Schizophrenia Diagnosis using Electroencephalographic Data

David Calhas
Instituto Superior Técnico
Lisbon, Portugal
david.calhas@tecnico.ulisboa.pt

Abstract

The increasing access to brain signal data using electroencephalography creates new opportunities to study electrophysiological brain activity and perform ambulatory diagnoses of neuronal diseases. This work proposes a pairwise distance learning approach for Schizophrenia classification relying on the spectral properties of the signal. Given the limited number of observations (i.e. the case and/or control individuals) in clinical trials, 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. Results on a case-control population show that the features extracted using the proposed neural network lead to an improved Schizophrenia diagnosis (+10pp in accuracy and sensitivity) against spectral features, thus suggesting the existence of non-trivial, discriminative electrophysiological brain patterns. An analysis of the electroencephalographic signal is also done, in order to motivate the use of more complex techniques.

1 Introduction

The recording of increasingly affordable and precise electroencephalographic (EEG) data is creating unprecedented opportunities to understand brain activity, aid personalized prognostics, and promote health through wearable biofeedback systems [Nan et al., 2012]. 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, risks and restrictions on the periodicity of recordings [Fowle and Binnie, 2000]. 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 [Owen et al., 2016].

Despite the inherent advantages of monitoring electrophysiological brain activity, its use for diagnosing neuronal diseases is still capped by the limited size of case-control populations [Litjens et al., 2017], as well as by the intrinsic difficulties of mining brain signals. Brain signal data is high-dimensional, multivariate, susceptible to noise/artefacts, rich in temporal-spatial-spectral content, and highly-variable between individuals [da Silva, 2013].

This work proposes a dedicated class of neural networks to extract discriminative features of Schizophrenia from electrophysiological brain data. 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 extract features from a limited number of observations of a possible data distribution under study [Gorbachevskaya and Borisov, 2002]. The features produced by these networks have proven to be useful to perform classification as they rely on either the homologous or discriminative properties of observation-pairs in a pairwise distance domain [Koch et al., 2015];

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, 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 under a cosine-loss. Gathered evidence show less susceptibility to artefacts and the inherent variability of electrophysiological potentials composed of continuously changing overlapping electrical fields created by localized neurons [da Silva, 2013];

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 weights of shared connections in the network;

5. Ability to handle the extremely-high dimensional nature of the gathered spectral content of brain signal data (high-resolution spectral image per scalp location) under L1 regularization;

6. Applicability of the proposed EEG-based diagnostics to alternative populations or diseases, evidenced by the: i) placed Bayesian optimization step [Snoek et al., 2012] for hyperparameter tuning and fixing feature numerosity; 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 traditional stance to 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, often motivated by disease rarity, capped size of control population, trial eligibility requirements, or the facultative nature of EEG assessments. It also dives into a statistical analysis complemented by a clustering exploration.

The manuscript is organized as follows. After formalizing the problem, Section 2 surveys existing contributions on 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 Description

Problem. A EEG recording or brain signal observation is a multivariate time series \( X = \{ x^k_j \mid k \in \{1..m\}, j \in \{1..t\} \} \), where \( x^k_j \) is a measure of the electrophysiological activity in scalp channel \( k \) and instant \( j \), \( t \) is the number of time points, and \( m \) is the multivariate order (number of channels). Given a brain signal dataset \( \{(X_i, c_i) \mid i = 1..N\} \), i.e. \( N \) EEG recordings \( X_i \) annotated with a label \( c_i \in \Sigma \), we aim to identify a discriminative feature space to classify (unlabeled) observations. Specifically, we are interested in the classification of Schizophrenia against case-control populations.

Background. The electrophysiological signal produced by a specific channel \( k \) 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, time and frequency.

2 Related Work

Dvey-Aharon et al. [2015] claim mostly changes in functional connectivity are seen in schizophrenia patients, as well as differences in theta-frequency activity. A classification approach was applied
on 1-minute signals recorded by a single electrode. The developed system consists of four stages: performing several 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 specific time frames following a given set of stimuli between the time/frequency matrix representations of the healthy subjects and the schizophrenia patients. The results obtained for the model were: 88.7% ± 4% discrimination accuracy rate, 100% specificity rate, 77.4% ± 7.1% sensitivity rate and 0.0078 p-value significance. Despite promising results, the approach requires the performance of cognitive tasks by the individuals under assessment throughout the recording. \cite{2017} introduced another way of looking at the EEG signal using connectivity maps derived from the brain activity. 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. In terms of accuracy, the connectivity maps method achieved 93.8% ± 4%; 100% specificity rate; 87.6% ± 5.1% sensitivity rate; and 0.0041 significance p-value. 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.

\cite{2009} introduces another approach to classify schizophrenia was employed 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, two algorithms were implement to perform the classification: Adapative boost (Adaboost) and Linear Discriminant Analysis (LDA). Both algorithms were validated with and without the feature selection step. The recordings were done with eyes open. Not relying on the definition of a task, but it can still be biased by the environment surrounding the patient. Without feature selection, the accuracy reached 0.8578 ± 0.1207 and 0.8961 ± 0.1077, for LDA and Adaboost respectively. With Genetic programming for feature selection, the accuracy reached 0.8875 ± 0.0946 and 0.9126 ± 0.0924. These results were obtained using leave-one-out cross validation.

