EEGminer: discovering interpretable features of brain activity with learnable filters

Siegfried Ludwig\(^1\), Stylianos Bakas\(^2\), Dimitrios A Adamos\(^4\), Nikolaos Laskaris\(^2\), Yannis Panagakis\(^3\), and Stefanos Zafeiriou\(^1\)

\(^1\) Department of Computing, Imperial College London, London SW7 2RH, United Kingdom
\(^2\) School of Informatics, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
\(^3\) Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens 15784, Greece
\(^4\) Cogitat Ltd, London, United Kingdom
\(^5\) Equal contribution.

E-mail: siegfried.ludwig20@imperial.ac.uk

Keywords: electroencephalography (EEG), functional connectivity, learnable filters, feature mining

Abstract

Objective. The patterns of brain activity associated with different brain processes can be used to identify different brain states and make behavioural predictions. However, the relevant features are not readily apparent and accessible. Our aim is to design a system for learning informative latent representations from multichannel recordings of ongoing EEG activity.

Approach: We propose a novel differentiable decoding pipeline consisting of learnable filters and a pre-determined feature extraction module. Specifically, we introduce filters parameterized by generalized Gaussian functions that offer a smooth derivative for stable end-to-end model training and allow for learning interpretable features. For the feature module, we use signal magnitude and functional connectivity estimates.

Main results. We demonstrate the utility of our model on a new EEG dataset of unprecedented size (i.e. 721 subjects), where we identify consistent trends of music perception and related individual differences. Furthermore, we train and apply our model in two additional datasets, specifically for emotion recognition on SEED and workload classification on simultaneous task EEG workload. The discovered features align well with previous neuroscience studies and offer new insights, such as marked differences in the functional connectivity profile between left and right temporal areas during music listening. This agrees with the specialisation of the temporal lobes regarding music perception proposed in the literature.

Significance. The proposed method offers strong interpretability of learned features while reaching similar levels of accuracy achieved by black box deep learning models. This improved trustworthiness may promote the use of deep learning models in real world applications. The model code is available at [https://github.com/SMLudwig/EEGminer/](https://github.com/SMLudwig/EEGminer/).

1. Introduction

Several conventional measures have been proposed for estimating underlying connectivity from EEG recordings [1], such as signal correlations and phase locking values (PLVs) [2]. Tuning these measures, however, involves extensive manual feature engineering. Indeed, in practice, the search space of informative features explodes exponentially when one considers coordinated brain activity and employs functional interactions between distinct brain areas. Therefore, manual feature engineering is a suboptimal process. In contrast, deep neural networks [3] enable learning features from large amounts of annotated training data. Concretely, such deep models involve trainable parameters (usually arranged into a series of processing layers), the values of which are estimated using backpropagation of gradients from the training data [4]. However, features extracted by deep models often lack in interpretability, which can present significant detriments in real-world applications [5]. Besides interpretability,
off-the-shelf deep learning models require a significant amount of annotated data, hindering their applicability in data-scarce scenarios such as EEG analysis.

The aforementioned limitations of deep neural networks regarding interpretability and the need for large training datasets can be reduced by introducing appropriate inductive biases. The most common approach in the EEG domain is to use convolutional neural networks (CNN) \[6, 7\]. A prominent example is EEGNet, which takes cues from conventional EEG signal processing techniques to create a compact and relatively interpretable model \[8\]. There are very limited constraints on the model however, such as fully trainable temporal filter functions. Convolutional filters have been constrained by parameterized sinc functions in the audio domain \[9\], an approach gaining some popularity for EEG decoding \[10–13\]. This constrains the model but could reduce trainability, especially in the low signal-to-noise ratio regime of the EEG domain. Further, CNNs are usually thought of as processing spectral-temporal features. An attempt to introduce constraints towards using EEG functional connectivity features has been undertaken with SyncNet \[14\], which however does not make these constraints fully explicit and thereby limits its interpretability. Other interpretability approaches use class activation maps of deep learning models to reveal the selection of different electrodes and frequencies \[15\], or use gradient analysis to quantify the impact of changes in specific parts of the input w.r.t. the classification output \[16\].

Distinct from the above models, we propose to implement an end-to-end differentiable EEG decoding pipeline (coined as EEGminer), making use of explicit neuroscientific measures of brain activity, such as signal correlations. The type of feature itself is therefore untrainable, but by using a differentiable implementation, the frequencies over which the connectivity is computed can be learned. The proposed model consists of learnable temporal filters parameterized by generalized Gaussians, a differentiable feature module and a linear layer for classification (figure 1). The EEGminer can easily incorporate multiple types of features (local activity, covariation, phase synchrony, etc) and keeps close relation with the topological information about the sensors, which in turn guarantees the direct interpretation of the trained model. Our approach can be viewed both from the perspectives of conventional neuroscience on the one hand and deep learning on the other. The steps that make up an EEG pipeline (e.g. signal filtering, connectivity measure, classification) are analogous to consecutive layers in a deep neural network. Conversely, viewed from a deep learning perspective, our proposal for a differentiable EEG pipeline can be seen as introducing strong inductive biases into an end-to-end trainable model. By choosing an appropriate model architecture and training procedure, the

---

**Figure 1.** Visualization of the EEGminer architecture, applied on a batch \(B\) of EEG signals. Independently trainable filters are applied on each of \(C\) input channels (signals from \(C = 5\) electrodes are shown here over \(T = 4\) s). As an example of a differentiable feature module used in this paper, the filtered signals are then correlated with each other to get connectivity matrices. As they are symmetric, the upper triangular is extracted as a feature vector. Multiple filters \(F\) per input channel result in multiple independent feature maps (shown here are three connectivity feature maps). The feature vectors resulting from all feature maps are concatenated, standardized (not shown) and used by the linear classifier with dimension \(D\) for logistic regression.
Figure 2. Flowchart of the proposed method. First, the dataset is prepared and a feature module is selected. The model is then trained by iteratively making predictions on the training data and updating weights via backpropagation of loss gradients. The mined features can then be extracted from the trained model.

interpretability of learned features can be made explicit while maintaining competitive decoding accuracy (figure 2). A PyTorch implementation is available on GitHub6.

Traditionally defined frequency bands are often used in EEG analysis, but we argue that using a priori band limits can impose a somewhat arbitrary structure on the obtained results. Instead, we propose to use trainable bandpass filters to find the most discriminative frequency ranges in a data-driven way. Learnable Gaussian filters have been used before as part of deep learning models, both as a trainable filterbank [17] as well as in their time-domain representation as Morlet wavelets to parameterize a convolutional filter [18]. In this work, we propose to design a more general filter function, based on the generalized Gaussian, which introduces a shape parameter. During training, the filter shape can adapt from a Gaussian (i.e. Morlet wavelet) with good training behaviour to a rectangular filter (i.e. sinc) with good frequency selectivity.

To demonstrate the generalizability of our approach, we apply it on three datasets and four distinct classification tasks, in all cases searching for subject-independent features. The chosen tasks put a focus on spontaneous brain activity that does not exhibit event-locked responses. The main dataset used was collected in collaboration with the Science Museum London, providing recordings from an unprecedented number of participants during music listening. We train our model to discriminate between resting state and music listening, and to differentiate female and male subjects based on their brainwaves during music listening. Generalizing to different settings, we use the SEED dataset to classify positive and negative emotions elicited by watching emotional film clips [19, 20], and the simultaneous task EEG workload (STEW) dataset to classify task-induced workload against resting state brain activity [21]. For all four tasks we compare model classification performance to conventional and deep learning approaches and extract the learned magnitude and functional connectivity features and the importance attributed to them by the classifier.

