

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

# EEG-ITNet: An Explainable Inception Temporal Convolutional Network for motor imagery classification

Abbas Salami<sup>1</sup>, Student Member, IEEE, Javier Andreu-Perez<sup>1</sup>, Senior Member, IEEE, and Helge Gillmeister<sup>2</sup>

<sup>1</sup>School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ UK

<sup>2</sup>Department of Psychology, University of Essex, Colchester, CO4 3SQ UK

Corresponding author: Javier Andreu-Perez (e-mail: javier.andreu@essex.ac.uk).

**ABSTRACT** In recent years, neural networks and especially deep architectures have received substantial attention for EEG signal analysis in the field of brain-computer interfaces (BCIs). In this ongoing research area, the end-to-end models are more favoured than traditional approaches requiring signal transformation pre-classification. They can eliminate the need for prior information from experts and the extraction of handcrafted features. However, although several deep learning algorithms have been already proposed in the literature, achieving high accuracies for classifying motor movements or mental tasks, they often face a lack of interpretability and therefore are not quite favoured by the neuroscientific community. The reasons behind this can be the high number of parameters and the sensitivity of deep neural networks to capture tiny yet unrelated discriminative features. We propose an end-to-end deep learning architecture called EEG-ITNet and a more comprehensible method to visualise the network learned patterns. Using inception modules and causal convolutions with dilation, our model can extract rich spectral, spatial, and temporal information from multi-channel EEG signals with less complexity (in terms of the number of trainable parameters) than other existing end-to-end architectures such as EEG-Inception and EEG-TCNet. By an exhaustive evaluation on dataset 2a from BCI competition IV and OpenBMI motor imagery dataset, EEG-ITNet shows up to 5.9% improvement in the classification accuracy in different scenarios with statistical significance compared to its competitors. We also comprehensively explain and support the validity of network illustration from a neuroscientific perspective. We have also made our code freely accessible at <https://github.com/AbbasSalami/EEG-ITNet>.

**INDEX TERMS** Brain-computer interface, deep learning, deep neural network visualisation, inception module, motor imagery, temporal convolutional network

## I. INTRODUCTION

BRAIN-COMPUTER interfaces (BCIs) were developed to provide direct communication between the human brain and external devices through the acquisition of neural signals and their translation and transmission as control commands to assistive devices [1]. Although initially designed to help people with cognitive or physical impairments, BCIs have a wide range of applications nowadays, ranging from medical and neuroergonomics to entertainment, intelligent environments, and even security and authentication [2]. From different BCI systems based on varying neuroimaging modalities such as fNIRS, fMRI, MEG, and Pet, electroencephalogram (EEG) is of great importance and one

of the popular techniques due to its convenient and inexpensive nature. The electrical activities of a population of neurons can generate adequately high electrical fields that can be measured by electrodes placed on the surface of the scalp. However, the recorded signals have a low signal-to-noise ratio (SNR) due to several factors such as interference from other physiological signals, electrode detachment, or signal distortion resulting from the cortex's behaviour as a volume conductor [3]. Therefore, advanced signal processing and machine learning algorithms have been extensively applied in this area to increase the SNR and the quality of the extracted neural patterns of interest [4].

Notably, we can recall many feature extraction algorithms that were applied successfully to classify motor imagery movements, including common spatial pattern (CSP) and its variants, Riemannian approaches, Fourier transform, Wavelet transforms, and Autoregressive models followed by different classification algorithms, such as LDA, SVM, MLP, and KNN [5, 6]. Moreover, deep learning algorithms have recently accompanied these state-of-the-art algorithms to obtain even more promising and reliable outcomes [4, 7-9]. For instance, Sakhavi et al. [10] proposed a novel temporal representation of a multi-channel EEG signal based on Hilbert transform and a convolutional neural network (CNN) to classify motor imagery movements. Wang et al. [11] trained CNN and long short-term memory (LSTM) models using multi-dimensional features derived from the representation of EEG signal using their Short Time Fourier Transform (STFT). Although the performances of such algorithms have been superior to the conventional ones, end-to-end deep learning architectures that eliminate the need for prior information are more favoured than traditional approaches that require the extraction of hand-crafted features or signals transformation pre classification. A powerful *end-to-end neural network* that permits the *explainability* of its inference can significantly impact the development of robust BCI systems and reduce the need for expertise. Hence, this paper proposes a CNN structure called EEG-ITNet based on inception modules and causal convolutional layers with dilation in the form of residual blocks.

Convolution with dilation is the foundation of temporal convolutional networks (TCNs). The TCN term was first coined by Lea et al. [12] and used as a new and effective class of temporal models for action segmentation and detection. Since then, TCNs have been suggested as a promising alternative to RNNs for sequence modelling by providing a unified approach to encode spatial-temporal information and capture high-level temporal information hierarchically. With new ways of network visualisation for interpretation, we will demonstrate how inception modules along with convolutional layers with dilation can extract rich and meaningful spectral, spatial, and temporal features from multi-channel EEG signals. The **main highlights of this paper** are:

- The proposed deep architecture has few trainable parameters, allowing its transformation into intuitive visualisation structures such as topographic maps (topo-maps). We provide a methodology for achieving this.
- A notable gain has been achieved for signal classification on the motor imagery BCI dataset in various scenarios compared to other end-to-end architectures.
- The conventional vanilla convolutions have been upgraded to inception modules and depthwise causal convolutions with dilation to achieve even higher performance with less complexity compared to similar deep learning models.

The rest of the paper will be structured as follows. In section II, we review the existing similar end-to-end neural network architectures for EEG analysis, particularly for motor imagery BCI systems. Then in section III, we extensively describe our method and architecture by explaining the fundamentals first. The section also contains full details of our visualisation techniques. In section IV, we confirm the effectiveness of our proposed network on EEG signal classification in three different scenarios followed by network visualisation. Finally, we discuss and conclude the results and extracted features in sections V and VI, respectively, including our future work plan.

## II. RELATED WORK

In the literature of end-to-end architectures, Schirrneister et al. [13] proposed ShallowNet and DeepConvNet for EEG decoding based on a raw dataset, which consisted of a shallow and deep sequence of convolutional layers, respectively. Although they reported comparable results with the state-of-the-art methods, the high number of parameters, especially in DeepConvNet, made the networks too complex to analyse and interpret. Later, Lawhern et al. [14] introduced EEGNet, a more compact and efficient CNN architecture with few parameters and fast training nature. EEGNet showed encouraging results on several types of EEG datasets with significantly fewer parameters than ShallowNet and DeepConvNet. In addition, the simple architecture of EEGNet has made it a notable candidate for EEG analysis in different scenarios. Although the authors attempted to visualise and interpret the features extracted from the network, the outcome is not very intuitive. In a more recent study, EEG-Inception [15] was introduced as an end-to-end architecture utilising inception modules originally proposed by Szegedy et al. [16] in computer vision. However, although EEG-Inception reached superior results than other CNN architectures, the network lacked interpretability as the authors also made no attempt in explaining the learning process and extracted features. Similarly, we should also recall MI-EEGNET [17] developed based on Inception and Xception architectures for motor imagery classification. Although MI-EEGNET performed better than previous architectures, the high number of trainable parameters made it challenging to interpret. Besides these CNN architectures investigated for EEG signal classification, several studies can be found in the literature that used recurrent neural network (RNN) and its variants [18-20], such as LSTM and gated recurrent units (GRU), for the classification of mental tasks based on EEG signals [21, 22]. However, RNNs are less prevalent in this area due to their exploding/vanishing gradient or lack of memory problems [23].