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. 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 cohorts and the inherent stochastic complexities associated with electrophysiological data.

### 2.1 Siamese Neural Network

First introduced by \cite{1994}, as a novel model used in the task of signature classification whose aim was to distinguish signature forgeries from the real ones, Siamese Neural Networks (SNN) 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, \cite{2015} 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.
3 Our Approach

Our architecture is inspired by the model proposed in Koch et al. [2015]. 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. The proposed 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) [Snoek et al., 2012].

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 consider to monitor the underlying brain patterning at the brain cortex, independently of the surrounding environment/undertaken task. Subsequently, this avoids any additional interference on the EEG signal recorded. Howells et al. [2018] 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.

The target dataset consists of a total of 84 individuals: 39 healthy controls and 45 schizophrenia subjects. 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 minute 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. The dataset is available online, sourced from Gorbachevskaya and Borisov, 2002.

3.2 Data analysis and Clustering

In order to perform a separation of the population in this dataset, a statistical analysis was performed taking into account the frequency domain of the EEG signal, as well as Hierchical Clustering methods, such as Dendogram and Community Finding algorithms (Girvan Newman) to build partitions of the population.

3.3 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.

![Base network from the SNN.](image)

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 minute recording of a channel of an individual. The DSTFT is taken with 2 seconds length windows in order to capture frequencies as low as 0.5Hz, corresponding to the delta wave frequencies (in Howells et al. [2018], it is pointed out that frequencies lower than 2Hz are relevant to differentiate Schizophrenia). 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 [Hahnloser et al., 2000], while the fully connected
layer uses the softmax activation function, normalizing the domain of the feature representations, $f \in \mathbb{R}^q, i \in [1, q] : f_i \in [0, 1]$.

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.

![Figure 2: SNN architecture.](image)

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 Hadsell et al. [2006], the Contrastive Loss function is introduced to that end, defined as:

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

where $(X_1, X_2)$ is the input pair, $Y = 1$ if $X_1$ and $X_2$ are neighbors and $0$ otherwise, $D_W$ the distance between the predicted values of $X_1$ and $X_2$, and $m$ is the margin value of separation. Minimization of 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.

The distance metric considered in our case is the cosine distance. This metric was chosen with the belief that a pattern based metric (cosine) would perform better than a magnitude based one.
We want the network to learn a valid transformation that generalizes to all channels. To that end, the features 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 was performed with a euclidean metric, in order to shed light on how the schizophrenia pathology expresses itself through the EEG.

Besides the type of layers and the distance metric, the following techniques are integrated in the model: L1 regularization and Dropout layers. The L1 regularization is useful because it helps remove features that are not useful for the 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 Srivastava et al. [2014], and is applied after each convolutional layer. Adam [Kingma and Ba 2014] is used to optimize the network during the training session.

3.3.1 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 250 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 whole dataset. 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 $\in [10^{-5}, 10^{-3}]$, 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 $\in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}$. 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 $[100.0, 500.0]$.

3.3.2 Pairwise Dataset Structure

We want the network to learn a valid transformation that generalizes to all channels. To that end, the pairs are set such that only equal channels are paired. Pairs of different channels are not considered, since we see different channels as correlated spaces with different properties. Fortunately, the SNN is capable of learning different spaces/classes, as shown in Koch et al. [2015], where the proposed system is able to learn a similar setup. This pairwise schema can be seen as a data augmentation technique being performed with the addition of noise to the dataset. This noise is present by mixing all 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 by the creation of a whole new optimization space where the variability contained in the data can be exploited in a different way.

3.4 Validation

After the SNN has been 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 was performed with a...
Leave-One-Out Cross Validation (LOOCV). For each of these classifiers, BO hyperparameter tuning is also performed, setup with a maximum of 10 acquisitions and 5 iterations for initial exploration. 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 $k \in \{2, 3, 4, 5, 6, 7, 8\}$

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

For the sake of comparison, the results of the classification obtained using the extracted features from the SNN are compared with the results of the classification using a competitive baseline Fast Fourier Transform (FFT) [Welch, 1967] frequency features, by performing a BO for each classifier in the same conditions described in 3.4. Further, a comparison with a Convolutional Neural Network (CNN) optimized for classification is made. This CNN differs from the BN, with the addition of a ReLU layer in the end to output a single label/classification value. The hyperparameters are obtained using BO, through use of the SNN hyperparameters domains (except for the margin value which is not included in this CNN architecture).

