EEG to fMRI Synthesis for Medical Decision Support: A Case Study on Schizophrenia Diagnosis

David Calhas
david.calhas@tecnico.ulisboa.pt
INESC-ID, Instituto Superior Tecnico
Lisbon, Portugal

Rui Henriques
rmch@tecnico.ulisboa.pt
INESC-ID, Instituto Superior Tecnico
Lisbon, Portugal

Abstract

Electroencephalography (EEG) measures the neuronal activity at the scalp, while functional magnetic resonance imaging (fMRI) provides a sub-cortical view of blood supply in the human brain. Although fMRI is known for providing rich spatial information, it is expensive and of restricted use. EEG to fMRI synthesis is a cross modal research area that bridges the gap between the two and has recently received attention. Although these studies promise lower healthcare costs and ambulatory assessments, their utility in diagnostic settings is still largely untapped. Using simultaneous EEG and fMRI recordings, this study combines a state-of-the-art synthesis model with a modified contrastive loss, and subsequent prediction layering, to unprecedentedly assess its predictive power in schizophrenia diagnosis. In addition, we perform an exhaustive search for the (synthesized) hemodynamic brain patterns able to discriminate schizophrenia. Schizophrenia diagnosis using synthesized hemodynamics yield an area under the ROC curve of 0.77, confirming the validity of the undertaken neuroimaging synthesis. Experiments further revealed schizophrenia-related patterns in frontal, left temporal and cerebellum regions of the brain. Altogether, our results suggest that a synthesized fMRI view is able to discriminate this pathology, and it contains discriminative patterns of brain activity in accordance with related work on schizophrenia.

Keywords Electroencephalography, Functional magnetic resonance imaging, Machine learning, Schizophrenia

1 Introduction

Schizophrenia is a neurological condition that affects a significant portion of the world population [1]. Yet, to this day, the diagnosis of schizophrenia generally involves several clinical tests, making it time consuming and exhausting for the patient. Diagnostic tests range from psychological symptoms to neuroimaging [2], molecular [3], and natural speech markers [4]. In particular, functional magnetic resonance imaging (fMRI) scans have shown potential for an automated diagnostic [5, 6, 7, 2], still their availability is limited. For instance, Ogbole et al. [8] report the density of MRI machines in Africa, claiming Nigeria to be the most critical country, where 0.3 MRI machines are made available for each one million people. As a consequence, the quality of these health care systems is low. Can alternative cheaper modalities be considered as a proxy to replace MRI machines? Recent advances in cross modal synthesis [9, 10, 11] show promise. And although, these do not enable the replacement for cost reductions, there are claims that electroencephalography (EEG) has predictive power over fMRI features [12, 13]. Opportunely, *EEG to fMRI synthesis* is an emerging research area that focuses on the synthesis of fMRI, an expensive modality, from EEG, a lower cost modality [14, 15, 16, 17]. In spite of these great efforts, we are yet to see their applicability in an health care setting.

The main idea of this paper is to show the ability to classify schizophrenia using a synthesized fMRI from EEG, ultimately enabling better interpretation, while ensuring competitive predictive accuracy. By doing this, we provide a prediction in a feature space with fine spatial resolution. In addition, to assess its fidelity, we compare the predictive value of fMRI estimates for the target schizophrenia classification task against (state-of-the-art) time-based and frequency-based EEG views. A view is defined as a set of related features with EEG as a source (see Section 1.2). In addition to the classification problem, the synthesized fMRI is subject to a data mining tool, known to extract biclusters [18]. These biclusters identify which areas of the brain are associated with schizophrenia. In a sense, EEG, which is recorded at the scalp level, is projected to an fMRI, where sub-cortical activity patterns are retrieved. Therefore, EEG to fMRI Synthesis enriches the EEG modality with fMRI learned features [17]. This additional information increases interpretability, which we measure through a pattern discovery algorithm. The contributions of this study are:

- integration of a state-of-the-art EEG to fMRI synthesis model [17] (see Section 1.2.3) in a schizophrenia classification setting (see Section 2);
- combination of a modified contrastive loss along with layer normalization, that together allow separation of data and maintain fMRI volume style (see Section 2.1);
- biclustering analysis of a synthesized modality, which allows the retrieval of statistical significant patterns for schizophrenic individuals (see Section 3.3).

We list the main findings of this work as:

- EEG frequency features are highly discriminative of schizophrenia, which goes in accordance with previous studies [19, 20]. Nonetheless, despite its predictive power, it lacks in interpretability. Restricting its use for medical diagnostics;
- the decision making of the linear classification is correlated with synthesized frontal lobe activity. It is known that the functions, related to this region, are impaired by schizophrenia [21, 22]. This suggests, that the classifier founded its predictions on frontal activity;
- the cerebellum region of the synthesized fMRI is present in several biclusters associated with schizophrenia. This finding goes in accordance with MRI studies on this pathology [23];
- schizophrenia diagnosis from the synthesized fMRI, in the absence of any other complementary clinical records, is able to reach the 0.77 AUC mark, offering preliminary evidence of the validity of the cross-modal synthesis process.

1.1 Problem description

Let $\mathcal{V} = \{v_1, \ldots, v_n\}$ be a set of different views, such that $\forall i \in \{1, \ldots, n\} : v_i = \{F_i, \theta_i\}$. A view v_i is characterized by a function

structure, F_i, and its parameters, θ_i. Each projection $F_i : \mathbb{R}^{D_0} \rightarrow \mathbb{R}^{D_i}$ is performed from the original feature space, $\mathbb{R}^{D_0}$, to its view space, $\mathbb{R}^{D_i}$. The original view is defined as $v_0 = \{I \cdot \vec{x}_0, I \in \mathbb{R}^{D_0}\}$, being I the identity matrix. Each view has a dedicated section (1.2.1, 1.2.2 and 1.2.3), where the structure of the view space, D_i, is defined. Each instance, $\vec{x}$, is paired with a label $y \in \{0,1\}^K : \sum_k y_k = 1$, where K is the number of classes described in the data. A view, v_i, optimizes its parameters, θ_i, in order to minimize

$$\mathcal{L}_C(\vec{x}_0, y) = -y \times \log(\sigma(W_f^i \cdot v_i(\vec{x}_0) + b_f^i)), \tag{1}$$

where $W_f^i \in \mathbb{R}^{K \times D_i}$ and $b_f^i \in \mathbb{R}^K$ are the parameters of a linear classifier, referred to as $\theta_C^i = W_f^i \bigcup b_f^i$, using the softmax activation, σ. Then the objective, for each view i, is defined as

$$\text{argmin}_{\theta_C^i} \mathcal{L}_C(\vec{x}_0, y). \tag{2}$$

1.2 Background: Defining the EEG views

This study is concerned with providing an accurate view that allows a straightforward interpretation. The interpretation of a prediction is typically made by a human being, with knowledge in the application domain [24]. However, when a prediction is made by a model, it should be supported with an explanation consisting on a subset of the features $\vec{x}_i$ that justify the decision of the model. In this section, we describe each EEG view that is considered in our experiments. Starting with the raw (channel-temporal) view, *raw*, moving to the frequency (channel-time-frequency) representation, referred to as *stft*, and finishing with the synthesized fMRI which is the *fmri* view. For each of these, we discuss their ability to provide good explanations, as well as their overall fit to discriminate schizophrenia.