In summary the contributions of this paper are as follows:

- We introduce generalised Gaussian filters that allow for end-to-end learning of frequency bands of interest.
- We propose a modular architecture for decoding EEG signals with explicitly interpretable features, without sacrificing accuracy.
- We provide novel insights into brain states underlying music listening, emotional responses and mental workload.

The remainder of the paper is organized as follows: section 2 describes related work in the literature and section 3 outlines the proposed method. Section 4 describes the baselines and datasets employed in this study and presents results on performance benchmarks as well as interpretations of the learned features and their statistical validation. The last section summarizes and concludes our work.

2. Related work

Two widely cited CNNs developed for EEG decoding are ShallowNet and DeepNet [22]. ShallowNet employs a single temporal filtering layer with trainable convolution kernels shared across electrode channels, followed by a spatial filter, temporal pooling and linear classifier. The composition of temporal and spatial filters is inspired by the filter bank common spatial pattern algorithm [23]. Using similar operations, DeepNet targets more complex feature extraction with additional network layers. Both models have a large number of trainable parameters due to missing constraints and are relatively difficult to interpret as a consequence.

EEGNet is another important convolutional decoding model developed for the EEG domain [8]. The EEGNet architecture is comprised of trainable convolutional temporal filters very similar to ShallowNet. This is followed by spatial filters, which create projections from signal channels filtered in the same frequencies. Constraining the spatial filter to work within a single temporal filter channel makes the model more parameter efficient and facilitates

6 https://github.com/SMLudwig/EEGminer/.
The types of temporal filters should be explicitly constructed as the difference between two sinc lowpass filters as

\[ k(n) = 2f_2 \text{sinc}(2\pi f_2 n) - 2f_1 \text{sinc}(2\pi f_1 n), \quad (1) \]

where the sinc function is defined as \( \text{sinc}(x) = \sin(x)/x \). The trainable parameters are here only the low and high cutoff frequencies \( f_1 \) and \( f_2 \) as opposed to the full filter function. Limiting the temporal filters to sinc functions improves interpretability, since each filter is defined by a specific frequency band, while still allowing for learnable filters. However, an important difference between audio and EEG signals is the significantly lower signal-to-noise ratio in the latter, which is expected to make training of the bandpass filters more difficult. Looking at sinc bandpass filters in the frequency domain, a weakness with regard to its use in deep learning is apparent: the ideal sinc bandpass filter has a rectangular frequency response, which has a flat pass-band and essentially no transition band and therefore a zero derivative at essentially every point, meaning it cannot give useful gradients when training a deep learning model (see figure 3). The convolutional sinc filter still works, since it has to be truncated in the time domain, which introduces a transition band in the frequency domain and therefore some useful gradients, but this leaves room for improvement.

Applying a similar constraint on the first-level filters, recent work on inductive biases for the EEG domain has been done with SyncNet, a simple deep learning model targeting EEG synchrony with claims towards interpretability [14]. The model is based on a trainable convolutional layer parameterized as the real part of a complex Morlet wavelet. The time-domain filter is defined as

\[ k(n) = b \cos(\omega n + \phi) \exp(-\beta n^2), \quad (2) \]

where \( b \) is the amplitude of the wavelet, \( \omega \) is the centre frequency, \( \phi \) is the phase-shift and \( \beta \) controls the frequency-time precision tradeoff. An approximation of cross-spectral density is then achieved by a linear combination of the filtered input channels. This linear combination is identical in its nature to the spatial projection used in EEGNet and is thereby limited in its approximation of synchrony. Furthermore, it insufficiently constrains the feature space towards an approximation of synchrony.

\[ G(x) = \frac{\beta}{2\alpha \Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta}, \quad (3) \]

CNNs are commonly thought to work with spectral features. Recently there has been growing interest in adapting deep learning architectures to work on symmetric positive definite (SPD) matrices, such as spatial covariance matrices of EEG signals. SPDNet was developed as a simple deep learning model to be applied on SPD matrices [24] and has been used for EEG decoding [25]. The focus with this model is however on the classification of SPD matrices, not on creating the specific SPD matrix from input signals in the first place. In the case of EEG decoding, usually a covariance matrix derived from wide-band signals is used. The interpretation of trained features is difficult.

Considering the limitations of the aforementioned models, a well trainable and directly interpretable model should have the following characteristics:

- The types of temporal filters should be explicitly defined and have good gradient support on all relevant frequencies.
- The feature extraction should be a close approximation or direct computation, and well constrained to the specific feature.
- The training procedure should enable direct interpretation of classifier weights.

### 3. Proposed method

The proposed model consists of trainable temporal filters with specific constraints, followed by a predefined differentiable feature module (band magnitude, functional connectivity) and leading to a linear classifier (see figure 1). Each of these components will be presented in the following subsections, accompanied by our strategy for feature interpretation and the training procedure.

#### 3.1. Temporal filters

As mentioned in section 2, a prominent example of using parameterized filter functions in a deep learning model are sinc bandpass filters [9]. Sinc filters however have a very narrow transition band, which means the derivative of the filter function with regard to most frequencies is zero (see figure 3). The filter will therefore get limited gradients during training. In order to improve trainability, we propose generalized Gaussian filters in the frequency domain, which have a wide transition band for good gradient support.

Filtering in the frequency domain allows for full control of the filter function [26] and has previously been used in a deep learning model [27], confirming that it does not impede differentiability. The generalized Gaussian is given by

\[ G(x) = \frac{\beta}{2\alpha \Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta}, \quad (3) \]
where $\Gamma(n) = (n-1)!$ is the gamma function, $\mu$ is the centre frequency, $\alpha$ is the scale and $\beta$ is the shape parameter. We fix the gain of all filters at one and accordingly drop the normalization factor of the generalised Gaussian.

The bandwidth of a filter is conventionally defined as the point where the filter function reaches half height. Generalizing Cohen’s reparameterisation of the normal distribution to replace the standard deviation with a full-width half-maximum (FWHM) parameter [28], we reparameterise the scale $\alpha$ with the FWHM parameter $h$. To derive the new form, we set the function value equal to 0.5 and solve for the solutions of $x$ (i.e. points/frequencies at which the function is at half height), depending on the centre frequency $\mu$ and function shape $\beta$:

$$
\begin{align*}
  \frac{1}{2} &= e^{-\frac{1}{2}((x-\mu)/\alpha)^\beta} \\
  2^{-1} &= e^{-\frac{1}{2}((x-\mu)/\alpha)^\beta} \\
  -\ln(2) &= -((x-\mu)/\alpha)^\beta \\
  \ln(2) &= \frac{|x-\mu|}{\alpha^\beta} \\
  \alpha^\beta \ln(2) &= |x-\mu| \\
  \alpha \ln(2)^{1/\beta} &= |x-\mu|,
\end{align*}
$$

resulting in the two points $x_1$ and $x_2$

$$
\begin{align*}
  x_1 &= \mu - \alpha \ln(2)^{1/\beta} \\
  x_2 &= \mu + \alpha \ln(2)^{1/\beta}.
\end{align*}
$$