In response to the slow training essence of RNNs for EEG signal analysis, TCNs have been recently applied and demonstrated promising results for temporal analysis of EEG time series with faster computation. In this regard, Ingolfsson et al. [23] and Musallam et al. [24] have reported

TABLE I  
COMPARISON OF THE KEY FACTORS OF EEG-ITNet WITH OTHER END-TO-END ARCHITECTURES

| Property<br>Work   | Inception | Dilation | Visualisation | Interpretability | Deepness<br>(# layers) | Parametric<br>complexity |
|--------------------|-----------|----------|---------------|------------------|------------------------|--------------------------|
| EEGNet 8,2 [14]    | No        | No       | Yes           | Partial          | Low (7)                | Low (~2k)                |
| EEG-TCNet [23]     | No        | Yes      | No            | No               | Medium (17)            | High (~5k)               |
| EEG-Inception [15] | Yes       | No       | No            | No               | High (23)              | Very high (~15k)         |
| EEG-ITNet          | Yes       | Yes      | Yes           | Yes              | Very high (31)         | Medium (~3k)             |

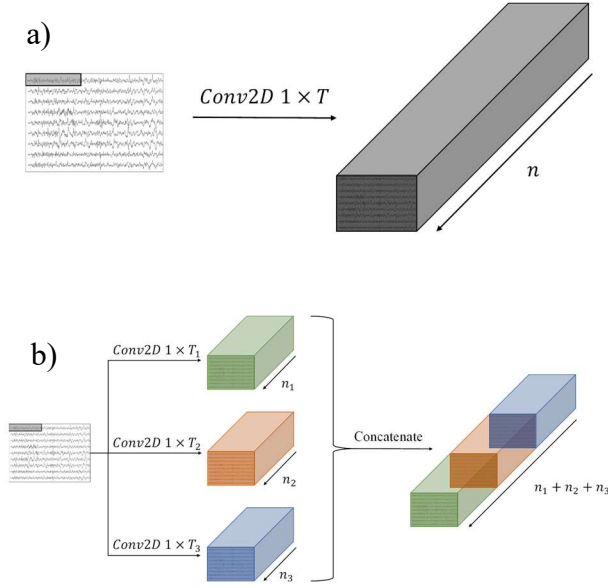


Fig. 1. Conventional convolutional layer (a) and its alternative with inception modules (b)

superior results for motor imagery signal classification by adding TCN in their structure. TCNs have also been successfully applied before to analyse different EEG datasets, including mental workload assessment [25], pathological versus non-pathological EEG classification [26], and recently for quantitative scoring of depression [27]. Table I compares the key factors of EEG-ITNet with some other existing end-to-end neural networks, which we later use as comparison models.

### III. METHOD

This section extensively explains the fundamental concepts and blocks of EEG-ITNet, followed by a detailed illustration of the proposed architecture. This section also contains a description of our visualisation and interpretation process.

#### A. BACKGROUND

This subsection defines the theories used to form the fundamental blocks of EEG-ITNet.

1) *Inception*: In order to break the multi-channel EEG signal into its informative frequency sub-bands, this study uses

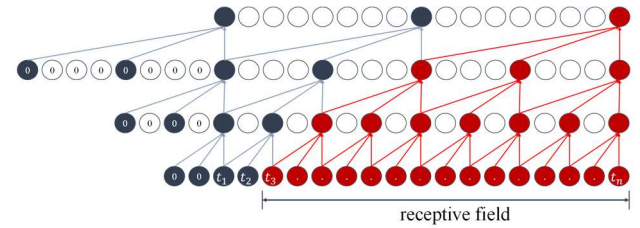


Fig. 2. Three causal convolutional layers with leading zero padding and the corresponding receptive field. In this example, the kernel size is 3 and the dilation based is 2.

several parallel (rather than sequential) convolutional layers with various filter sizes and stacks their output along the convolutional channel dimension (Fig. 1-b), rather than adopting a single convolutional layer with a fixed-length kernel (Fig. 1-a). This approach is based on the idea of the naive inception module, initially proposed in computer vision [16] to tackle the drawbacks of using deeper, wider, and more complex networks for image classification. Inception modules are effective, especially in the case of small training datasets or limited computational capacity, where it is not feasible to use a deep, complex neural network with a high number of parameters. This is because the network gets either more prone to overfitting or goes without the computational resources [16].

2) *Dilation*: Adding dilation to a series of convolutional layers can increase the receptive field of the network. Therefore, we applied this modification of CNN consisting of a series of causal convolutional layers with varying dilation (powers of a base dilation) to take advantage of a more extended history coverage and extract rich temporal features at each time step. By causal convolutional, we mean that the output of the convolutional layer at each time step depends only on the earlier time steps of the input time series (Fig. 2). A leading zero padding has been applied to achieve this and ensure an output sequence with equal length to the input sequence. Using CNN with dilation is the basis of the temporal convolutional network (TCN), which was initially proposed by Bai et al. [28] as a valuable alternative tool to RNNs for sequence modelling. Mathematically, if we have  $n$  convolutional layers stacked on one another with a dilation base of  $b$  and a kernel size of  $T$ , then the TCN receptive field  $r$  can be calculated as follows:

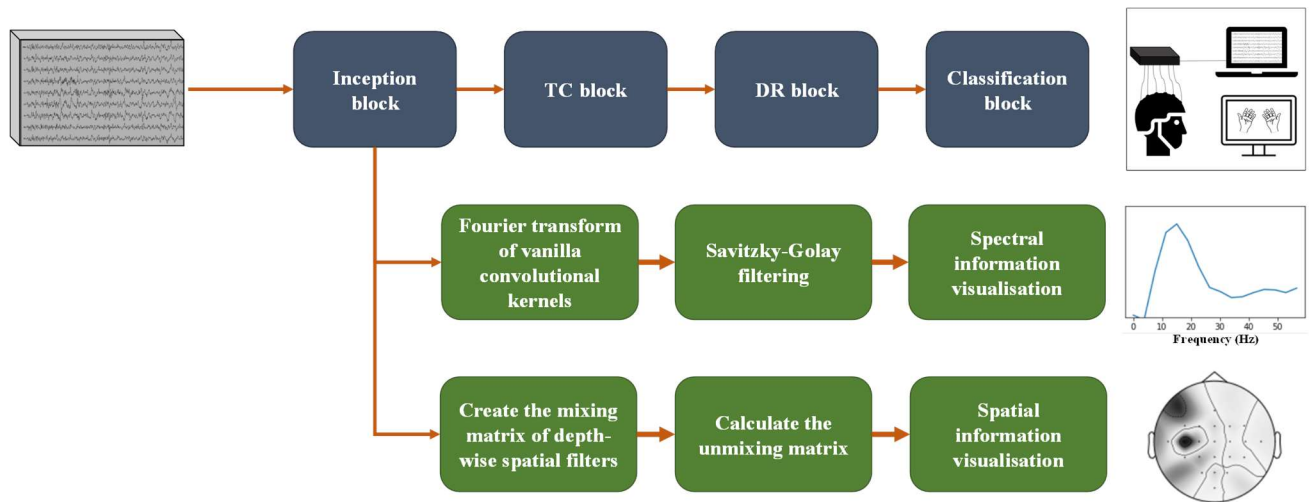


Fig. 3. General schematic of EEG-ITNet

$$r = 1 + (T - 1) \cdot \frac{b^n - 1}{b - 1} \quad (1)$$

Notice that the kernel size  $T$  needs to be selected as a number greater than the dilation base  $b$  to avoid having holes in the receptive field. By holes, we mean samples in the input sequence that do not affect the output sequence. To further enhance the performance of TCN, its representation in the form of residual blocks was proposed [28]. For the extraction of temporal features, we designed a TCN as a series of residual blocks, each consisting of multiple causal convolutional layers with the same dilation. The dilation rate increases for each consecutive residual block as the powers of the dilation base. In this case, with having  $n$  residual blocks with  $m$  convolutional layers in each, a dilation base of  $b$ , and a kernel size of  $T$ , the receptive field of the TCN will be calculated as follows:

$$r = 1 + m(T - 1) \cdot \frac{b^n - 1}{b - 1} \quad (2)$$

So (2) can be used to determine the number of residual blocks and causal convolutions in each, dilation base, and the kernel size needed to design a TCN that can account for  $r$  time steps back in time. Of course, the residual blocks also contain activation functions and dropout to account for nonlinearity and avoid overfitting in the network, respectively. We also applied batch normalisation after convolutional layers to tackle the problem of exploding gradient.