1.2.1 raw view

In its raw form, an EEG recording consists of a set of electrodes, also called channels, that contain the electrical activity present at the scalp. This activity has its source at the neurons, where the action potentials occur. Formally this view is defined as $v_0 = \{F_0, \theta_0\}$, where $F_0 = I \cdot \vec{x}_0$ and $\theta_0 = I \in \mathbb{R}^{D_0}$.

With EEG being a multivariate time series representation, researchers are able to study functional properties of the brain [25]. However, the application of this specific view is mainly useful for epilepsy detection [26]. Structurally, this representation is defined as $\vec{x}_0 \in \mathbb{R}^{D_0}$, with $D_0 = C \times T$. C stands for the number of channels and T defines the temporal dimension. It is worth noting that this view is limited to the channel-temporal features, characterized by noise and difficult medical interpretation [27]. In consequence, it is unsuitable for diagnosis. In the scope of schizophrenia classification, even when processed by models with a high level of feature engineering [28], it falls short from other more fit views [29].

1.2.2 stft view

Research studies, along the years, have had a common trend of extracting time frequency features from the EEG signal. From functional connectivity relations [30, 31, 32] to statistical significance of bands associated with pathologies [20, 33], several studies have used the frequency domain to uncover new findings. A time frequency representation of EEG can be achieved via a transform [34]. In this study we consider the short time Fourier transform (STFT) [35], whose kernel consists on sine and cosine waves with different shifts and frequencies. As such this view is defined as $v_1 = \{F_1, \theta_1\}$, where F_1 is the sum of sinusoids and θ_1 are the sinusoids themselves.

Similarly to *raw*, this view is a multivariate time series representation, with the channel, C, and temporal, T, dimensions belonging to its structure. On top of that, each frequency, with the dimension F, is represented at different time steps and channels. Hence, the function consists of $F_1 : \mathbb{R}^{D_0} \rightarrow \mathbb{R}^{D_1}$ and the *stft* view is structurally defined as $\vec{x}_1 \in \mathbb{R}^{D_1}$, where $D_1 = C \times F \times T$. On the interpretation side, we consider it not just similar to the *raw* view, but also more limited due to the requirement of frequency domain knowledge. Though, its discriminative ability is one of the best for schizophrenia [29].

1.2.3 EEG to fMRI synthesis: fmri view

We now describe the EEG to fMRI synthesis model proposed by Calhas et al. [17]. It is a neural network, whose flow is illustrated in Figure 1. Its input is the EEG time-frequency representation view, $\vec{x}_1$. On the opposite side, the output is the *fmri* view, $\vec{x}_2 \in \mathbb{R}^{M_1 \times M_2 \times M_3}$, characterized by the referential axes dimensions (M_1, M_2 and M_3). Zooming out of the neural network, we can describe this view as $v_2 = \{F_2, \theta_2\}$. Where $F_2 = F_1 \bigcup F$ denotes the union of two functions F_1, the short time Fourier transform view $v_1 = \{F_1, \theta_1\}$, and F, the neural network. Accordingly, its parameters are $\theta_2 = \theta_1 \bigcup \theta_F$, where θ_F are the parameters of the neural network.

In terms of structure, this network processes the input with an attention mechanism. For this, an attention weight matrix $A \in \mathbb{R}^{C \times F \times T}$ is used to compute $E \in \mathbb{R}^{C \times C}$, where $\forall i \in \{1, \ldots, C\} : E_i = \left[e_{i1}^\top, \ldots, e_{iC}^\top\right] = \left[\vec{x}_i^\top \cdot A_1, \ldots, \vec{x}_i^\top \cdot A_C\right]$. Then E is softmax normalized, such that

$$\forall i \in \{1, \ldots, C\} :$$
$$B_i^\top = \left[b_{i1}, \ldots, b_{iC}\right] = \left[\frac{\exp(e_1)}{\sum_j \exp(e_j)}, \ldots, \frac{\exp(e_C)}{\sum_j \exp(e_j)}\right].$$

Then we can compute the attention processed representation as

$$T_i = \sum_j^C \vec{x}_i \odot b_{ji}, \forall i \in \{1, \ldots, C\}. \tag{3}$$

Following, T is processed by a set of 2 Resnet-18 blocks [36], with kernel and stride sizes tuned by a neural architecture search algorithm [37]. [1] These blocks contain a collection of kernels represented by θ_k. Then, the representation is processed by a fully connected layer. All the parameters until this last layer are referred to as θ_E, which parameterize the encoder of the neural network.

After T is processed by the encoder, it is then used to project random Fourier features [38, 39], parameterized by ω and β. Before decoding, the sinudoids are multiplied with a learned latent style vector, $W \in \mathbb{R}^L$, as $\vec{z} \odot W$. The result is processed by an affine transformation, mapping it to the fMRI volume space $\mathbb{R}^{D_2}$. With this, we formulate the function structure $F : \mathbb{R}^{C \times F \times T} \rightarrow \mathbb{R}^{D_2}$. Regarding interpretability, no previous studies have assessed the quality of EEG to fMRI synthesis in decision making settings, such as schizophrenia classification. Therefore, no assumptions can be made at this stage. We try to answer this question in the end of this manuscript.

[1] The kernels of these (downsampling) layers are $k_1 \in \mathbb{R}^{7 \times 37 \times 2}$ and $k_2 \in \mathbb{R}^{7 \times 7 \times 2}$. The stride sizes are $3 \times 5 \times 1$ and $2 \times 2 \times 1$, respectively.