The FWHM can then be defined as the distance between the two points at which the function reaches half height as

$$
\begin{align*}
  h &= x_2 - x_1 \\
  h &= \mu + \alpha \ln(2)^{1/\beta} - \left(\mu - \alpha \ln(2)^{1/\beta}\right) \\
  h &= 2\alpha \ln(2)^{1/\beta} \\
  \alpha &= \frac{h}{2 \ln(2)^{1/\beta}}.
\end{align*}
$$

This allows us to use the filter bandwidth to reparameterize the scale ($\alpha$) of the generalised Gaussian, which is more intuitive for signal processing applications. The filter $F$ is then finally defined as

$$
F(x) = e^{-\frac{1}{2}((x-\mu)/\alpha)^\beta}.
$$

At $\beta = 2$ the generalised Gaussian equals the normal distribution, which corresponds to a Morlet wavelet in the time domain [28]. For higher values of the shape parameter, the filter moves towards a rectangular function, similar to a sinc filter in the frequency domain (see figure 3). In fact, the sinc filter is a special case of the generalized Gaussian in the limit of $\beta \to \infty$. This can be shown by inserting the FWHM reparameterization based on the bandwidth $h$ (equation (6)) for the scale parameter $\alpha$ into the generalised Gaussian function centred around $\mu$ as

$$
\begin{align*}
  F(x) &= e^{-\frac{1}{2}((x-\mu)/\alpha)^\beta} \\
  &= e^{\left((-\frac{x-\mu}{2 \ln(2)^{1/\beta}})^\beta\right)} \\
  &= e^{-\ln(2)((x-\mu)/\frac{2}{\ln(2)})^\beta}.
\end{align*}
$$

It is then straightforward to analyse the function in the limit of $\beta \to \infty$:

$$
\lim_{\beta \to \infty} e^{-\ln(2)((x-\mu)/\frac{2}{\ln(2)})^\beta} = \begin{cases} 
  1, & \text{if } |x-\mu| < \frac{\ln(2)}{2} \\
  0.5, & \text{if } |x-\mu| = \frac{\ln(2)}{2} \\
  0, & \text{if } |x-\mu| > \frac{\ln(2)}{2}.
\end{cases}
$$

This precisely coincides with the frequency domain representation of an ideal sinc bandpass filter given by two heavy-side step functions centred around $\mu$ with bandwidth $h$.

We clamp the shape parameter at a minimum of $\beta = 2$, since the generalised Gaussian moves towards the Laplace distribution for lower values, which is expected to interfere with training due to its non-continuous derivative.

The cost of a wide transition band is reduced frequency selectivity, which makes the function a less ideal filter. This potential problem can be addressed by making the shape parameter of the generalised
Gaussian trainable, which can then move the filter from the Gaussian towards a rectangular shape over the course of training.

Designing the filter in the frequency domain also allows for full control over the phase response. For simplicity we use linear-phase filters here with an arbitrarily chosen group delay of 20 ms, but a zero-phase or any linear or non-linear phase response could be used depending on the feature in question. The group delay of a linear-phase setup could also be added as a trainable parameter to allow for phase alignment between signals (not reported here).

We initialize all filters at 23 Hz with a wide bandwidth (44 Hz) and shape $\beta = 2$, forming a normal distribution. The filter then has a non-zero derivative for all relevant frequencies. During training, filters narrow down to specific frequencies of interest. We accelerate the training of the shape parameter by rescaling it to $\beta_{\text{rescale}} = 8\beta - 14$. For numerical stability in the generalised Gaussian function, we clamp the shape parameter at a maximum of $\beta = 3$ (corresponding to $\beta_{\text{rescale}} = 10$). The filters are trained independently for each electrode, except when using PLV connectivity, where it to $\beta_{\text{rescale}} = 8\beta - 14$. For numerical stability in the generalised Gaussian function, we clamp the shape parameter at a maximum of $\beta = 3$ (corresponding to $\beta_{\text{rescale}} = 10$). The filters are trained independently for each electrode, except when using PLV connectivity, where we take the absolute value of $|r_{XY}|$.

As an alternative estimate of functional connectivity, with a particular focus on the phase relationships between signals, we test PLVs [2]. PLV is determined by the instantaneous phases $\Delta \phi$ of two signals as

$$\text{PLV} = \frac{1}{T} \sum_t e^{i\Delta \phi(t)}.$$  \hfill (12)

Two signals with a strong coupling exhibit stable phase differences, leading to a high PLV, defined in the range $[0, 1]$. We build on the computationally efficient implementation proposed in [31] (see also [32]), while explicitly handling calculations with complex numbers as required by the PyTorch version used in this study. To decompose PLVs into efficient inner matrix products, we make use of the identity between the polar and Cartesian representation of the complex analytical signal, given by

$$Ae^{i\phi} = x + iH(x),$$  \hfill (13)

where $H(x)$ is the Hilbert transform and the envelope of the analytical signal is given by $A = \sqrt{x^2 + H(x)^2}$. We obtain the oscillatory component $e^{i\phi}$ by normalizing the analytical signal by the amplitude $A$ as

$$e^{i\phi} = \frac{x + iH(x)}{A},$$  \hfill (14)
and then define the real component $u$ and the imaginary component $iv$ of $e^{i\phi}$ as

$$
u = \frac{x}{A},$$

$$iv = \frac{H(x)}{A}.$$  (15)

The differences in instantaneous phase $\Delta \phi$ used in the PLV metric can then be expressed making use of the complex representation $u + iv$ as

$$PLV = \frac{1}{T} \left| \sum_{t} e^{i\Delta \phi} \right|$$

$$= \frac{1}{T} \left| \sum_{t} e^{i\phi_{t} - i\phi_{1}} \right|$$

$$= \frac{1}{T} \left| \sum_{t} e^{i\phi_{t} e^{-i\phi_{1}}} \right|$$

$$= \frac{1}{T} \left| \sum_{t} (u_{t} + iv_{t}) (u_{2} - iv_{2}) \right|$$

$$= \frac{1}{T} \left| \sum_{t} (u_{1}u_{2} - u_{1}iv_{2} + u_{2}iv_{1} - iv_{1}v_{2}) \right|$$

$$= \frac{1}{T} \left( \sum_{t} u_{1}u_{2} + \sum_{t} v_{1}v_{2} \right)$$

$$- i \left( \sum_{t} u_{1}v_{2} - \sum_{t} u_{2}v_{1} \right) \right|.  \quad (16)$$

The resulting equation can be implemented based on four inner matrix products, handling the real and imaginary components separately.