## B. PROPOSED DEEP ARCHITECTURE

The general architecture of EEG-ITNet is depicted in Fig. 3 and consists of 4 main blocks: inception block, temporal convolution (TC) block, dimension reduction (DR) block, and classification block.

1) *Inception block*: The learning process starts with three parallel sets of layers; each includes a 2D convolutional layer along the time axis, which acts as frequency filtering, followed

by a 2D depthwise convolutional layer acting as spatial filtering. Adding inception modules with varying convolutional kernel sizes eliminates the need for a fixed-length kernel [15] and allows the network to learn filters representing various frequency sub-bands. We will later show that the longer the kernel size, the more likely to learn features in low-frequency components. The kernel size for the depthwise convolution equals the number of electrodes in the dataset to design a spatial filter that combines all the electrodes to find the sources. Hence, the tensor obtained after the inception modules represent the signals of sources in different frequency sub-bands. This block ends with a nonlinear activation function and dropout to allow the network to learn more complex nonlinear spatial information and avoid overfitting, respectively.

2) *Temporal convolution (TC) block*: After extracting sources in different informative frequency sub-bands, the TCN architecture is applied to extract the discriminative temporal features while taking the history of the time series into account. The TC block consists of several residual blocks, and each is formed by depthwise causal convolutional layers with leading zero padding, followed by activation function and dropout. Using depthwise causal convolution followed by batch normalisation instead of weight normalisation is a modification that we made in the original TCN structure. Since the output of the inception block represents signals in the source domain, depthwise causal convolution has been used to ensure that the temporal information of each source is extracted separately. With this modification, our observation showed more robust and superior performance than the conventional TCN. Depending on the number of previous time steps intended to be considered, the number of residual blocks, kernel length, and dilation base can be chosen based on (2). An average pooling layer also precedes this block to reduce the data dimensions and avoid overfitting.

3) *Dimension reduction (DR) block*: The output of the TC block essentially contains temporal information extracted



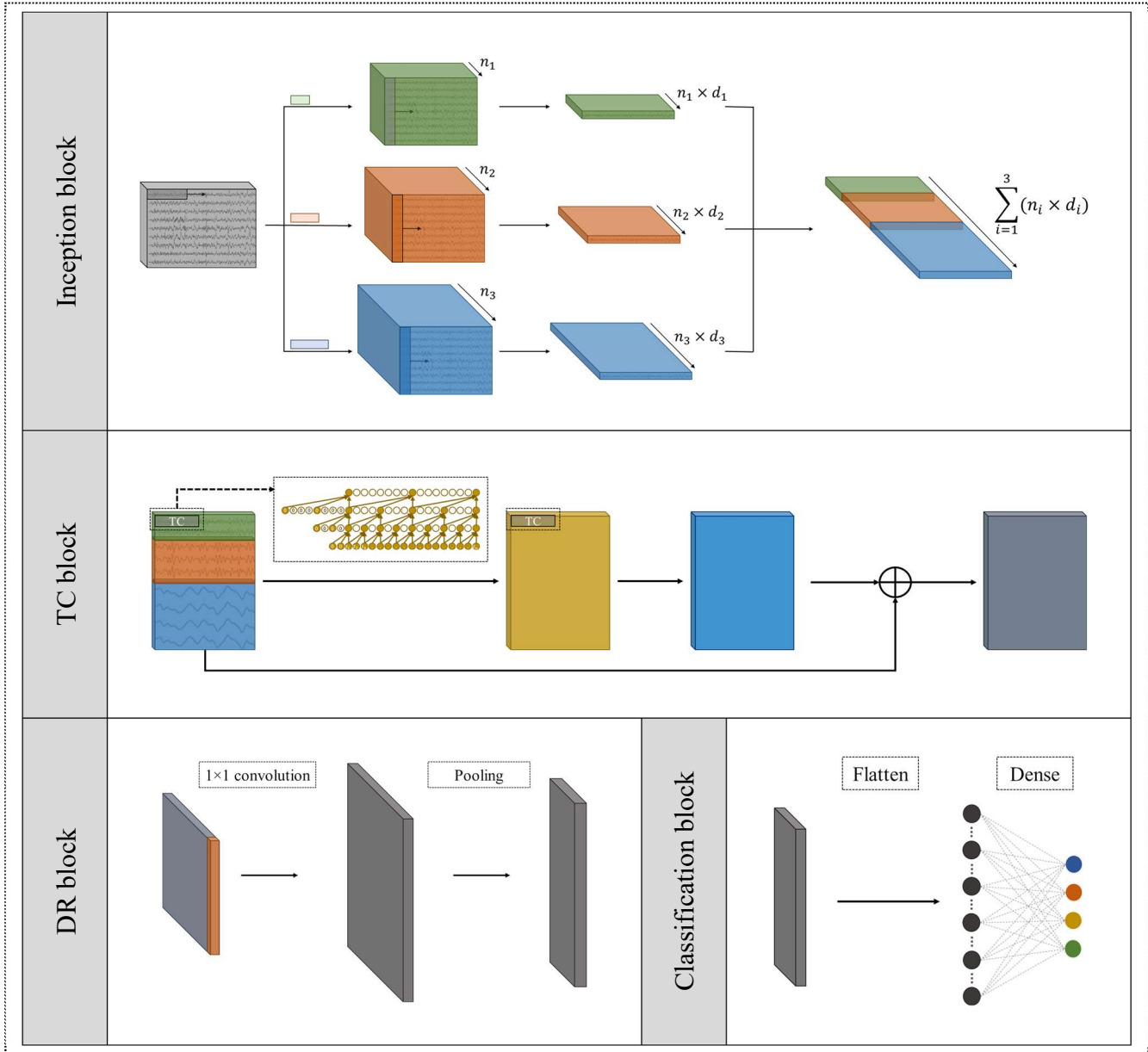


Fig. 4. Details of different blocks in EEG-ITNet architecture

from sources with various frequency spectrums. Thus, we utilised a  $1 \times 1$  convolutional layer to combine these temporal features and control the number of final features used to perform the classification task. The number of convolutional filters is a hyperparameter and needs to be adjusted to avoid overfitting. As well as an activation function and a dropout layer, this block also ends with an average pooling layer to further reduce the tensor dimension.

4) *Classification block*: This block is the final piece of EEG-ITNet and contains a fully connected layer with a "softmax" activation function that follows a flatten layer. Although we call it the classification layer, it can be easily modified depending on the desired output and the problem set.