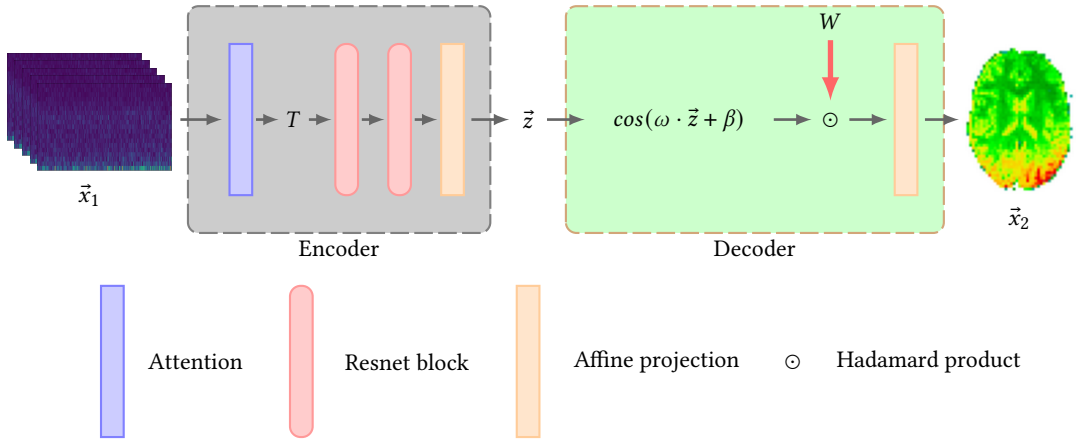
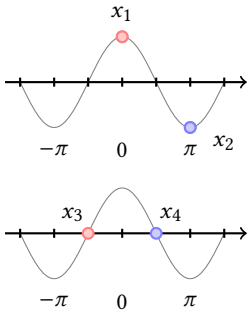


Figure 1. The neural architecture has two components: an Encoder (shaded in grey) and a Decoder (shaded in green). The input is the *stft* v_1 representation. The output is the synthesized *fmri*. The Encoder begins with a simple attention mechanism on the channels dimension of the *stft*. After, it is processed by two Resnet blocks and an affine layer. This produces the latent representation $\vec{z}$. Following, comes the Decoder, which picks this representation and builds the cosine bases through the projection $\omega \cdot \vec{z} + \beta$. The sinusoids are style induced with $cos(\omega \cdot \vec{z} + \beta) \odot W$. The vector W is a style fixed pretrained vector of an *fmri* representation, learned from a simultaneous EEG and fMRI dataset (described in Section 3.1). Finally, an affine layer projects it to the *fmri* space.

2 Creating discriminative fMRI views of EEG

To assess which view is better, we need to choose a classifier without a high feature engineering ability, in order to truly see the discriminative power of the view. A linear classifier fits this requirement, because it bases its decision solely on the multiplication of features with coefficients (weights), without additional feature rearrangement. To further motivate its use, a linear classifier has been used as a metric to assess explanation quality [24], exactly due to its lack of feature engineering ability.

As such, a linear classifier is used for the considered views. But a special case for v_2 arises. Since this view is produced by sinusoids, it disables the separation of data according to the input $y|\vec{x}_1$. We consider that, for generalization sake, the EEG modality contains information to separate the data, $\vec{x}_1$, according to its labels, y. Though, if this representation has its distribution broken at the sinusoids, we can no longer use it. To avoid this, we propose a method that maintains the sinusoids in the neural architecture flow and manipulates the representation, so that the separation before the sinusoid, $y|\omega \cdot \vec{z} + \beta$, is identically distributed to $y|\vec{x}_2$.



$$x_1, x_2 \sim \mathcal{X}_1 : ||x_1 - x_2||_1 = \pi$$

$$x \sim \mathcal{X}_1 : \mathbb{E}[x] = \tfrac{\pi}{2} \wedge \mathrm{Var}[x] = \tfrac{\pi}{2}$$

$$x_3, x_4 \sim \mathcal{X}_2 : ||x_3 - x_4||_1 = \pi$$

$$x \sim \mathcal{X}_2 : \mathbb{E}[x] = 0 \wedge \mathrm{Var}[x] = \tfrac{\pi}{2}$$

Figure 2. Description of how two similarly distributed samples, taken from $\mathcal{X}_1$ and $\mathcal{X}_2$, can lead to different portions of the cosine function image, since a sinusoid is periodic. Two distributions $\mathcal{X}_1$ and $\mathcal{X}_2$, may be mapped to the same image (second/bottom example). This is why a cosine is a shift invariant function. However, there are intervals a shift can be made and it is not invariant. Such intervals take the form $\forall i \in \mathbb{Z} : [i\pi, (i+1)\pi]$.

2.1 Sinusoid separation

In order to separate data along the projection of cosines, we have to operate in a sub-domain interval where the *cos* is not periodic. Layer normalization [40], with center in $\frac{\pi}{2}$ and standard deviation of $\frac{\pi}{2}$, maps most of the data to such an interval of the cosine. See Figure 2 for an illustration example.

Nonetheless, the assurance that the projections are likely in a non periodic sub-domain of the cosine function, does not alone complete separation of the data. This is because the distribution $y|\omega \cdot \vec{z} + \beta$ needs to be separated accordingly. We propose a modified contrastive loss defined as

$$\mathcal{L}_D(p_1, p_2, y_p) = y_p \times D(p_1, p_2) + (1 - y_p) \times ||D(p_1, p_2) - m||_1, \quad (4)$$

where p_1, p_2 formulate a pair of two instances and $y_p \in \{0, 1\}$ defines the pairwise label, being 0 when the label of p_1, p_2 mismatch and 1 when they match. $D(p_1, p_2)$ is the $l1$ distance $||p_1 - p_2||_1^1$. Setting $m \in [\epsilon, \pi] : \epsilon > 0$, together with the described layer normalization, allows separation of the data within a non periodic sub domain of the cosine. In Figure 3, the effects of this methodology are illustrated, before and after the learning session.