### 3.3. Volume conduction

At this point, it is important to notice that functional brain connectivity features, derived from EEG recordings in sensor space, are generally expected to be contaminated by volume conduction effects [33], which make signals recorded at nearby sensors to appear correlated even in the absence of actual neural connectivity [34]. Spurious connectivity, due to volume conduction, is expected to have only minimal impact on the features extracted via EEGminer for the following reasons. First, by training the classifiers to discriminate between two recording conditions (or groups), we extract only discriminative connectivity features and hence disregard spurious connectivity due to volume conduction, which is expected to behave like a constant/common factor [35]. Aligned with this consideration, in our connectivity plots we present the mined edges weighted by classifier weights and not weighted by the absolute measure of connection strength. Second, by encompassing cross-frequency correlations in our correlation models, the influence of volume conduction is further reduced [35]. Particularly for the results included in section 4, it is the low spatial resolution of the electrode array in the case of STEW used for workload classification and the MyBrainTunes dataset used for music vs rest and female vs male classification (see section 4.1), as well as the cross-frequency character of the detected dependencies in the SEED dataset that make dispensable the further signal manipulation for overcoming volume conduction. Additional steps that could be considered for mitigating the effects of volume conduction are appropriate preprocessing (e.g. by means of Laplacian transform [36] or beamformers [37]) or modifying the differentiable feature module so as to overlook zero-lag correlations (e.g. by adopting the imaginary PLV [31] version of equation (16)).

### 3.4. Classifier and feature interpretation

As a classifier we use a linear layer with sigmoid activation, essentially performing logistic regression. This has the advantage that the trained classifier weights can directly be used as importance weights attributed to the features. For non-linear classifiers the interpretation strategy would have to be adapted, e.g. by using prediction gradients with respect to the features [38]. Since different features can have different distributions, they need to be standardized in order to allow for the usage of the regression coefficients directly as an importance measure. In a machine learning setting with mini-batch gradient descent, standardization is performed via batch normalization [39], which keeps a running mean and variance over mini-batches to approximate dataset statistics. In our case we use non-affine batch normalization, meaning the target mean and standard deviation are not trainable. We use a batch size of 256, since a larger batch size leads to a better estimation of the dataset statistics. To further improve model convergence and feature standardization, we perform cosine decay to zero on the learning rate [40]. This ensures that the magnitude of model updates is reduced over time, allowing the batch normalization to capture final statistics of the discovered features.

For the band magnitude model we can investigate the full search space by plotting the relative difference in frequency magnitudes between the two classes. This allows us to check whether the discovered features are actually represented in the dataset. This is not possible for the connectivity model due to the vastly larger search space. We therefore save class-wise feature activations over the dataset and perform statistical t-tests to see whether the difference of feature means between the classes is statistically different from zero.

### 3.5. Training procedure

The training of EEGminer models is performed via mini-batch stochastic gradient descent with Nesterov momentum [41]. We found that heavy momentum on the bandpass filters helps training and improves performance. The momentum on the filters is set to 0.99, while it is set to the default value of 0.9 on the
classifier. The increased momentum on the generalized Gaussian filters was beneficial for their convergence during training. We performed only very minimal hyperparameter optimization. The learning rate is initialized at $2 \times 10^{-3}$ and decayed to zero according to a cosine schedule. The learning rate decay is necessary for the batch normalization to get final estimates of statistical feature distributions and thereby provide standardized classification features.

We train the EEGMiner models for 5000 epochs on music vs rest and workload vs rest, 300 on female vs male classification and 1200 epochs on positive vs negative emotion (these roughly correspond to the same number of total update steps). For the loss $\mathcal{L}$ we use the mean squared error function between the targets $t$ and model predictions $\hat{t}$ for simplicity. To sparsify discovered features, we use lasso regularization, which adds the L1 norm of the classifier weights $w$ to the loss with scaling factor $\gamma$, as defined by

$$\mathcal{L} = \frac{1}{N} \sum_{i} (p_i - t_i)^2 + \gamma \sum_{k} |w_k|. \quad (17)$$

The L1 penalty is tuned manually per model and classification task (EEGMiner magnitude $\gamma = 2 \times 10^{-3}$ on all tasks, EEGMiner signal correlations on positive vs negative emotion $\gamma = 3 \times 10^{-2}$, and on all other tasks $\gamma = 3 \times 10^{-3}$). The L1 penalty is mostly tuned for the desired sparsity of extracted features, while having only minor effects on classification performance. Note that no parameter regularization is performed on the bandpass filters.

4. Experimental evaluation

The usefulness of discovered features can be evaluated by the classification performance of EEGMiner models on specific tasks. We benchmark the performance of the proposed models against conventional machine learning models and deep learning models that are commonly used in the EEG domain. This serves to illustrate whether the discovered features have similar discriminative power as conventional approaches with manual feature engineering as well as using modern black box classifiers. The aim is not to directly improve on decoding performance, but to provide a research tool for discovering features of brain activity that have discriminative power close to current state-of-the-art decoding models.

4.1. Datasets

The dataset introduced here was formed from the data recorded during a very recent large-scale experiment, organized by two of the authors of this study and collected in cooperation with the Science Museum London. The data is available upon request [42]. The experiment was approved by the Science, Engineering and Technology Research Ethics Committee at Imperial College London. It consists of brain activity recorded from 761 sessions of 721 volunteer subjects (387 female, 334 male) aged 12 and above (mean = 29, std = 11). The experiment followed a music-listening design, similar to [43, 44]. A 30 second EEG baseline was recorded for each subject, after which 30 second segments of 30 songs were played in random order while recording EEG activity, for a total of 30×30 = 900 s of music listening EEG per subject. During EEG recording, participants were asked to remain still to reduce recording artifacts, to keep their eyes open and to focus on a dot in the middle of a screen to avoid introducing noise from the muscular activity of eye movements. The recordings were conducted with Emotiv EPOC+ wet electrode headsets with 14 channels at 256 Hz (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 in accordance with the 10–20 system). The song playlist was pre-defined and comprised of songs from the top of the UK charts of the last decades including songs from various genres (e.g. pop, rock, ballads, hip-hop, etc).

The female vs male classification task uses all recordings of the MyBrainTunes dataset. For the music vs rest task a single song recording is randomly chosen per subject to provide an equal number of song and resting state trials.

The SEED dataset contains EEG recordings from 15 participants while watching emotional film clips [19, 20]. Each subject participated in 3 recording sessions, viewing the same 15 clips with negative, neutral or positive emotional content. The EEG data provided include recordings from 62 electrodes at 200 Hz sampling rate. The power line noise at 50 Hz has been removed. For the positive vs negative emotion task we discard clips with neutral emotion.

The STEW dataset contains recordings from 48 subjects during resting state and while performing the Simultaneous Capacity (SIMKAP) test module of the Vienna Test System [45], serving as a mental workload condition. For each subject, a single recording of 2.5 min is provided in each condition. The EEG signals were recorded with Emotiv EPOC wet electrode headsets with 14 channels at 128 Hz.

For SEED and STEW we break up the long recordings into windows of 20 s. Windows are sampled randomly and treated independently for training samples. To account for this windowing, every recording is sampled 9 times per epoch in SEED and 8 times in STEW, which increases the total number of update steps. For validation samples, each recording is split into consecutive non-overlapping windows and the model predictions are averaged across these windows.