### C. INTERPRETABILITY AND VISUALISATION OF EEG-ITNet

As depicted in Fig. 4, the inception block consists of 2D convolutional layers applied over the time axis of multi-channel EEG with varying kernel sizes. Adding an extra dimension to the input signal preserved for convolutional channels, these initial 2D convolutional layers act as frequency filters and extract the signal in different informative sub-bands. To demonstrate this, assume the input signal in each electrode  $X[n] \in R^{1 \times s}$ , where  $s$  is the number of samples. Since a convolutional layer is simply the dot product of the input signal and a kernel, the output  $X'[n] \in R^{1 \times s}$  of the initial vanilla 2D convolutional layer

with a kernel  $\mathbf{K}$  of size  $2l + 1$  and "same" padding over each EEG channel can be calculated as follows:

$$\mathbf{X}'[n] = \sum_{i=-l}^l \mathbf{X}[i] \mathbf{K}[i - n] \quad (3)$$

which represents the convolution between the EEG time series of each electrode and the reverse of the kernel ( $\mathbf{X}[n] * \mathbf{K}[-n]$ ). We also know that the convolution in the time domain acts as multiplication in the frequency domain. Therefore, taking the Fourier transform of convolutional kernels, we can find out the frequency sub-bands selected by the network during the training phase. However, since the deep learning algorithms are sensitive to even tiny discriminative features, the Fourier transform of convolutional kernels can be highly varied. So for the sake of visualisation, we use Savitzky-Golay filtering [29] to smooth the Fourier transforms. Savitzky-Golay is a type of digital filtering consisting of a series of least-square polynomial approximations applied on fixed-length time windows swept over the time series. For a sequence of samples  $\mathbf{Y}[n]$ , the mean-squared approximation error for a time window centred at  $c$  can be calculated as follows:

$$E = \sum_{n=c-l}^{c+l} \left( \sum_{m=0}^p a_m n^m - \mathbf{Y}[n] \right)^2 \quad (4)$$

where  $a_m$ ,  $m = 0, \dots, p$  are polynomial coefficients,  $p$  is the polynomial order, and  $l$  is the half-width of the filter window. To minimise the mean-squared approximation error and find the optimal polynomial coefficients, we should take the derivative of (4) with respect to all the polynomial coefficients and set them equal to 0, which yields a set of  $p + 1$  normal equations [30]. Accordingly, the polynomial coefficients can be found as follows:

$$\mathbf{A} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{Y} \quad (5)$$

where  $\mathbf{A} = [a_0, a_1, \dots, a_m, \dots, a_p]^T$  is the vector of polynomial coefficients. The matrix  $\mathbf{D} = \{d_{n,m}\}$  is called the design matrix with the size of  $2l + 1$  by  $p + 1$  and the elements as follows:

$$d_{n,m} = n^m, \quad c - l \leq n \leq c + l, \quad m = 0, \dots, p \quad (6)$$

The Savitzky-Golay smoothing process can also be explained as a shift-invariant discrete convolution process, which is why it is referred to as Savitzky-Golay filtering (details can be found in [30]).

Besides batch normalisation and activation functions, each inception module in EEG-ITNet also contains a depthwise convolutional layer. The depthwise convolutional layer with "valid" padding acts as spatial filtering. It linearly combines signals in different electrodes to transform the surface domain to the source domain and find discriminative sources. Note that this convolutional layer is applied separately in each convolutional channel, as each

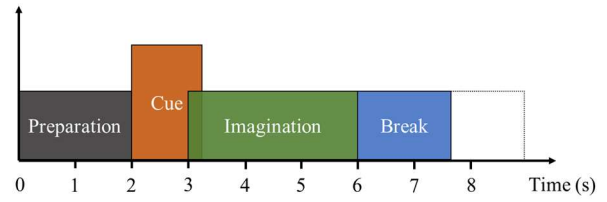


Fig. 5. Timing scheme of the motor imagery dataset 2a from BCI competition IV

convolutional channel represents the signal in a different frequency sub-band. Also, the nonlinear activation function that follows this layer further improves the spatial filtering by giving it nonlinearity. Thus, for an input EEG signal  $\mathbf{X}[n] \in \mathbb{R}^{c \times s}$  with  $c$  channels and  $s$  number of samples in each channel, the transformation from the surface domain to the source domain  $\mathbf{S}[n] \in \mathbb{R}^{c' \times s}$ , with  $c'$  being the number of sources, can be formulated as follows:

$$\mathbf{S}[n] = \mathbf{W} \mathbf{X}[n] \quad (7)$$

where  $\mathbf{W}$  is a  $c' \times c$  matrix containing spatial filters; however,  $\mathbf{W}$  in (7) represents the mixing matrix and does not represent the spatial locations of sources. Therefore, we must find and visualise the unmixing matrix  $\mathbf{W}^{-1}$  to find the correct location of sources learned during the training phase.

## IV. EXPERIMENTS AND RESULTS

### A. DATASET AND EVALUATION SCENARIOS

Dataset 2a of BCI competition IV and OpenBMI motor imagery dataset [31] were used in this study to evaluate the performance of EEG-ITNet exhaustively. The description of each dataset is as follows:

1) *BCI competition IV dataset 2a*: It is a multi-class motor imagery dataset containing EEG recordings of 9 participants during imagination of left hand, right hand, both feet, and tongue movements. The dataset has been collected in two separate sessions, and each consists of 288 trials with an equal number of trials for each class. The actual labels were provided for the data in the first session, while the second session was reserved for testing the classification algorithms in the competition. Each trial in the paradigm started with 2 seconds of preparation time, followed by a cue that lasted for 1.25s, representing the imagination class (Fig. 5). The imagination period continued for 4s after the cue onset and was terminated by an inter-trial break. The EEG signals were recorded using 22 electrodes and sampled with 250Hz. Since we aimed to propose an end-to-end neural network, the signals were only downsampled to 125Hz, scaled to have zero mean and unit variance, and epoched to a 3-seconds time window after the cue onset. It should also be noted that the 3 EOG channels available in the dataset were excluded from our analysis.

2) *OpenBMI motor imagery dataset*: Collected by Lee et al. [31], OpenBMI is a large EEG dataset containing EEG recordings from 54 subjects across multiple sessions

performing three different BCI tasks: motor imagery (MI), event-related potential (ERP), and steady-state visually evoked potential (SSVEP). In this study, we used EEG recordings from the first session of the MI task, which represent neural activities while performing the imagery task of grasping with either left or right hand. The experiment consists of balanced train and test datasets with 100 trials each. The experimental paradigm of this dataset is depicted in Fig. 6. It starts with 3s of preparation period followed by 4s of imagination period based on the movement direction of the fixation cross. 62 Ag/AgCl electrodes were used to record the signals at a 1kHz sampling frequency. Similar to BCI competition IV dataset 2a, the dataset was only downsampled to 125Hz, standardised to have zero mean and unit variance, and epoched to a 3-seconds time window after the cue onset. We should note that similar to what the authors proposed in the original paper [31], only 20 electrodes located on the motor cortex (FC-5/3/1/2/4/6, C5/3/1/z/2/4/6, and CP-5/3/1/z/2/4/6) were selected for the analysis.

The performance of EEG-ITNet was evaluated in three different scenarios, corresponding to within-subject, cross-subject, and cross-subject with fine-tuning, detailed below. Note that, since the data from the second sessions were used for testing the system and the labels were unknown to the participants in the competition, we did not use them as training or validation sets in our analysis.