3 Experimental setting

In our experiments we considered 3 EEG views: 1) natural/raw EEG view in $\mathbb{R}^{C \times T}$, referred to as v_0; 2) time-frequency representation of EEG in $\mathbb{R}^{C \times F \times T}$, referred to as v_1; 3) synthesized fMRI representation in $\mathbb{R}^{M_1 \times M_2 \times M_3}$, referred to as v_2. The optimization of the parameters, of each of these, consists on the minimization of $\mathcal{L}_C$,

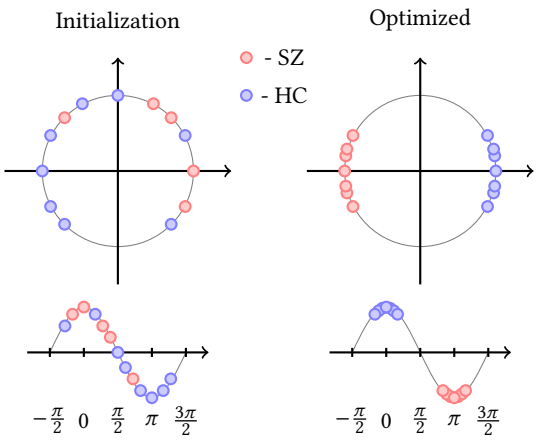


Figure 3. Normalization of data points inside the unit circle, using layer normalization, along with the optimization of a contrastive loss lead to correct separation of sinusoids. Data points belong to two classes, *HC* and *SZ*, that are separated after the minimization of $\mathcal{L}_D$. Because we separate false pairs, according to $(1 - y_p) \times ||D(x_1, x_2) - m||_1$, all points are placed within a shift variant interval of the cosine. The variance needed for classification.

such that the gradients are $\nabla_{\theta^i_C} \mathcal{L}_C$. The special case $v_2 = \{F_2, \theta_2\}$ requires a pretraining session in a simultaneous EEG and fMRI dataset, so it learns fMRI representations. Following, comes the minimization of the classification objective, where $\nabla_{\theta^2_C} \mathcal{L}_C$ is computed along with $\nabla_{\theta_E} \mathcal{L}_D$, being θ_E the parameters of the encoder of F (recall Section 1.2.3). All these objectives require the respective data for the computation of the gradients described above. Namely EEG data for classification and, in particular for v_2, the simultaneous EEG and fMRI data, for EEG to fMRI synthesis.

3.1 Datasets

***Classification: EEG only,* Fribourg dataset.** Padée et al. [41] gathered 43 individuals, 24 healthy controls and 19 with diagnosed schizophrenia, and performed a task-based recording session. In this recording session each individual played a game. The EEG was set with 128 electrodes, distributed according to the 10-20 system. The sampling rate was 2048Hz.

***Synthesis: simulatenous EEG and fMRI,* NODDI dataset.** Deligianni et al. [42] performed resting state recordings and a total of 8 individuals are considered. The EEG was setup with 64 channels, placed according to the 10-20 system. Recorded at a sampling frequency of 250Hz. The fMRI setup was recorded with a 2.160 Time Response and 30 millisecond echo time. Each voxel is $3 \times 3 \times 3$ millimeters, consequently making the resolution of a volume $M_1 \times M_2 \times M_3 = 64 \times 64 \times 30$ voxels.

3.2 Validation

To validate the hypotheses drawn, each view, v_i, will do a validation process. To this end, the steps of this process are:

1. leave-one-individual-out cross validation (LOOCV);
2. cross validation (5 folds) hyperparameter optimization with 25 iterations for each fold of the LOOCV step;
3. particularly for fMRI synthesized views, v_2, a pretraining session is done, where F is optimized on simultaneous EEG and fMRI data.

The hyperparameters used to train F are in accordance with the original work [17]. For the cross validation (step 2.), Bayesian optimization [43] is performed and the hyperparameters subject to optimization are: $l1$ regularization constant $\in [1e{-}10, 2.0]$, learning rate $\in [1e-5, 1.0]$ and batch size $\in \{1, 2, 4, 8, 16, 32\}$. The learning session is fixed with 10 epochs and gradients are propagated using the Adam optimizer [44].

3.3 Biclustering

With the synthesized fMRI view, $v_2(\vec{x}_0) = \vec{x}_2$, it is important to address which synthesized regions are most discriminative for schizophrenia. As well as compare the reported regions with related studies. For this, we ran BicPAMS [45] to find cluster subspaces composed of a subset of rows (individuals) and columns (voxels) that discriminate a target (pathology or healthy). Two settings are considered: 1) clustering $\vec{x}_2$ features with ground truth labels y; and 2) clustering $\vec{x}_2$ features with its linear classifier predictions $\hat{y} = \sigma(W^2_f \cdot v_2(\vec{x}_0) + b^2_f)$. This two settings allow us to find discriminative and explainability patterns from the synthesized fMRI view.

Discriminative patterns provide us sets of individuals and features that support a certain target. This analysis is similar to classification, but instead of having a classifier, we have a set of association rules. In turn, *Explainability* patterns give us interpretable information about the decision made by the linear classifier. Although it does not provide discriminative patterns, it gives us association rules for the decision making process of the classifier. The biclustering of the features, $\vec{x}_2$, and the predictions, $\hat{y}$, produces global explanations, forming a novel explainability algorithm. Jointly, these settings allow us to have a better comprehension from the synthesized fMRI view, potentially able to uncover advantages and pitfalls. An important analysis when working with data driven projections.

Biclusters were found in three resolutions: $5 \times 5 \times 3$; $10 \times 10 \times 5$; and $14 \times 14 \times 7$. These resolutions enable us to assess patterns at different granularities. In a $5 \times 5 \times 3$ clusters represent big regions of the brain, as big as entire lobes; a $10 \times 10 \times 5$ resolution can still retrieve regions of interest, but at a finer granularity; and $14 \times 14 \times 7$ goes even more detailed. Altogether, this analyses give us patterns, with statistical assurances for a target (schizophrenia).

The parameters given to BicPAMS to search for biclusters in the different resolutions are shown in Table 1. Note that, for the lowest resolution $5 \times 5 \times 3$, the number of *columns*, which translates to the number of voxels is only 1. This is because 1 voxel in a $5 \times 5 \times 3$ corresponds to a big region in the original resolution, $64 \times 64 \times 30$. As the resolution increases, we also increased the number of voxels required to appear in a bicluster. The number of *voxels* pretains to the discretization of the data. It is important to not increase this parameter, as the *items-boundary* problem can arise [46]. This problem manifests when the number of *bins* is high and similar values are put in different bins. The number *biclusters* requires that all biclusters obtained are bigger than this value. The *lift* tells us how discriminative is the cluster for a target. In other words, the *lift* is the ratio of how much an item occurs in the bicluster by how much it occurs in the dataset. Typically, $lift \gg 1$ indicates that the selected bicluster is able to discriminate the given target.

