Trans Neural Syst Rehabil Eng 2016;24(1):28–35. https://doi.org/10.1109/TNSRE.2015.2441835.

[14] Zhang P, Wang X, Zhang W, Chen J. Learning spatial-spectral-temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment. IEEE Trans Neural Syst Rehabil Eng 2019;27(1):31–42. https://doi.org/10.1109/TNSRE.2018.2884641.

[15] Fazli S, Danóczy M, Schelldorfer J, Müller K-R. L1-penalized linear mixed-effects models for high dimensional data with application to BCI. NeuroImage 2011;56(4):2100–8. https://doi.org/10.1016/j.neuroimage.2011.03.061.

[16] Destrero A, Mosci S, De Mol C, Verri A, Odone F. Feature selection for high-dimensional data. Comput Manag Sci 2009;6(1):25–40.

[17] Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. Advances in neural information processing systems 2012:2951–9.

[18] Bhandari A, Voineskos D, Daskalakis ZJ, Rajji TK, Blumberger DM. A review of impaired neuroplasticity in schizophrenia investigated with non-invasive brain stimulation. Front Psychiatry 2016;7:45.

[19] Liu KK, Bartsch RP, Lin A, Mantegna RN, Ivanov PC. Plasticity of brain wave network interactions and evolution across physiologic states. Front Neural Circuits 2015;9:62.

[20] Shaikhina T, Khovanova NA. Handling limited datasets with neural networks in medical applications: a small-data approach. Artif Intell Med 2017;75:51–63. https://doi.org/10.1016/j.artmed.2016.12.003.

[21] Barz B, Denzler J. Deep learning on small datasets without pre-training using cosine loss. 2019. arXiv:1901.09054.

[22] Hu G, Peng X, Yang Y, Hospedales TM, Verbeek J. Frankenstein: learning deep face representations using small data. IEEE Trans Image Process 2017;27(1):293–303.

[23] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. J Neural Eng 2018;15(5):056013http://stacks.iop.org/1741-2552/15/i=5/a=056013.

[24] Schirrmeister RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggensperger K, Tangermann M, et al. Deep learning with convolutional neural networks for EEG decoding and visualization. Hum Brain Mapp 2017;38(11):5391–420.

[25] Dvey-Aharon Z, Fogelson N, Peled A, Intrator N. Schizophrenia detection and classification by advanced analysis of EEG recordings using a single electrode approach. PLoS One 2015;10(4):e0123033.

[26] Niu W, Huang X, Xu K, Jiang T, Yu S. Pairwise interactions among brain regions organize large-scale functional connectivity during execution of various tasks. Neuroscience 2019;412:190–206. https://doi.org/10.1016/j.neuroscience.2019.05.011.

[27] Charles P. pyriemann. 2013 [accessed 22.10.19]. https://github.com/alexandrebarachant/pyRiemann.

[28] Sengupta S, Singh A, Leopold HA, Gulati T, Lakshminarayanan V. Ophthalmic diagnosis using deep learning with fundus images – a critical review. Artif Intell Med 2020;102:101758. https://doi.org/10.1016/j.artmed.2019.101758.

[29] Ieracitano C, Mammone N, Bramanti A, Hussain A, Morabito FC. A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings. Neurocomputing 2019;323:96–107. https://doi.org/10.1016/j.neucom.2018.09.071.

[30] Oh SL, Vicnesh J, Ciaccio EJ, Yuvaraj R, Acharya UR. Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals. Appl Sci 2019;9(14):2870.

[31] Lih OS, Jahmunah V, San TR, Ciaccio EJ, Yamakawa T, Tanabe M, et al. Comprehensive electrocardiographic diagnosis based on deep learning. Artif Intell Med 2020;103:101789. https://doi.org/10.1016/j.artmed.2019.101789.

[32] Bromley J, Guyon I, LeCun Y, Säckinger E, Shah R. Signature verification using a “Siamese” time delay neural network. Advances in neural information processing systems 1994:737–44.

[33] Petrie A, Sabin C. Medical statistics at a glance. John Wiley & Sons; 2019.

[34] Hahnloser RH, Sarpeshkar R, Mahowald MA, Douglas RJ, Seung HS. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature 2000;405(6789):947.

[35] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR’06), vol. 2 2006:1735–42.

[36] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15:1929–58http://jmlr.org/papers/v15/srivastava14a.html.

[37] Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. arXiv preprint arXiv:1412.6980.

[38] Matthews B. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta 1975;405(2):442–51. https://doi.org/10.1016/0005-2795(75)90109-9.

[39] Hindarto H, Sumarno S. Feature extraction of electroencephalography signals using fast Fourier transform. CommIT J 2016;10(2):49–52.

[40] Itil TM. Qualitative and quantitative EEG findings in schizophrenia. Schizophr Bull 1977;3(1):61.

[41] Na SH, Jin S-H, Kim SY, Ham B-J. EEG in schizophrenic patients: mutual information analysis. Clin Neurophysiol 2002;113(12):1954–60. https://doi.org/10.1016/S1388-2457(02)00197-9.

[42] Bakhshi K, Chance S. The neuropathology of schizophrenia: a selective review of past studies and emerging themes in brain structure and cytoarchitecture. Neuroscience 2015;303:82–102. https://doi.org/10.1016/j.neuroscience.2015.06.028.

[43] Wynn JK, Roach BJ, McCleery A, Marder SR, Mathalon DH, Green MF. Evaluating visual neuroplasticity with EEG in schizophrenia outpatients. Schizophr Res 2019;212:40–6. https://doi.org/10.1016/j.schres.2019.08.015.

[44] Zomorrodi R, Rajji T, Blumberger D, Daskalakis Z. The association between cross-frequency coupling and neuroplasticity via paired associative stimulation: TMS-EEG study. Brain Stimul 2019;12(2):512. https://doi.org/10.1016/j.brs.2018.12.680.