1) *Within-subject*: For each subject independently, the EEG recordings of their first session were divided into train and validation set to find the best parameters of the network. Then, the trained network was evaluated on the EEG recordings of the second session, and the results were reported for each subject.

2) *Cross-subject*: The network parameters were learned using the training set of all other subjects rather than the test subjects themselves, meaning the system had never seen any data from the test subject. The subjects preserved for training were divided to form our train and validation sets. Similar to the within-subject case, unique network parameters were obtained for each individual subject in this case.

3) *Cross-subject with fine-tuning*: In the final scenario, the labelled EEG recordings of the training session of each target subject were used to fine-tune the parameters of the system developed for each target subject in the second scenario.

## B. COMPARISON MODELS

We have selected three similar existing end-to-end architectures for an exhaustive evaluation of the EEG-ITNet performance, including EEGNet, EEG-TCNet, and EEG-Inception. This section explains each of these networks and their implementation in this study.

1) *EEGNet*: As a compact convolutional neural network, EEGNet showed promising results for analysing different EEG datasets [14]. Like DeepConvNet, EEGNet starts with temporal and spatial convolutions, followed by a separable convolution.

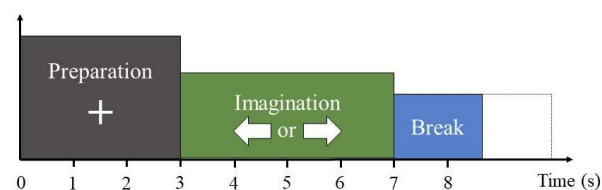


Fig. 6. Timing scheme of the OpenBMI motor imagery dataset

1) *EEG-TCNet*: Proposed as an extension of EEGNet, EEG-TCNet applied a TCN structure after feature extraction layers of EEGNet to further improve its performance [23].

3) *EEG-Inception*: Designed as a classification tool for ERP-based spellers, EEG-Inception is a CNN architecture that comprises two inception modules and an output module [15]. Each inception module consists of three parallel sets of layers, including 2D vanilla convolutions.

Since our dataset and preprocessing stages are similar to Lawhern et al. [14], we used the same implementation of EEGNet. The EEG-Inception was also implemented based on table IV in [15] as they used a dataset with the same sampling frequency as ours (after downsampling). For EEG-TCNet, the authors introduced their network as an extension to the shallow EEGNet. So in order to reproduce EEG-TCNet to be compatible with our preprocessing stages, we used the same implementation of EEGNet provided in [14] followed by a TCN structure. The parameters of TCN were selected based on the first formula in [23] as follows:  $K_T = 3, L = 2, F_T = 16, p_t = 0.3$ .

## C. TRAINING PROCESS AND HYPERPARAMETERS SELECTION

The inception block in EEG-ITNet consists of three 2D convolutional layers with 2, 4, and 8 filters and kernel sizes of 16, 32, and 64 samples, respectively, followed by three depthwise convolutional layers with a depth of 1. Since the input signal's sampling rate is 125Hz, the largest kernel was selected with the size of 64 samples as it can capture frequency components as low as  $125/64 \approx 2\text{Hz}$  of the input signal.

For example, Fig. 10 contains the learned frequency filters with the 16, 32, and 64-sample kernels for subject 3. As it can be seen, the larger the filter size, the more it is prone to capture lower frequencies. The reason behind selecting more filters for larger kernel sizes is the fact that motor imagery is often associated with activities in the lower alpha (8–12Hz) and beta (13–30Hz) bands [32], rather than high frequency bands, e.g. gamma (30–100 Hz) band. Besides, the lack of samples in short filter sizes makes it difficult to calculate their Fourier transform and visualise them. Notwithstanding, gamma activity can still be observed during motor imagery with open eyes due to changes in subjects' spatial attention toward the target limb [33].

The inception block is followed by an average pooling layer with a pool size of 4. Afterwards, to extract temporal features

TABLE II  
PERFORMANCE EVALUATION FOR WITHIN-SUBJECT CASE IN TERMS OF CLASSIFICATION ACCURACY FOR BCI COMPETITION IV DATASET 2A

|                | EEG-Inception      | EEGNet 8,2         | EEG-TCNet          | EEG-ITNet          |
|----------------|--------------------|--------------------|--------------------|--------------------|
| <b>S1</b>      | 77.43 (4th)        | 81.94 (3rd)        | 82.29 (2nd)        | <b>84.38 (1st)</b> |
| <b>S2</b>      | 54.51 (4th)        | 56.94 (3rd)        | <b>64.24 (1st)</b> | 62.85 (2nd)        |
| <b>S3</b>      | 82.99 (4th)        | <b>90.62 (1st)</b> | 88.89 (3rd)        | 89.93 (2nd)        |
| <b>S4</b>      | <b>72.22 (1st)</b> | 67.01 (3rd)        | 60.76 (4th)        | 69.1 (2nd)         |
| <b>S5</b>      | 73.26 (2nd)        | 72.57 (4th)        | 72.92 (3rd)        | <b>74.31 (1st)</b> |
| <b>S6</b>      | <b>64.24 (1st)</b> | 58.68 (3rd)        | 62.5 (2nd)         | 57.64 (4th)        |
| <b>S7</b>      | 82.64 (3rd)        | 76.04 (4th)        | 83.33 (2nd)        | <b>88.54 (1st)</b> |
| <b>S8</b>      | 77.78 (4th)        | 81.25 (2nd)        | 79.51 (3rd)        | <b>83.68 (1st)</b> |
| <b>S9</b>      | 76.39 (3rd)        | 78.12 (2nd)        | 76.39 (3rd)        | <b>80.21 (1st)</b> |
| <b>Average</b> | 73.50              | 73.69              | 74.54              | <b>76.74</b>       |
| <b>Std</b>     | 9.11               | 11.12              | 10.09              | 11.48              |
| <b>p-value</b> | <b>0.043*</b>      | <b>0.010*</b>      | 0.055              | -                  |

\*Significant at level of 0.05

in the TC block, we decided to use four residual blocks, each consisting of two convolutional layers, with a dilation base of 2 capable of covering the whole signal history. For this purpose, the TC block requires convolutional layers with a filter size of 4, which can be calculated based on (2). Finally, the number of filters in the  $1 \times 1$  convolutional layer used as a dimension reduction technique was selected to be 14, followed by another average pooling with a pool size of 4. The dropout rate of 0.4 and 0.2 were also picked throughout the network for within-subject and cross-subject scenarios, respectively.