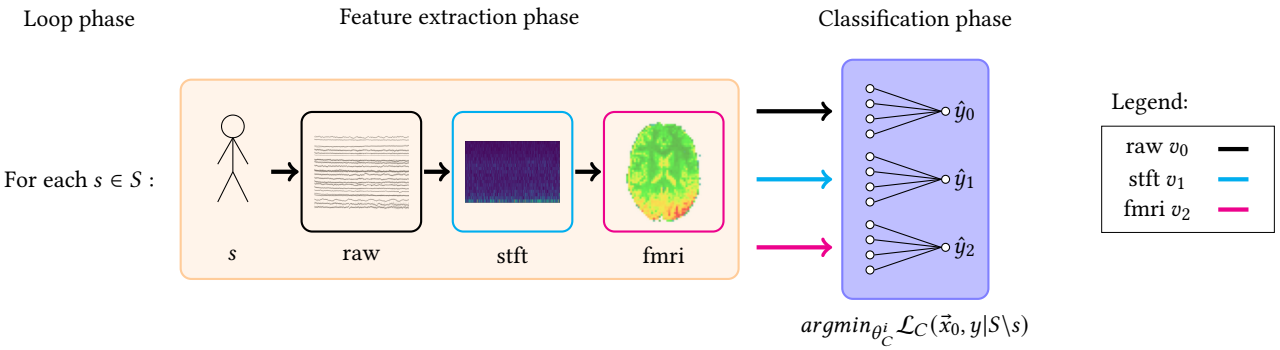


Figure 4. We do a leave-one-individual-out validation, where for each fold we either train a linear classifier with *raw*, *stft* or *fmri* representations. Each representation has its own validation. The arrows inside the feature extraction phase indicate dependency, that is: an *fmri* representation needs an *stft*; *stft* needs the *raw*; and the *raw*, of course, needs an individual's recording, denoted with a human figure. For each fold, $s \in S$, we train a linear classifier without s, $argmin_{\theta_C^i} \mathcal{L}_C(\vec{x}_0, y|S\backslash s)$. In addition, the *fmri* setting also optimizes $argmin_{\theta_E} \mathcal{L}_D$. For all individuals/folds the predictions are saved to compute the are under the curve (AUC) against the ground truth.

Parameter	$5 \times 5 \times 3$		$10 \times 10 \times 5$		$14 \times 14 \times 7$	
	y	$\hat{y}$	y	$\hat{y}$	y	$\hat{y}$
lift	1.3	1.3	1.3	1.2	1.3	1.2
# biclusters	100	100	100	100	100	100
# bins	3	3	5	5	5	5
min # voxels	1	1	10	10	15	15

Table 1. Parameters for the biclustering algorithm.

4 Results

The results for the classification experiment are shown in Table 2. The *stft* representation had the best performance with an AUC of 0.933, followed by *fmri* with 0.765 and *raw* performed below random with 0.225. The performance of the *raw* representation is low and shows that an EEG recording without preprocessing steps is not able to be applied in schizophrenia classification settings. On the other hand, the *stft* is capable of it, showing the time-frequency domain features are discriminative of schizophrenia. Nonetheless, we expected *fmri* to be closer to the performance of *stft*. An AUC of 0.765 shows it has a good prediction power, however, it was outperformed by its preceding representation, according to Figure 4. We found no statistical significance between *raw* and *fmri*, on the other hand *stft* outperformed *raw* and *fmri* with statistical significance.

View	Fribourg (AUC)
raw v_0	0.225
stft v_1	**0.933**
fmri v_2	0.765

Table 2. AUC of a linear classifier with different views as input.

In terms of synthesis quality, the synthesized fMRI were well defined, meaning that they appeared as fMRI volumes to the human eye. Though, there are ill defined predictions with activity present in the background of the volume. The latter may be due to each model at a fold, converging to different suboptimal parameters. Consequently, those parameters may lead to different distributions. Nevertheless, the success of the synthesis is encouraging, since it demonstrates the ability of $\mathcal{L}_D$, along with the proposed layer normalization, to maintain the style of the fMRI and at the same time separate the data. The proposed loss also demonstrated to work well in a joint training with an additional classification loss, $\mathcal{L}_C$. We were able to take advantage of a shift invariant function, *cosine*, and process a different EEG dataset by an encoder, that enabled the decoder to project to an fMRI with the learned distribution (the distribution of the NODDI fMRI).

In terms of biclustering, we were able to find biclusters with the parameters for all the resolutions, using the ground truth and the predictions. From the gathered biclusters, we report the best biclusters according to the *lift*. All biclusters have statistical significance, an assurance of the BicPAMS algorithm. In contrast with the classification setting, the biclustering analysis shows us there are patterns with high discriminative power for schizophrenia. Suggesting that models with a better feature engineering would take advantage of these patterns.

5 Discussion

Castanho et al. [47] study the application of several biclustering algorithms in fMRI data to uncover statistically significant patterns. One of the algorithms studied was BicPAMS. The biclusters were of the type voxels by time, $D_2 \times T$. The authors claim this setup has the leverage of finding patterns that correlate/connect different brain regions over the temporal dimension, a.k.a. functional connectivity. Indeed, it is known that distant brain regions communicate between each other through neuronal pathways. This