The datasets have been preprocessed as follows: the multichannel signals were filtered (1–45 Hz for our dataset and STEW, 0.5–99.5 Hz for SEED) using a zero-phase band-pass filter (3rd order Butterworth). To remove artifacts we resorted to wavelet-enhanced
independent component analysis (wICA) [46, 47]. Artifact suppression was carried out separately for each trial, based on an in-house implementation of wICA. Specifically, independent components (ICs) were extracted from the multichannel signal by means of the EEGLAB Matlab toolbox. Subsequently, wavelet decomposition based on wavelets of the biorthogonal family and wavelet shrinkage with a hard threshold based on false discovery rate [48] were applied to each one of the ICs. The multichannel signal was then reconstructed based on the artifact-free ICs. Hence, in this study, the use of ‘wICA-cleaned’ brain activity is implied. The EEG signals were re-referenced by common average referencing. We downsample the MyBrainTunes dataset to 128 Hz. For the deep learning baseline models, the SEED dataset was also resampled to 128 Hz. For all models, each input is standardized by its total trial mean and standard deviation. For SPDNet, we apply a bandpass filter in the 4–40 Hz range before computing the spatial covariance matrix.

### 4.2. Evaluation procedure

To equalize small class imbalances during training, we perform oversampling of underrepresented classes by using relevant trials from the training set multiple times. Cross-validation was performed for all models, using ten random folds with held-out subjects for the respective validation set. This means that for each fold, none of the samples of the subjects in the corresponding validation set were seen during training (also considering oversampling and the three sessions per subject on SEED), and results should be seen as subject-independent. The main performance metric is given as the unweighted average recall (UAR) at the end of training, reporting the mean and standard deviation across folds. The UAR for classes \( c \in C, |C| = N \) is given by

\[
UAR = \frac{1}{N} \sum_{c} \frac{\text{true_positives}_{c}}{\text{true_positives}_{c} + \text{false_negatives}_{c}}.
\]

The UAR is also known as balanced accuracy, and in the case of perfect class balance corresponds to the simple accuracy metric.

### 4.3. SVM and deep learning baselines

As a performance baseline from the conventional neuroscience perspective, we employ a linear support vector machine (SVM) on standardized features spanning the whole search space. Regarding band magnitude, the features are the signal magnitudes of all channels filtered in canonical EEG bands. In the signal correlation case, the features are all possible connections across all sensors filtered within and across all EEG bands (3486 for our dataset and STEW, 69,006 for SEED). Performance was evaluated in the same way as for the EEGminer and deep learning models, using 10-fold subject-independent cross-validation.

To benchmark our approach against deep learning models, we train ShallowNet, EEGNet, SyncNet and SPDNet on all tasks. DeepNet is omitted from the baselines here due to inferior performance when compared to ShallowNet on the studied tasks. We adjust the filter kernel sizes of ShallowNet and EEGNet for input signals sampled at 128 Hz. All four deep learning models are trained for 100 epochs using an Adam optimizer. In the case of the SPDNet, an Adam optimizer with manifold constraint was used. The optimizer learning rate is \( 1 \times 10^{-3} \) for ShallowNet and EEGNet and \( 2 \times 10^{-3} \) for SyncNet, as proposed by the respective authors. The learning rate used for SPDNet is \( 1 \times 10^{-3} \). The loss function is binary cross-entropy between targets and predictions for all deep learning models. ShallowNet and EEGNet use a batch size of 16, while SyncNet uses 10 and SPDNet uses 30. The dropout rates for ShallowNet and EEGNet are 0.5 and 0.25 respectively. SyncNet uses dropout of input channels, which we employ with a rate of 0.25 only on the SEED dataset, as it gave inferior performance when used on the MyBrainTunes and STEW datasets (likely due to the small number of electrodes). No weight decay is used. For EEGNet we use 4 temporal filters on music vs rest, positive vs negative emotion and workload vs rest as the number of training samples is somewhat limited, while we use 8 temporal filters on female vs male classification. For SPDNet we use 3 BiMap layers, with hidden size of 16 on SEED and 14 on the other tasks.

### 4.4. Performance evaluation

We evaluate the performance of all tested models using cross-validation and the UAR metric. Table 1 illustrates the cross-validation performance on all four tasks. The performance of the proposed EEGminer models is on par with a linear SVM trained on pre-computed signal magnitude or signal correlations using canonical EEG bands. Performing a \( t \)-test between the strongest EEGminer and SVM models on each task reveals no significant differences in performance (Music vs Rest comparing Ours mag. vs SVM mag.: \( t = 1.27, p = 0.24 \); Female vs Male comparing Ours corr. vs SVM corr.: \( t = -1.75, p = 0.11 \); Pos vs Neg Emotion comparing Ours mag. vs SVM mag.: \( t = 0.00, p = 1.0 \); Workload vs Rest comparing Ours PLV vs SVM corr.: \( t = 0.37, p = 0.72; n = 10 \)). This shows that EEGminer models can discover the relevant features to match the performance of a conventional machine learning model that operates on a
large pre-computed search space, while having a focus on clear feature interpretability.

Furthermore, the classification performance of EEGminer models is on a comparable level to the deep learning baselines. Performing a t-test between the strongest EEGminer and deep learning baseline models on each task reveals no significant differences in performance, except on the Female vs Male classification task, where EEGNet performs better than the EEGminer at the 5% level (Music vs Rest comparing Ours mag. vs ShallowNet: $t = -0.54$, $p = 0.60$; Female vs Male comparing Ours corr. vs EEGNet: $t = -2.63$, $p = 0.03$; Pos vs Neg Emotion comparing Ours mag. vs SyncNet: $t = 0.64$, $p = 0.54$; Workload vs Rest comparing Ours PLV vs EEGNet: $t = -0.27$, $p = 0.79$; $n = 10$). The temporal resolution that CNNs offer is well-suited for stimulus-locked recordings, such as event-related potentials or motor imagery, but might have relatively limited use on the spontaneous nature of the classification tasks studied here. This indicates that our proposed model limits trainable components to the relevant feature space, cutting unnecessary capacity and thereby reducing the potential for overfitting.

Regarding connectivity features, both signal correlations and PLV achieved similar performance. PLV is limited to within-band connectivity, while signal correlations are also defined across frequencies and accordingly have a larger feature space. We therefore focus on band magnitudes and signal correlations when discussing the mined features.

### 4.5. Mined features

Benefiting from the clear interpretability of EEGminer models, we can investigate the trained features in terms of relevant frequencies and functional connectivity. On the music vs rest classification task we present the full feature space as utilized by the respective models, while we focus on the most impactful features for female vs male classification during music listening, positive vs negative emotion recognition and workload vs rest classification.

When plotting top features on the latter tasks, averaging across folds is difficult. We reverted to the fold with highest validation performance. Using two filter channels can result in duplicate features, which we manually removed. An analysis of the sensitivity of the classification loss with regard to small changes in the learned filter parameters can be found in appendix A.2.

Looking at the trained magnitude EEGminer model on music vs rest classification (figure 4), it is apparent that the proposed generalised Gaussian filters exhibit strong trainability of the centre frequency, bandwidth and shape, departing substantially from their initialized state. Many of the filters with high model weight narrow down to specific EEG bands (e.g. alpha in O2, theta in P7 and high beta in AF3). Multiple bands of interest can be found for the same sensor when using a model with multiple filter channels (e.g. P7, F8). Features are sparsified by L1 regularization on the linear classifier, forcing the model to disregard locations with minor contribution (e.g. FC5, FC6 and F3). The trained filters also exhibit good stability over folds, as seen in the small difference of mean and max filter importance, meaning the discovered features are consistent across reruns on other training folds.