In all scenarios, we first used 10-fold cross-validation with 500 (150 in cross-subject) epochs in each to select the best model parameters, with the help of early stopping with the patience of 100 (15 in cross-subject) to avoid overfitting. Then we trained the networks for a maximum of 50 extra epochs with the learning rate of  $10^{-4}$  on the combined training and validation sets to benefit from the information of all the labelled samples. The number of extra epochs needs to be carefully investigated to ensure no overfitting. For this purpose, we extracted the curves representing train and test losses versus a range of extra epochs used in the training process, concatenated to the loss patterns in the selected best fold of cross-validation. For instance, Fig. 7 shows this curve for subject 3 in BCI competition IV dataset 2a while training EEG-ITNet in the cross-subject scenario. Notice that the sudden jump in the curves is due to the fact that in each fold, we split all labelled data to form train and validation sets, while during extra epochs, we use all labelled data as the train

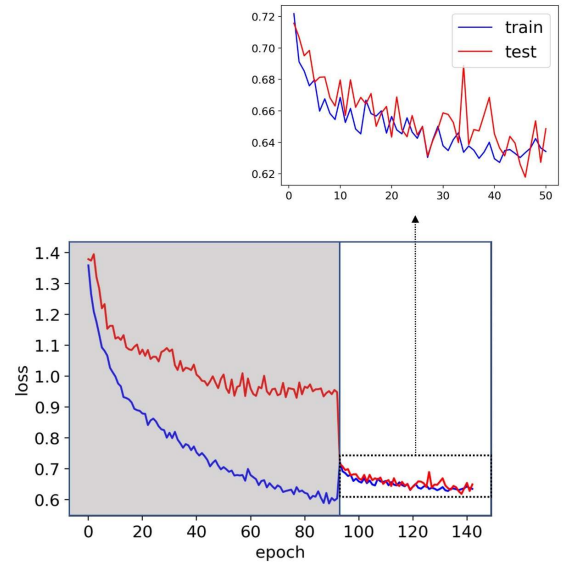


Fig. 7. The effect of the number of extra epochs in the training of EEG-ITNet for subject 3 from BCI competition IV dataset 2a in the cross-subject scenario. Left (grey box) - training process in the selected best fold of cross-validation. Right (white box) - training process for 50 extra epochs using all labelled data

set, resulting in a different range of values for our losses. It should be reminded that in the competition, the participants did not have access to the test data labels to evaluate the performance of their proposed algorithm. However, in this paper, for the sake of fair comparison, we selected the best test



TABLE III  
PERFORMANCE EVALUATION FOR CROSS-SUBJECT CASE  
IN TERMS OF CLASSIFICATION ACCURACY FOR BCI  
COMPETITION IV DATASET 2A

|         | EEG-<br>Inception | EEGNet<br>8,2 | EEG-<br>TCNet | EEG-<br>ITNet |
|---------|-------------------|---------------|---------------|---------------|
| S1      | 66.32             | 68.75         | 69.1          | <b>71.88</b>  |
| S2      | 48.26             | 50            | 52.08         | <b>62.85</b>  |
| S3      | 73.61             | 80.21         | <b>81.94</b>  | <b>81.94</b>  |
| S4      | 56.6              | 59.38         | 61.81         | <b>65.62</b>  |
| S5      | <b>65.62</b>      | 64.24         | 60.42         | 63.19         |
| S6      | <b>56.25</b>      | 48.26         | 51.39         | <b>56.25</b>  |
| S7      | 73.61             | 72.57         | 76.39         | <b>80.21</b>  |
| S8      | 70.49             | 77.43         | 74.31         | <b>78.12</b>  |
| S9      | 61.11             | 55.56         | 58.68         | <b>64.93</b>  |
| Average | 63.54             | 64.04         | 65.12         | <b>69.44</b>  |
| Std     | 8.69              | 11.59         | 10.86         | 8.98          |
| p-value | <b>0.009*</b>     | <b>0.008*</b> | <b>0.006*</b> | -             |

\*Significant at level of 0.05

TABLE IV  
PERFORMANCE EVALUATION FOR CROSS-SUBJECT WITH FINE-  
TUNING CASE IN TERMS OF CLASSIFICATION ACCURACY FOR  
BCI COMPETITION IV DATASET 2A

|         | EEG-<br>Inception | EEGNet<br>8,2 | EEG-<br>TCNet | EEG-<br>ITNet |
|---------|-------------------|---------------|---------------|---------------|
| S1      | 77.43             | 84.38         | <b>86.46</b>  | 84.03         |
| S2      | 54.86             | 54.86         | 64.93         | <b>65.28</b>  |
| S3      | 87.85             | <b>92.36</b>  | 90.28         | 92.01         |
| S4      | 72.57             | 67.01         | 71.53         | <b>73.96</b>  |
| S5      | 74.65             | 66.67         | 73.26         | <b>75.35</b>  |
| S6      | <b>66.32</b>      | 61.46         | 58.68         | 64.93         |
| S7      | 79.17             | 79.86         | 80.56         | <b>84.72</b>  |
| S8      | 83.33             | 82.99         | 82.99         | <b>84.72</b>  |
| S9      | 79.17             | 75.69         | 73.61         | <b>83.68</b>  |
| Average | 75.04             | 73.92         | 75.81         | <b>78.74</b>  |
| Std     | 9.76              | 12.19         | 10.24         | 9.4           |
| p-value | <b>0.008*</b>     | <b>0.010*</b> | <b>0.022*</b> | -             |

\*Significant at level of 0.05

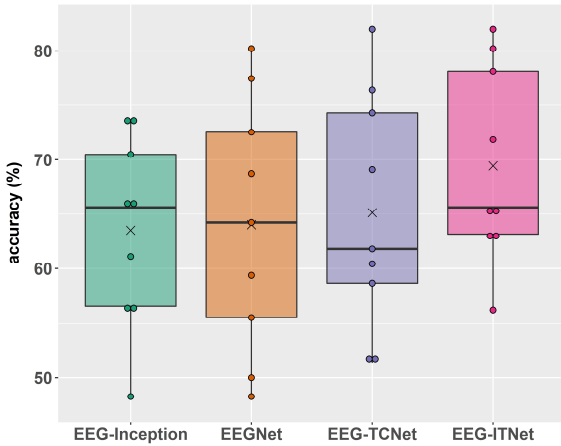


Fig. 8. Box plot of the classification performance on BCI competition IV dataset 2a in the cross-subject scenario

accuracy for each deep learning model and each subject, meaning the number of extra epochs varies over models and subjects.

#### D. PERFORMANCE EVALUATION

Table II summarises the classification accuracies in the within-subject case for EEG-ITNet and our comparison models on BCI competition IV dataset 2a. The performance of EEG-ITNet has been superior to other classification algorithms in terms of mean classification accuracy over all the subjects. Nevertheless, a one-sided Wilcoxon signed-rank test with a significance level of 0.05 was used to verify the significance of improvement [34]. The Wilcoxon signed-rank test is a non-

parametric statistical test, especially suitable for small sample sizes. Non-parametric statistical tests generally do not require large sample sizes and make fewer assumptions about the data distribution, including normality. The performance of EEG-ITNet is very promising, reaching the highest accuracy for five out of nine subjects and the second best in three cases.

Table III shows the performance of EEG-ITNet in the cross-subject scenario and its comparison with other end-to-end architectures on BCI competition IV dataset 2a. It also contains the p-value needed to reject the null hypothesis (no difference in the classification accuracies obtained by EEG-ITNet and each comparison model). In this scenario, each model is trained without using any data from the target subject. Therefore, due to the dynamic nature of EEG signals and their variation from one subject to another, significantly lower performance have been observed in this case. However, our proposed model has managed to reach the highest mean accuracy in the cross-subject case, with statistically significant improvement over EEG-Inception, EEGNet, and EEG-TCNet. Fig. 8 also shows this comparison and the data distribution of the results reported in Table III using a box plot.