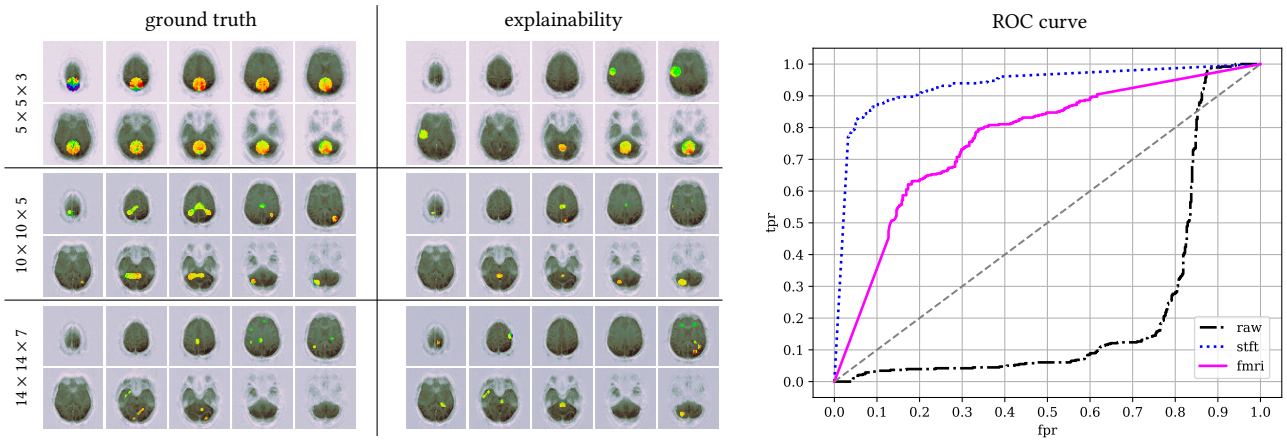


Figure 5. We analyzed resolutions $\in \{5 \times 5 \times 3, 10 \times 10 \times 5, 14 \times 14 \times 7\}$ and gathered the biclusters retrieved for the ground truth and predicted labels. Only the best biclusters (with the best *lift*) are shown in this figure for each setting. On the right side, the receiver operating characteristic (ROC) curve is shown. The *raw* view performed below random, while the *stft* and *fmri* were above. The *stft* is close to a perfect classifier.

phenomena can be observed with similar frequencies present in distant EEG electrodes [48]. Note that, we do not consider the temporal dimension of fMRI in our pattern search made by BicPAMS. Thus, we can not make these functional connectivity claims from the synthesized fMRI. Nonetheless, clusters of the form individuals by voxels, $S \times D_2$, allow us to assess if there are *constant*[2] spatial patterns in the fMRI volume that can discriminate schizophrenia. **Outside of brain volume areas, such as background, are not discriminative.** We performed ablation experiments, to ensure no patterns were being found in regions where they were not supposed to exist, such as the background. For this, we collected all these regions of the synthesized volumes, and BicPAMS did not find biclusters with statistical assurances. This experiment rejected the possibility of discriminative information out of brain regions. **Yet, in in brain regions, BicPAMS found several biclusters**, all of them with lift greater than 1.38 for the schizophrenia class. This means, the synthesized fMRIs are able to represent schizophrenia related patterns and do not build patterns in healthy brains. The latter, is particularly encouraging, since healthy brains in this task should not have patterns present.

The synthesized cerebellum region is present in several biclusters associated with schizophrenia. In the biclusters of the ground truth labels, we found heterogeneous regions in the resolutions. Meaning, finer granularities uncover information that entire lobes, as a whole, do not. This suggests higher resolution volumes may contain relevant patterns that would not be discovered otherwise. To put in perspective, these biclusters in $5 \times 5 \times 3$ volumes were present in the parietal lobe, left temporal lobe and cerebellum; in $10 \times 10 \times 5$ resolution volumes patterns were found in parietal, occipital and left temporal lobes, as well as in the cerebellum region; and in $14 \times 14 \times 7$ biclusters had voxels present in parietal, occipital, frontal and left temporal lobes. Figure 5 illustrates the best biclusters found according to lift. This patterns go in accordance with previous findings reporting that prefrontal and temporal lobes are affected by schizophrenia [49]. On another note, Rahaman et al. [23] made a significant contribution on the application of biclustering in MRI data. They were able to find discriminative MRI patterns for schizophrenia patients. While MRI measures white matter, fMRI records blood supply levels. MRI is not fMRI, but we find it pertinent to relate this study with ours, since biclusters present in regions associated with schizophrenia (gyrus, brainstem) are relevant, independently of the modality. Also, comparing our results, of a synthesized fMRI modality, with an MRI, lets us assess if the synthesis is veracious in a spatial perspective. In their study, a biclustering algorithm was ran on individuals considering nine components (taken from a division of 30 regions of interest using independent component analysis). The biclusters found observed patterns in the gyrus and brainstem parts of the brain. Our different granularity experiments go in accordance with biclusters containing patterns in the cerebellum which is connected to the brainstem. However, we did not report any pattern in the gyrus region. Of course no major claim can be made about the spatial veracity of this inter study correlation, since both our view is synthesized and datasets are different. Additional analyses are needed to check if the activity synthesized in the different regions of the fMRI volume goes in accordance with the dynamics of a real fMRI.

In the explainability analysis, we found different biclusters at higher resolutions, but still present in the same regions. In $5 \times 5 \times 3$ resolution volumes, patterns were observed in parietal, occipital, left temporal lobes and cerebellum. At finer granularities, voxels in the frontal lobe were present in the biclusters retrieved for $10 \times 10 \times 5$ and $14 \times 14 \times 7$ resolutions. Areas uncovered in $5 \times 5 \times 3$ were also present at higher resolutions, except for the cerebellum at $14 \times 14 \times 7$. These biclusters all reported lifts above 1.2. Lifts in explainability were lower than the ground truth. In general high lift patterns were harder to find in this setting, and as a consequence the *minimum lift* parameter (see Table 1) had to be lowered (relative the ground truth) to relax the search.

The linear classifier decision making is correlated with the synthesized frontal lobe activity. Doing explainability on

[2] Constant patterns are patterns that are equal for every individual.

a leave-one-individual-out validation schema is difficult. We only have access to the model at the prediction time of the fold. In addition, different folds have different suboptimal parameters. Altogether, these are challenges that we tackled using a pattern mining tool. Pinto et al. [50] used the apriori algorithm to retrieve the explanations of predictions. The idea is that, by finding patterns of data that have statistical assurances for a target, we are explaining predictions. Note however, that the model may not be looking at those patterns to make its decision. What the explainability analysis tells us, is how the synthesized fMRI correlates with the predictions made by the linear classifier. We see them as biclusters that explain the predictions, with statistical assurances. In this setting, BicPAMS found different row sets, because the predictions differ from the ground truth. So, how can we view these biclusters? They discriminate the predictions, a difference seen in the different regions gathered by the biclusters. For instance, explainability biclusters reported the frontal lobe presence, while the ground truth ones did not report this region. Nonetheless, the frontal lobe is associated with problem solving and attention functions, which are recognized as impairments provoked by schizophrenia [21, 22]. This pathology affects the cognitive ability and signatures of the human brain [51]. There is extensive research on the different discriminative patterns that are able to identify this pathology and a lot of research is performed using MRI technologies.