Relevant band magnitude features as discovered by the EEGminer model on music vs rest classification are decreased activity of lower frequencies in frontal electrodes (AF3, AF4, F7 and F8) during music listening and increased activity in high beta and gamma (AF3, AF4 and F8). Also used by the model are increased alpha activity in occipital electrodes (O1 and O2) during music listening and decreased gamma as well as increased theta activity in P7. Decreased gamma activity in T7 and in low frequencies in T8 are also associated with music listening.

---

**Table 1.** Validation performance and number of trainable parameters for each of the tested models on four different tasks (mean ± std across 10 folds).

|                                | Music vs Rest | Female vs Male | Pos vs Neg Emotion | Workload vs Rest |
|--------------------------------|---------------|----------------|-------------------|------------------|
|                                | params val UAR | params val UAR | params val UAR | params val UAR |
| Ours mag.                      | 113 ± 0.03    | 113 ± 0.05     | 497 ± 0.11       | 113 ± 0.08       |
| Ours corr.                     | 267 ± 0.02    | 267 ± 0.03     | 4155 ± 0.15      | 267 ± 0.10       |
| Ours PLV                       | 189 ± 0.02    | 189 ± 0.03     | 3789 ± 0.13      | 189 ± 0.05       |
| SVM mag.                       | 84 ± 0.04     | 84 ± 0.03      | 434 ± 0.12       | 84 ± 0.12        |
| SVM corr.                      | 348 ± 0.03    | 348 ± 0.02     | 69 ± 0.16        | 348 ± 0.16       |
| ShallowNet                     | 42 ± 0.05     | 42 ± 0.03      | 112 ± 0.10       | 35 ± 0.07        |
| EEGNet                         | 157 ± 0.05    | 157 ± 0.03     | 1637 ± 0.15      | 125 ± 0.10       |
| SyncNet                        | 126 ± 0.04    | 126 ± 0.02     | 1901 ± 0.10      | 94 ± 0.11        |
| SPDNet                         | 785 ± 0.03    | 785 ± 0.04     | 1761 ± 0.12      | 785 ± 0.11       |

* The number of parameters for the SVM is given by the number of input features.
When using signal correlations as the feature module to classify music listening vs resting state (figure 5), the most relevant features are cross-hemispheric connections between the left temporal/parietal with the right frontal areas. Most specifically during music listening there is increased coupling between high-frequency activity in T7 and low frequencies in F8. Also relevant is decreased local connectivity of gamma brainwaves between FC5 and F7 and of high frequencies in FC6 with low frequencies in F8. It can be observed that the model identifies a hemispheric specialisation of the temporal areas, focusing on cross-hemispheric connections of the left temporal area and within-hemisphere connections of the right temporal area. This agrees with the previously proposed respective specialisation of the temporal lobes regarding music perception [50], and specific findings for the role of left temporal to right frontal connectivity in music processing [51].

Regarding female vs male classification during music listening, the magnitude EEGminer model mostly focuses on decreased connectivity in lower frequencies for females and increased connectivity in the gamma band, particularly in cross-hemispheric symmetric connections. Specifically, reduced connectivity between alpha in P7 and low beta in O1 is associated with female subjects, as well as reduced connectivity between FC5 and FC6 and reduced beta connectivity between O1 and O2 and FC5 and F7. Increased connectivity in female subjects is found in gamma activity between F3 and F4. In the relevant literature, overall higher cross-hemispheric coherence was found for female subjects during music listening [53]. While this is less evident in the top features found by the EEGminer model on the specific cross-validation fold, except for a connection between F3 and F4, this general increase in cross-hemispheric connectivity can be observed in the full feature space learned by the model (see supplementary figure 4).

The magnitude EEGminer model uses differences in gamma activity to classify emotion (see figure 6). Positive emotion is associated with increased gamma activity in temporal areas (T7, T8 and FT8) compared to negative emotion, as well as decreased gamma activity in parietal electrodes (CP3 and P6). The association of emotional state and high frequency activity in the temporal areas has previously been reported on the SEED dataset [54, 55]. The most important couplings detected by the correlation EEGminer model are increased within-hemisphere connectivity during positive emotion between central and frontal areas.

Figure 4. Trained magnitude EEGminer model on music listening vs resting state classification, indicating important features of brain activity for 721 subjects. Filter lines show the averaged filters across folds scaled by the model weights, with shaded areas denoting the maximum weight and colors indicating the feature direction. Since two freely moving filters were used per electrode, averaging of the filters over folds is performed according to the respective centre frequency (i.e. average all low-centre filters and all high-centre filters). Each subplot shows the relative filter importance plotted against the frequency in Hz. The inset shows the filter initialization. The limits of canonical EEG bands are given as dotted lines.

Figure 5. Signal correlations between resting state and music listening for 721 subjects. The colors indicate the direction of the correlation and the weight of the correlation is denoted by the size of the box. The limits of canonical EEG bands are given as dotted lines.
Figure 5. Trained correlation EEGminer model on music listening vs resting state classification, indicating important features of brain activity for 721 subjects. Connections are scaled by the mean model weights across folds, with the colors indicating the feature direction. Filter lines show the averaged filters across folds scaled by the model weights on the strongest adjacent connection, with shaded areas denoting the maximum weight. Since two freely moving filters were used per electrode, averaging of the filters over folds is performed according to the respective centre frequency (i.e. average all low-centre filters and all high-centre filters), and the stronger of the two model weights is used for the connection. Each subplot shows the relative filter importance plotted against the frequency in Hz. The inset shows the filter initialization. The limits of canonical EEG bands are given as dotted lines.

(C3 with FC5 and C4 with F8) and decreased within-hemisphere connectivity of temporal with adjacent electrodes (FT8 with T8, FT7 with T7 and TP7 with T7). These findings are supported by recent literature [56].

To classify workload vs rest, the magnitude EEGminer model uses mostly changes in very specific frequency bands (see figure 6). These include decreased magnitude of alpha and delta bands in occipital electrodes (O1 and O2) during the workload condition, as well as increased magnitudes in the delta and theta bands of the right frontal-temporal area (FC6 and T8) and in the alpha to low beta band in the left frontal area (F7). These findings mostly agree with those reported by the authors of the STEW dataset, which include significant activity in frontal and occipital areas [21]. The correlation EEGminer model finds reduced long-range connectivity in all bands from delta to high beta during the workload condition, mostly in the left hemisphere, and increased high-beta to gamma connectivity between O1 and O2. Reduced frontal-parietal connectivity in the beta band during the workload condition, particularly in the left hemisphere, agrees with findings in the literature [57].

4.6. Feature validation
To validate the features found by EEGminer models on our dataset (music vs rest, female vs male classification during music listening), we created the relative change of grand averaged magnitude profiles for the magnitude models, and violin plots with t-test statistics showing the distributions of the features used by signal correlation models. Since the search space for connectivity models is exceedingly large, we limit ourselves to the top features here.

Figure 7(a) illustrates the relative change in grand averaged magnitude profiles during music listening relative to resting states for all sensors and available frequencies. Most of the features mined by the model are strongly represented in the grand average (reduced gamma in P7, increased alpha in O1 and O2, decreased delta to alpha in F7 and F8 and increased high beta and gamma in F8). Some mined features are ambiguous in the grand average, specifically increased theta in P7, which is shown in the alpha band by the grand average, and increased high beta in AF4, which is shown in low beta by the grand average.