Finally, for each subject, their EEG recordings from the training session were used to fine-tune the parameters of the systems designed for them in the cross-subject scenario. The classification results of this case can be found in Table IV, which confirm the superior performance of EEG-ITNet compared to other deep learning models with statistical significance. We have also investigated the improvement in the classification results compared to the within-subject

TABLE V  
SUMMARY OF CLASSIFICATION PERFORMANCE IN DIFFERENT SCENARIOS FOR OPENBMI MOTOR IMAGERY DATASET

|               | Within-subject |       |                | Cross-subject |       |                | Cross-subject with fine tuning |       |                |
|---------------|----------------|-------|----------------|---------------|-------|----------------|--------------------------------|-------|----------------|
|               | acc            | t (n) | <i>p-value</i> | acc           | t (n) | <i>p-value</i> | acc                            | t (n) | <i>p-value</i> |
| EEG-Inception | 69.3           | 3.88  | <0.001*        | 71.15         | 3.71  | <0.001*        | 75.11                          | 1.52  | 0.067          |
| EEGNet 8,2    | 69.61          | 4.23  | <0.001*        | 71.2          | 3.35  | <0.001*        | 74.04                          | 3.35  | <0.001*        |
| EEG-TCNet     | 68.35          | 5.23  | <0.001*        | 70.83         | 4.55  | <0.001*        | 73.85                          | 4.32  | <0.001*        |
| EEG-ITNet     | <b>71.91</b>   | -     | -              | <b>73.52</b>  | -     | -              | <b>76.19</b>                   | -     | -              |

acc: Classification accuracy, t: Test statistic T, n: Degree of freedom  
\*Significant at level of 0.05

scenario in Fig. 9. Comparing the improvement of shallow EEGNet (5 out of 9 subjects with an average improvement of 0.23%) with deep EEG-ITNet (7 out of 9 subjects with an average improvement of 2%) proves the ability of the proposed deeper architecture to handle extra information from other subjects more effective than a shallow one. The improvement is also more significant for complex networks (in terms of the number of parameters) such as EEG-Inception, which showed improvement in 7 out of 9 subjects with an average improvement of 1.54%. However, deep EEG-ITNet is more favoured than the deep and complex EEG-Inception in this area, as it has more potential to be visualised and interpreted.

Because of the high number of subjects in the OpenBMI motor imagery dataset, results have been summarised in Table V. The table shows superior performance for EEG-ITNet in all of our cases compared to other deep models in terms of mean classification accuracy. In order to evaluate the significance of improvement, we performed right-tailed paired t-test after checking the test assumptions. We confirmed data normality by Shapiro–Wilk test.

### E. VISUALISATION

Fig. 10 shows the spectral and spatial filters learned for subject 3 from BCI competition IV dataset 2a in the inception block of EEG-ITNet using the visualisation techniques proposed in section III. This has been selected as an example to show how each set of frequency response and scalp topo-map represents the power spectrum and spatial location of a group of sources activating uniquely among motor imagery classes. Notice that the sign of spatial coefficients does not carry any information about oscillatory synchronisation or desynchronisation processes. The network is trained to perform a classification task by finding discriminative neural patterns among the four motor imagery classes. The spatial activations' location and the absolute value of coefficients are important only as they represent the occurrence and significance of distinct neural

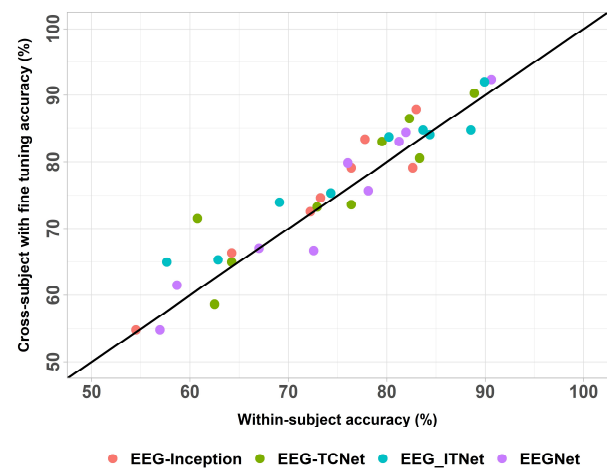


Fig. 9. Classification improvement in cross-subject with fine-tuning case compared to the within-subject scenario for different deep learning models on BCI competition IV dataset 2a

activities in at least one of the motor imagery classes. So in order to avoid confusion, the spatial patterns are plotted in grayscale symmetric to zero.

As shown in Fig. 10, larger filter sizes capture relatively lower frequencies while smaller filters tend to learn higher frequency components. To explain the learned filters, we should consider the dataset to be composed of two tasks, including motor imagery and sustained spatial attention toward the target limb. Motor imagery is associated with a decrease in alpha and beta power (event-related desynchronisation) over the sensory and motor cortex in the contralateral hemisphere [32, 35, 36]. This modulated alpha activity has been clearly captured by filters 4, 5, 7, and 8 in the left panel of Fig. 10 and filter 4 in the middle panel. Notice that among these filters, spatial activation is located at the vertex only in filter 5 of the left panel, which is likely to be associated with the imagination of bilateral foot movement. In other cases, the filters represent modulated neural patterns