The synthesized fMRI view is discriminative of schizophrenia. The statistical assurances (lift, support) were higher for the biclusters found in the ground truth analysis. No major comment is made for this observation, as they are different settings, that can not be compared. However, the high discriminative power of the ground truth biclusters for schizophrenia, show that the produced fMRI views have potential for schizophrenia diagnostics. The low AUC, 0.765, of the linear classifier shows us that it is not sufficient for a reliable application of this view. All in all, a linear classifier has no feature engineering properties, but BicPAMS gave us statistical assurances about the discriminative patterns retrieved. Showing us that there is information present in this view, potentially uncoverable by powerful models. We refer this for future work.

Frequency features are highly discriminative of schizophrenia, yet lack interpretability. The linear classifier showed us that it better assessed schizophrenia using the *stft* representation. It is very well known that frequency features are highly discriminative of this pathology [19, 20]. However, we were not able to show the power of the *fmri* view, through this classification setup. Still, this makes sense. The distance from *stft* to the *fmri*, shown in Figure 1, inhibits the gradients of $\mathcal{L}_D$ w.r.t. θ_F to be significant at the top layers. These are the ones close to the *stft* representation. Since the number of epochs is fixed, we hypothesize that not enough time was given in the validation process to show that the *fmri* view is able to perform in comparison to *stft*. We hypothesize that finer regularization strategies, such as l_1-path-norm regularization [52], are needed to allow a faster convergence. Nonetheless, the superiority of *stft* features is due to it being better engineered for schizophrenia. In contrast, the *raw* representation lacks not only in interpretability, but also lacks this engineering property to allow a good performance of the linear classifier. We look at an *fmri* representation and see its potential applicability in a health care setting, since it has higher levels of interpretability and, as previously discussed, highly discriminative patterns. And although, *stft* had a better performance (see ROC curve in Figure 5), it still lacks in the interpretability level. It is essential that a clinical diagnostic be made of simple explanations. Not only has the doctor to understand, but also it is beneficial if the patient fully understands its diagnostic [53]. Explaining a diagnostic based on *stft* features is not tractable for people out of the EEG clinical scope. Not to say that the *fmri* representation is understandable by everyone, but it is easier since it is explained in the spatial (and temporal) domain. EEG is recognized to have low interpretability power [27], however it does not block its use in health care. Our main concern, is projecting this modality to a space where it can better be understood by a human, be it expert or not.

6 Conclusion

This study presented a decision making model that is able to produce an *fmri* view from an EEG representation, through the integration of a novel contrastive loss on a state-of-the-art EEG to fMRI synthesis network. The validity of this synthesis task for subsequent diagnosis from single neuroimaging modes/sources is demonstrated for the first in this work, showing delineate predictive accuracy against peer approaches. Our results show that time frequency features, corresponding to an *stft* view, remain as the best option for schizophrenia diagnosis, yet lack interpretability. In turn, we showed, through an exhaustive search framework, that the *fmri* view contains highly discriminative patterns, revealing notable associations of schizophrenia with left temporal, frontal and cerebellum regions. The role of cross-modal augmented neuroimaging views is highlighted as an important subsequent direction.

Limitations

Note that the synthesized fMRI view analysed in this study has some limitations. We consider the following: Sinusoid-based projections may not be sufficiently expressive to handle all relevant distribution shifts, specially when considering case-control populations; EEG records are mostly suggested for the diagnosis of a limited set of pathologies, for instance, according to our experiments, raw and stft EEG views show poor predictive performance for the classification of Alzheimer's and Parkinson's diseases; the synthesized fMRI is only as good as the quality of paired EEG-fMRI records use to train the model. Nowadays, in spite of the emerging set of cohorts with simultaneous EEG and fMRI data, learning limitations related to cohort size and undertaken recording protocols should be carefully noted.

Acknowledgements

This work was supported by national funds through Fundação para a Ciência e Tecnologia (FCT), under the Ph.D. Grant SFRH/BD/5762-/2020 to David Calhas and INESC-ID pluriannual UIDB/50021/2020. We want to thank João Torres for fruitful discussions regarding data imbalance and ways to tackle it. We also give a big thanks to Daniel Gonçalves and Pedro Orvalho for brainstorming along us some ideas of this work. A big thanks to André Patrício for helping us with setting the code in the biclustering analysis.

Code Availability

Simulation code was written in Python 3. Please refer to the eeg-to-fmri Python package and the notebook example (using a subset of the data used) found in this link: classification_contrastive.ipynb.

References

[1] Sukanta Saha et al. A systematic review of the prevalence of schizophrenia. *PLoS medicine*, 2(5):e141, 2005.

[2] Christina Andreou et al. Structural and functional imaging markers for susceptibility to psychosis. *Molecular psychiatry*, 2020.

[3] Benjamin M Neale et al. Genetic analysis of schizophrenia and bipolar disorder reveals polygenicity but also suggests new directions for molecular interrogation. *Current opinion in neurobiology*, 2015.

[4] Cheryl M Corcoran et al. Language as a biomarker for psychosis: A natural language processing approach. *Schizophrenia research*, 2020.

[5] Jihoon Oh et al. Identifying schizophrenia using structural mri with a deep learning algorithm. *Frontiers in psychiatry*, 11:16, 2020.

[6] Dennis Velakoulis et al. Hippocampal and amygdala volumes according to psychosis stage and diagnosis: A magnetic resonance imaging study of chronic schizophrenia, first-episode psychosis, and ultra–high-risk individuals. *Archives of general psychiatry*, 63(2):139–149, 2006.

[7] Unn Kristin Haukvik et al. Schizophrenia–what does structural mri show? *Tidsskrift for Den norske legeforening*, 2013.

[8] Godwin Inalegwu Ogbole et al. Survey of magnetic resonance imaging availability in west africa. *Pan African Medical Journal*, 30(1), 2018.

[9] Apoorva Sikka et al. Mri to fdg-pet: cross-modal synthesis using 3d u-net for multi-modal alzheimer's classification. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 80–89. Springer, 2018.

[10] Gianmarco Santini et al. Unpaired pet/ct image synthesis of liver region using cyclegan. In *16th International Symposium on Medical Information Processing and Analysis*, volume 11583, pages 247–257. SPIE, 2020.

[11] Qianye Yang et al. Mri cross-modality neuroimage-to-neuroimage translation. *arXiv preprint arXiv:1801.06940*, 2018.

[12] Rodolfo Abreu et al. Eeg synchronization measures predict epilepsy-related bold-fmri fluctuations better than commonly used univariate metrics. *Clinical Neurophysiology*, 129(3):618–635, 2018.