The relative change in grand averaged magnitude profiles for female relative to male subjects during music listening is given in figure 7(b). All five of
the top features discovered by the EEGminer model are strongly represented by the differences in grand averages (increased low beta in O1 and O2 for females, increased beta in AF3 and gamma in AF4, reduced theta and alpha in P8).

Violin plots showing the distributions of music listening and resting state EEG across all subjects (rest and one song each) of top features discovered by the EEGminer signal correlation model are given in figure 8. All feature directions agree between the EEGminer model and the feature distributions in the dataset, although some of the features show low statistical significance in the difference of means. This can be attributed to the limited number of training examples using only one randomly chosen song per subject to equalize class balance, as well as the somewhat limited classification performance of the model (0.68 ± 0.02).

Figure 9 shows the distributions of top features discovered by the EEGminer signal correlation model.
Figure 7. Relative change in grand averaged magnitude profiles for (a) music-listening relative to resting state and (b) female relative to male subjects during music listening. These profiles are derived purely from the EEG data without involvement of the EEGminer model and are used for visual feature validation.

Figure 8. Violin plots of unstandardized top features discovered by the correlation EEGminer model trained on music vs rest classification. Dashed lines indicate the quartiles of each feature distribution on all samples, including train and test set. The associated statistical significance is given above each distribution by a paired $t$-test between the music and resting state trial of each subject: $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$, $^{****} p < 0.0001$.

Figure 9. Violin plots of unstandardized top features discovered by the correlation EEGminer model trained on female vs male classification during music listening. Dashed lines indicate the quartiles of each feature distribution on all samples, including train and test set. The associated statistical significance is given above each distribution by an independent two-samples $t$-test between male and female subject groups: $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$, $^{****} p < 0.0001$. 
for male and female subjects during music listening. All feature directions agree between the model and the feature distributions and all of the top seven features show strong statistical mean differences in the t-tests.

5. Conclusion

We presented novel trainable filters parameterized by the generalised Gaussian function, which include Morlet wavelets and sinc filters as special cases and allow the model to select the appropriate filter during end-to-end training. Using differentiable implementations of neuroscientifically plausible EEG features like band magnitude and signal correlations, our proposed model can discover classification-relevant frequency bands and functional connectivity patterns among a large repertoire of possible features, generally matching the decoding accuracy of established deep learning models that have some aspect of interpretability. While a number of deep learning approaches with varying degrees of interpretability do exist, they make much less use of strong constraints on the trained features and thereby limit the precision in the interpretation. The constraints we place on the feature extraction in the model allow for direct interpretability, but could on the downside also overly restrict the feature space if the wrong feature type is chosen for the given task. In future work, going beyond the static nature of features used here, a feature mining mechanism with higher time resolution should be developed for event-related tasks.

The current application of deep learning models on SPD matrices, like SPDNet, in the context of EEG is limited by the choice of input matrix. In the context of this paper, a wide-band covariance matrix was used as a reasonable choice. This approach could benefit significantly from the trainable filters and connectivity estimation developed for EEGminer, which in the case of signal correlations produces SPD matrices.

Implementing an EEG pipeline as a differentiable deep learning model could bridge the gap between conventional research in neuroscience and difficult to interpret black-box methods commonly employed in the deep learning field. Such a model promises comparable performance with deep learning models as well as neuroscientific insights. This has important implications in practical applications as well, where model interpretability is often a strong requirement.

Data availability statement

The data is available upon request at the following URL: https://mybraintunes.doc.ic.ac.uk/ The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

We would like to thank the Science Museum London for hosting the data collection experiment.

ORCID iDs

Siegfried Ludwig  
https://orcid.org/0000-0001-9693-9482

Stylianos Bakas  
https://orcid.org/0000-0003-1054-0169

Dimitrios A Adamos  
https://orcid.org/0000-0001-6700-1057

Nikolaos Laskaris  
https://orcid.org/0000-0002-1960-394X

Yannis Panagakis  
https://orcid.org/0000-0003-0153-5210

Stefanos Zafeiriou  
https://orcid.org/0000-0002-5222-1740

References

[1] Kida T, Tanaka E and Kakigi R 2016 Multi-dimensional dynamics of human electromagnetic brain activity Front. Hum. Neurosci. 9 713

[2] Lachaux J-P, Rodriguez E, Martinerie J and Varela F J 1999 Measuring phase synchrony in brain signals Hum. Brain Mapp. 8 194–208

[3] LeCun Y, Bengio Y and Hinton G 2015 Deep learning Nature 521 436–44

[4] LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W and Jackel L D 1989 Backpropagation applied to handwritten zip code recognition Neural Comput. 1 541–51

[5] Rudin C 2019 Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead Nat. Mach. Intell. 1 206–15

[6] Roy Y, Banville H, Albuquerque I, Gramfort A, Falk T H and Faulbert J 2019 Deep learning-based electroencephalography analysis: a systematic review J. Neural Eng. 16 031001

[7] Craik A, He Y and Contreras-Vidal J L 2019 Deep learning for electroencephalogram (EEG) classification tasks: a review J. Neural Eng. 16 031001

[8] Lawhern V J, Solon A J, Waytowich N R, Gordon S M, Hung C P and Lance B J 2018 Eegnet: a compact convolutional neural network for EEG-based brain–computer interfaces J. Neural Eng. 15 056013

[9] Ravaneli M and Bengio Y 2018 Speaker recognition from raw waveform with sincnet 2018 IEEE Spoken Language Technology Workshop (SLT) (IEEE) pp 1021–8

[10] Borra D, Fantozzi S and Magosso E 2019 EEG motor execution decoding via interpretable sinc-convolutional neural networks Mediterranean Conf. on Medical and Biological Engineering and Computing (Springer) pp 1113–22

[11] Bria A, Marrocco C and Tortorella F 2021 Sinc-based convolutional neural networks for EEG-BCI-based motor imagery classification (arXiv:2101.10846)

[12] Priyasad D, Fernando T, Denman S, Sridharan S and Fookes C 2021 Interpretable seizure classification using unprocessed EEG with multi-channel attention feature fusion IEEE Sens. J. 21 19186–97

[13] Chen J, Wang D, Yi W, Xu M and Tan X 2023 Filter bank sinc-convolutional network with channel self-attention for high performance motor imagery decoding J. Neural Eng. 20 020001