The FFT-based feature classifiers are referred to as: (i) FFT-kNN, (ii) FFT-NB, (iii) FFT-RF, (iv) FFT-SVM, (v) FFT-XGB. The CNN classifier is referred to as (vi) DSTFT-CNN. The proposed classifiers based on the SNN extracted features are referred to as: (vii) DSTFT-SNN-kNN, (viii) DSTFT-SNN-NB, (ix) DSTFT-SNN-RF, (x) DSTFT-SNN-SVM, (xi) DSTFT-SNN-XGB.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) FFT-kNN</td>
<td>0.60 ± 0.31</td>
<td>0.56 ± 0.33</td>
<td>0.64 ± 0.30</td>
</tr>
<tr>
<td>(ii) FFT-NB</td>
<td>0.57 ± 0.32</td>
<td>0.33 ± 0.38</td>
<td><strong>0.85 ± 0.14</strong></td>
</tr>
<tr>
<td>(iii) FFT-RF</td>
<td>0.66 ± 0.28</td>
<td>0.69 ± 0.26</td>
<td>0.62 ± 0.30</td>
</tr>
<tr>
<td>(iv) FFT-SVM</td>
<td>0.66 ± 0.28</td>
<td>0.69 ± 0.26</td>
<td>0.63 ± 0.29</td>
</tr>
<tr>
<td>(v) FFT-XGB</td>
<td>0.69 ± 0.26</td>
<td>0.76 ± 0.21</td>
<td>0.61 ± 0.31</td>
</tr>
<tr>
<td>(vi) DSTFT-CNN</td>
<td>0.76 ± 0.21</td>
<td>0.76 ± 0.21</td>
<td>0.77 ± 0.20</td>
</tr>
<tr>
<td>(vii) DSTFT-SNN-kNN</td>
<td>0.82 ± 0.16</td>
<td>0.84 ± 0.14</td>
<td>0.79 ± 0.18</td>
</tr>
<tr>
<td>(viii) DSTFT-SNN-NB</td>
<td>0.83 ± 0.16</td>
<td>0.84 ± 0.14</td>
<td>0.81 ± 0.17</td>
</tr>
<tr>
<td>(ix) DSTFT-SNN-RF</td>
<td><strong>0.84 ± 0.14</strong></td>
<td><strong>0.87 ± 0.12</strong></td>
<td><strong>0.82 ± 0.17</strong></td>
</tr>
<tr>
<td>(x) DSTFT-SNN-SVM</td>
<td>0.82 ± 0.16</td>
<td>0.84 ± 0.14</td>
<td>0.79 ± 0.18</td>
</tr>
<tr>
<td>(xi) DSTFT-SNN-XGB</td>
<td>0.82 ± 0.16</td>
<td>0.84 ± 0.14</td>
<td>0.79 ± 0.18</td>
</tr>
</tbody>
</table>

According to Table[1], the SNN features outperform the FFT frequency baselines considered by an average of 19pp in accuracy. One particular observation is that among the FFT features there was a clear difference between classifiers: FFT-kNN and FFT-NB were worse than FFT-XGB, FFT-RF and FFT-SVM. The same was not found among the SNN features, with all of the variants performing comparably. This is due to most of the discriminative work of the problem being solved by the SNN transformation of the DSTFT representation of the EEG signal. Suggesting that the SNN alone is capable of generalizing better and have extremely discriminative features.
Further, as expected the DSTFT-CNN baseline underperformed by an average of -10pp than all of the classifiers that used SNN features. The latter is justified by the low amount of data provided to train the model, which in contrast is much smaller than the amount of data used to train the SNN model. Adding to this, the high dimensionality of the DSTFT features is a characteristic that DSTFT-CNN is not originally prepared to handle. On the other hand, the regularization techniques still present in the DSTFT-CNN baseline were key factors to outperform the FFT baselines.

![Figure 3: Performance of each classification model per channel. The SNN feature classifiers are plotted in green. The FFT feature classifiers are plotted in red. The CNN classifier is plotted in blue.](image)

In contrast with the Table 1, Figs. 3a, 3b and 3c compare the performance of the assessed classifiers for each channel individually. We can observe that, independently of the channel, SNN classifiers still outperform the baseline classifiers considered. The classification performance is considerably stable among channels and a few percentage points lower than the original feature space of all channels. The classifiers found to have more variability between channels were: FFT-XGB, FFT-RF, SNN-XGB and SNN-RF.

Further, regarding the analysis and the hierarchical clustering techniques performed, the results were very poor and no coherent conclusions were able to be drawn out of it.

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, had left the value of neural network approaches largely unexplored. This manuscript stresses the relevance of revisiting this problem. By reshaping the architecture, loss and applied regularization, we show that the use of neural networks to classify Schizophrenia can increase the accuracy of diagnostics by 15-to-20 percentage points against peer alternatives (without hampering sensitivity or specificity). Two master principles underlie these results: 1) the mapping of the 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 better account for the inherent noise of electrophysiologic data.

As future work, we aim to extend the experimental analysis towards alternative disorders and populations with potentially different EEG instrumentation or protocol; contrast the performance of the proposed EEG-based learners against state-of-the-art brain imaging learners on a population of individuals with (and without) neurodegenerative conditions being currently monitored; and to establish a method that is capable of performing a neurofeedback technique to tackle Schizophrenia symptoms, similarly to what has been previously proposed Nan et al. [2012].

Acknowledgments

This work is supported by national funds through FCT under project iLU DSAIPA/DS/0111/2018 and INESC-ID pluriannual UID/CEC/50021/2019.
References