[13] Ian Daly et al. Electroencephalography reflects the activity of subcortical brain regions during approach-withdrawal behaviour while listening to music. *Scientific reports*, 9(1):1–22, 2019.

[14] Xueqing Liu et al. A convolutional neural network for transcoding simultaneously acquired eeg-fmri data. In *NER*, pages 477–482. IEEE, 2019.

[15] Xueqing Liu et al. Latent neural source recovery via transcoding of simultaneous eeg-fmri. *arXiv preprint arXiv:2010.02167*, 2020.

[16] Paul Bricman et al. Eeg2fmri: Cross-modal synthesis for functional neuroimaging. 2021.

[17] David Calhas et al. Eeg to fmri synthesis benefits from attentional graphs of electrode relationships. *arXiv preprint arXiv:2203.03481*, 2022.

[18] Gérard Govaert et al. Block clustering with bernoulli mixture models: Comparison of different approaches. *Computational Statistics & Data Analysis*, 52(6):3233–3245, 2008.

[19] Turan M Itil. Qualitative and quantitative eeg findings in schizophrenia. *Schizophrenia bulletin*, 3(1):61, 1977.

[20] Brian J Roach et al. Event-related eeg time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia bulletin*, 34(5):907–926, 2008.

[21] Cynthia G Wible et al. Prefrontal cortex, negative symptoms, and schizophrenia: an mri study. *Psychiatry Research: Neuroimaging*, 108 (2):65–78, 2001.

[22] Daniel R Weinberger et al. The frontal lobes and schizophrenia. *The Journal of neuropsychiatry and clinical neurosciences*, 1994.

[23] Md Abdur Rahaman et al. N-bic: A method for multi-component and symptom biclustering of structural mri data: Application to schizophrenia. *IEEE Transactions on Biomedical Engineering*, 67(1):110–121, 2019.

[24] Marcos V Treviso et al. The explanation game: Towards prediction explainability through sparse communication. *arXiv preprint arXiv:2004.13876*, 2020.

[25] Zirui Huang et al. Temporal circuit of macroscale dynamic brain activity supports human consciousness. *Science advances*, 6(11):eaaz0087, 2020.

[26] William O Tatum IV. *Handbook of EEG interpretation*. Springer Publishing Company, 2021.

[27] Hans Op de Beeck et al. *Introduction to human neuroimaging*. Cambridge University Press, 2019.

[28] Shu Lih Oh et al. Deep convolutional neural network model for automated diagnosis of schizophrenia using eeg signals. *Applied Sciences*, 2019.

[29] Carla Barros et al. Advanced eeg-based learning approaches to predict schizophrenia: Promises and pitfalls. *Artificial intelligence in medicine*, 114:102039, 2021.

[30] Andrew A Fingelkurts et al. Impaired functional connectivity at eeg alpha and theta frequency bands in major depression. *Human brain mapping*, 28(3):247–261, 2007.

[31] Adrian P Burgess et al. Functional connectivity of gamma eeg activity is modulated at low frequency during conscious recollection. *International Journal of Psychophysiology*, 46(2):91–100, 2002.

[32] Matthew A Albrecht et al. Time and frequency dependent changes in resting state eeg functional connectivity following lipopolysaccharide challenge in rats. *PLoS One*, 13(11):e0206985, 2018.

[33] Alan T Bates et al. Low-frequency eeg oscillations associated with information processing in schizophrenia. *Schizophrenia research*, 115 (2-3):222–230, 2009.

[34] Ronald Newbold Bracewell et al. *The Fourier transform and its applications*, volume 31999. McGraw-Hill New York, 1986.

[35] Eric Jacobsen et al. The sliding dft. *IEEE Signal Processing Magazine*, 20(2):74–80, 2003.

[36] Kaiming He et al. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[37] David Calhas et al. Automatic generation of neural architecture search spaces. In *Combining Learning and Reasoning: Programming Languages, Formalisms, and Representations*, 2021.

[38] Ali Rahimi et al. Random features for large-scale kernel machines. In *NIPS*, volume 3, page 5. Citeseer, 2007.

[39] Matthew Tancik et al. Fourier features let networks learn high frequency functions in low dimensional domains. *arXiv preprint arXiv:2006.10739*, 2020.

[40] Jimmy Lei Ba et al. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[41] Anna Padée et al. "fribourg ultimatum game in schizophrenia study", 2022.

[42] Fani Deligianni et al. Relating resting-state fmri and eeg whole-brain connectomes across frequency bands. *Frontiers in Neuroscience*, 8: 258, 2014. ISSN 1662-453X. doi: 10.3389/fnins.2014.00258. URL https://www.frontiersin.org/article/10.3389/fnins.2014.00258.

[43] Jasper Snoek et al. Practical bayesian optimization of machine learning algorithms. *NIPS*, 25, 2012.

[44] Diederik P Kingma et al. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[45] Rui Henriques et al. Bicpams: software for biological data analysis with pattern-based biclustering. *BMC bioinformatics*, 18(1):1–16, 2017.

[46] Rui Henriques et al. Bicpam: Pattern-based biclustering for biomedical data analysis. *Algorithms for Molecular Biology*, 2014.

[47] Eduardo N Castanho et al. Biclustering fmri time series: a comparative study. *BMC bioinformatics*, 23(1):1–30, 2022.

[48] Gert Pfurtscheller et al. Event-related eeg/meg synchronization and desynchronization: basic principles. *Clinical neurophysiology*, 110(11): 1842–1857, 1999.

[49] Rowena Chin et al. Recognition of schizophrenia with regularized support vector machine and sequential region of interest selection

using structural magnetic resonance imaging. *Scientific Reports*, 8(1): 1–10, 2018.

[50] Mauro Pinto et al. Interpretable eeg seizure prediction using a multi-objective evolutionary algorithm. *Scientific reports*, 12(1):1–15, 2022.

[51] Jing Sui et al. Multimodal neuromarkers in schizophrenia via cognition-guided mri fusion. *Nature communications*, 9(1):1–14, 2018.

[52] Behnam Neyshabur et al. Path-sgd: Path-normalized optimization in deep neural networks. *NIPS*, 28, 2015.

[53] Ajay Kumar Yadav et al. Patients understanding of their diagnosis and treatment plans during discharge in emergency ward in a tertiary care centre: a qualitative study. *JNMA: Journal of the Nepal Medical Association*, 57(219):357, 2019.