[14] Li Y, Murias M, Major S, Davson G, Dzirasa K, Carin L and Carlson D E 2017 Targeting EEG/LFP synchrony with neural
nent Con. on Neural Information Processing Systems pp 4620–30
[15] Cui J, Lan Z, Sourina O and Müller-Wittig W 2022 EEG-based cross-subject driver drowsiness recognition with an interpretable convolutional neural network IEEE Trans. Neural Netw. Learn. Syst. 34 7921–33
[16] Borja D, Magosso E, Castelo-Branco M and Simões M 2022 A bayesian–optimized design for an interpretable convolutional neural network to decode and analyze the p300 response in autism J. Neural Eng. 19 046010
[17] Seki H, Yamamoto K and Nakagawa S 2017 A deep neural network integrated with filterbank learning for speech recognition 2017 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) (IEEE) pp 5480–4
[18] Zhao D, Tang F, Si B and Feng X 2019 Learning joint space–time–frequency features for EEG decoding on small labeled data Neural Netw. 114 67–77
[19] Duan R-N, Zhu J-Y and Lu B-L 2013 Differential entropy feature for EEG-based emotion classification 2013 6th Int. IEEE/EMBS Conf. on Neural Engineering (NER) (IEEE) pp 81–84
[20] Zheng W-L and Lu B-L 2015 Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks IEEE Trans. Auteur. Mental Dev. 7 162–75
[21] Lim W L, Sourina O and Wang L P 2018 Stew: simultaneous task EEG workload data set IEEE Trans. Neural Syst. Rehabil. Eng. 26 2106–14
[22] Tibor Schirmeister R, Tobias Springenberg J, Dominique Josef Fiederer L, Glasmeter M, Eggersperger K, Tangermann M, Hutter F, Burgard W and Ball T 2017 Deep learning with convolutional neural networks for EEG decoding and visualization and Simulating Brain Hum. Brain Mapp. 38 5391–420
[23] Keng Ang K, Yang Chin Z, Wang C, Guan C and Zhang H 2012 Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b Front. Neurosci. 6 39
[24] Huang Z and Van Gool L 2017 A riemannian network for spatio matrix learning 31st AAAI Conf. on Artificial Intelligence [51]
[25] Hajimoradi M, Zhang J and Huang Y 2017 Prediction of fatigue-related driver performance from EEG data by deep riemannian model 2017 39th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE) pp 4167–70
[26] Widmann A, Schröger E and Maess B 2015 Digital filter design for electrophysiological data—a practical approach J. Neurosci. Meth. 250 34–46
[27] Brosch T and Jam R 2015 Efficient training of convolutional deep learning networks in the frequency domain for application to high-resolution 2d and 3d images Neural Comput. 27 21–27
[28] Cohen M X 2019 A better way to define and describe morlet wavelets for time-frequency analysis NeuroImage 199 81–86
[29] Bakhshayesh H, Fitzgibbon S P, Janani A S, Grummett T S and Pope K J 2019 Detecting synchrony in EEG: a comparative study of functional connectivity measures Comput. Biol. Med. 105 1–15
[30] Pandit V and Schuller B 2019 The many-to-many mapping between concordance correlation coefficient and mean square error (arXiv:1902.05180)
[31] Bruna R, Maestu F and Pereda E 2018 Phase locking value revisited: teaching new tricks to an old dog J. Neural Eng. 15 056011
[32] Georgiadis K, Laskaris N, Nikolopoulos S and Kompatariatis I 2018 Exploiting the heightened phase synchrony in patients with neuromuscular disease for the establishment of efficient motor imagery BCIS J. Neuroeng. Rehabil. 15 1–18
[33] van den Broek S P, Reinders F, Donderswinkel M and Peters M J 1998 Volume conduction effects in EEG and MEG Electroencephalogr. Clin. Neurophysiol. 106 522–34
[34] Srinivasan R, Winter W R, Ding J and Nunez P L 2007 EEG and MEG coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics J. Neurosci. Methods 166 41–52
[35] Cohen M X 2014 Analyzing Neural Time Series Data: Theory and Practice (MIT Press)
[36] Nunez P L, Silverstein R B, Cadusch P J, Wijesinghe R S, Westdorp A F and Srinivasan R 1994 A theoretical and experimental study of high resolution EEG based on surface laplacians and cortical imaging Electroencephalogr. Clin. Neurophysiol. 90 40–57
[37] Gross J, Kujala J, Hämäläinen M, Timmermann L, Schnitzler A and Salmen R 2001 Dynamic imaging of coherent sources: studying neural interactions in the human brain Proc. Natl. Acad. Sci. 98 694–9
[38] Lundberg S M and Lee S-I 2017 A unified approach to interpreting model predictions Advances in Neural Information Processing Systems pp 4765–74
[39] Ioffe S and Szegedy C 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift (arXiv:1502.03167)
[40] Loshchilov I and Hutter F 2016 SGDR: stochastic gradient descent with warm restarts (arXiv:1608.03983)
[41] Sutskever I, Martens J, Dahl G and Hinton G 2013 On the importance of initialization and momentum in deep learning Int. Conf. on Machine Learning pp 1139–47
[42] Ludwig S, Bakas S, Adams D A, Laskaris N, Panagakis Y and Zaferiou S MyBrainTunes 2024 (available at: https://mybraintunes.doc.ic.ac.uk/)
[43] Bakas S, Adams D A and Laskaris N 2021 On the estimate of music appraisal from surface EEG: a dynamic-network approach based on cross-sensor pac measurements J. Neural Eng. 18 046073
[44] Adams D A, Dimitriades S I and Laskaris N A 2016 Towards the bio-personalization of music recommendation systems: a single-sensor EEG biomarker of subjective music preference Inf. Sci. 343 94–108
[45] Brattich O and Hagman E 2008 Simkap–multiskanpakkaitsa/ multi-tasking (available at: https://marketplace.schuwfried.com/en/SIMKAP) Modling: Schuwfried GmbH
[46] Castellanos N P and Makarov V A 2006 Recovering EEG signal brain signals: artifact suppression with wavelet enhanced independent component analysis J. Neurosci. Methods 158 300–12
[47] Dolemne A, Sejnowski T and Makeig S 2007 Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis NeuroImage 34 1443–9
[48] Abramovich F and Benjamini Y 1996 Adaptive thresholding of wavelet coefficients Comput. Stat. Data Anal. 22 351–61
[49] Cortes C and Vapnik V 1995 Support-vector networks Mach. Learn. 20 273–97
[50] Zatorre R J, Belin P and Penhune V B 2002 Structure and function of auditory cortex: music and speech Trends Cogn. Sci. 6 37–46
[51] Herojo Ruiz M, Koelsch S and Bhattacharya J 2009 Decrease in early right alpha band phase synchronization and later gamma band oscillations in processing syntax in music Hum. Brain Mapp. 30 1207–25
[52] Van Putten M J A M, Olbrich S and Arms M 2018 Predicting go from brain rhythms with deep learning Sci. Rep. 8 1–7
[53] Johnson J K, Petch H, Richter P, Von Stein A and Filz O 1996 The dependence of coherence estimates of spontaneous
EEG on gender and music training *Music Perception* 13 563–81

[54] Zheng W-L, Zhu J-Y and Lu B-L. 2017 Identifying stable patterns over time for emotion recognition from EEG *IEEE Trans. Affect. Comput.* 10 417–29

[55] Li X, Song D, Zhang P, Zhang Y, Hou Y and Hu B. 2018 Exploring EEG features in cross-subject emotion recognition *Front. Neursci.* 12 162

[56] Wu X, Zheng W-L and Lu B-L. 2019 Identifying functional brain connectivity patterns for EEG-based emotion recognition 2019 9th Int. IEEE/EMBS Conf. on Neural Engineering (NER) (IEEE) pp 235–8

[57] Dimitrakopoulos G N, Kakkos I, Dai Z, Lim J, deSouza J, Bezerianos A and Sun Y. 2017 Task-independent mental workload classification based upon common multiband EEG cortical connectivity *IEEE Trans. Neural Syst. Rehabil. Eng.* 25 1940–9