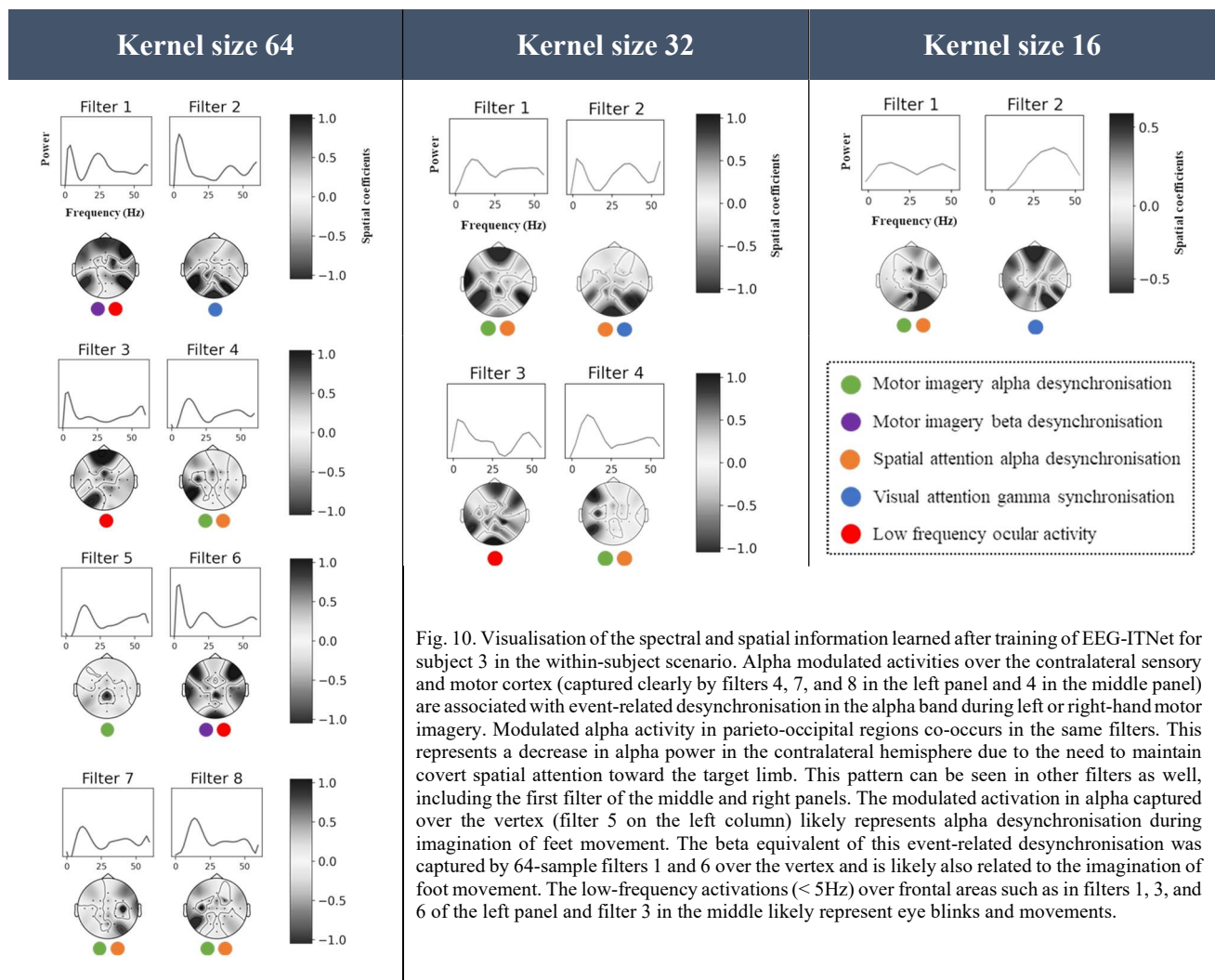


Fig. 10. Visualisation of the spectral and spatial information learned after training of EEG-ITNet for subject 3 in the within-subject scenario. Alpha modulated activities over the contralateral sensory and motor cortex (captured clearly by filters 4, 7, and 8 in the left panel and 4 in the middle panel) are associated with event-related desynchronisation in the alpha band during left or right-hand motor imagery. Modulated alpha activity in parieto-occipital regions co-occurs in the same filters. This represents a decrease in alpha power in the contralateral hemisphere due to the need to maintain covert spatial attention toward the target limb. This pattern can be seen in other filters as well, including the first filter of the middle and right panels. The modulated activation in alpha captured over the vertex (filter 5 on the left column) likely represents alpha desynchronisation during imagination of feet movement. The beta equivalent of this event-related desynchronisation was captured by 64-sample filters 1 and 6 over the vertex and is likely also related to the imagination of foot movement. The low-frequency activations ( $< 5\text{Hz}$ ) over frontal areas such as in filters 1, 3, and 6 of the left panel and filter 3 in the middle likely represent eye blinks and movements.

during left or right-hand motor imagery located in the contralateral hemisphere. The beta equivalent to the alpha desynchronisation seen over the vertex (filter 5 of the 64-sample kernel) during motor imagery was also captured by 64-sample filters 1 and 6, which may be similarly related to bilateral foot movement imagination.

The spatial activation in the parieto-occipital regions in the alpha band witnessed by filters 4, 7, and 8 of the left panel may be associated with the need to maintain spatial attention toward the target limb. Studies have shown that covert spatial attention can lead to alpha desynchronisation in the contralateral hemisphere alongside an increase in alpha power ipsilaterally over the visual cortex [37-39]. So, for instance, during right-hand motor imagery, when the subjects direct their visual attention to the right visual hemifield and start imagining movement, alpha desynchronisation happens contralaterally in motor and visual cortical regions that are associated with motor imagery and visual attention, respectively. Increased gamma activity has also been reported during visual attention contralateral to the visual hemifield [33], which would explain the increase in the power spectrum

between 30-40Hz in most of the filters, yet more detectable in the second filter of each inception module. Gamma-band activity can also be explained considering that the time window used for the analysis contains the period in which subjects attended to the visual cue. Please also note that visualisations  $>60\text{Hz}$  are not valid based on the Nyquist theorem as we have downsampled our signals to 125Hz.

Finally, spatial activations in low frequencies ( $< 5\text{Hz}$ ) over frontal regions most likely represent eye blinks and eye movements. This pattern can be seen in filters 1, 3, and 6 of the left panel and filter 3 in the middle. While eye blinks are unlikely to play a functional role, there might be a relationship between subjects' eye movement and how they attend to and perform each motor imagery task in the experiment, which is captured as a discriminative pattern by the neural network.

## V. DISCUSSION

This study proposed EEG-ITNet, an explainable CNN architecture based on inception modules and causal convolutions with dilation. Inception modules were proposed to eliminate the need to use fix-length kernels, allowing the

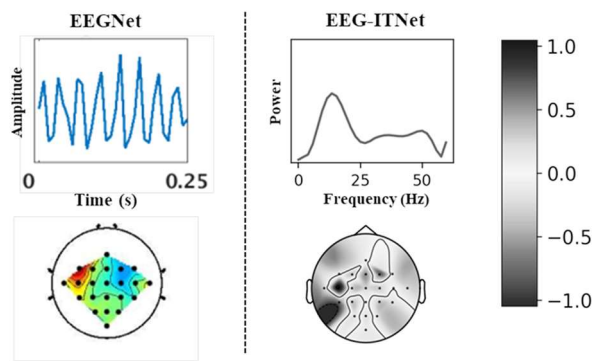


Fig. 11. Comparison of the outcome of EEGNet (left) and EEG-ITNet (right) feature visualisation techniques. In EEGNet, frequency information can be calculated based on the number of detectable cycles in the learned temporal kernel window, resulting in a single frequency component. So, for example, the above kernel from EEGNet represents 32Hz as it contains 8 cycles in 0.25 seconds.

network to learn input data patterns at different scales. In addition, causal convolutions with dilation were employed as an alternative to RNNs to learn and extract temporal dependencies and information in EEG time series. The series of depthwise causal convolutions with dilation in the form of residual blocks extracted informative discriminative features to perform the classification task in such a way that allowed EEG-ITNet to outperform well-known existing end-to-end neural networks in different scenarios.

Another significance of EEG-ITNet is its less complex structure (in terms of the number of trainable parameters) compared to other existing end-to-end architectures, including EEG-Inception and EEG-TCNet. Together with the proposed interpretable feature visualisation technique, it makes EEG-ITNet an ideal candidate for EEG analysis. Among our end-to-end comparison models, only the authors of EEGNet have attempted to explain and visualise the learned features. However, in our study, the Fourier transform combined with Savitzky-Golay filtering offered more interpretability and accuracy for network visualisation (Fig. 11). For instance, the weights of the convolutional layers applied along the time axis in EEGNet were translated as representing a single frequency component each. However, it is almost unlikely for the neural patterns of motor movements or mental tasks to be generated from a single frequency component. That is why the reader may find the smoothed power spectrum of sources generated in EEG-ITNet more intuitive and easier to interpret. Besides, the authors of EEGNet have made no attempt to explain the spatial filters from a neuroscientific viewpoint. Hence, we discussed the validity of the extracted features and supported them from a neuroscientific perspective. Our discussion was based on a fair assumption that motor imagery in our selected experiment relies on lateralised changes in the spatial attention toward the target limb and the imagination of motor execution.

## VI. CONCLUSION

We thoroughly evaluated our proposed network on two motor imagery BCI datasets in three different scenarios: within-

subject, cross-subject, and cross-subject with fine-tuning. As a result, EEG-ITNet showed statistically significant improvement in performing the classification task in most of the scenarios compared to other end-to-end architectures, including EEG-Inception, EEGNet, and EEG-TCNet. Furthermore, we introduced a visualisation technique based on Fourier transform and Savitzky-Golay filtering to explain the network learning criteria. Although initially showed promising results, the performance of EEG-ITNet may be improved even further by finding the optimum hyperparameters for each scenario. In addition, finding a way to visualise and explain temporal features and dependencies learned inside our TC block would be of great importance. This can further improve the interpretability of EEG-ITNet and help the neuroscientific community with implementing deep neural networks such as ours in their research. Besides, we are interested in further investigating the effect of network deepness on its performance. These will be our focus in future studies. Nevertheless, we believe that the concepts introduced in this study can help to develop more robust, interpretable, and high-accuracy BCI systems.

## REFERENCES

- [1] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain-computer interfaces for communication and rehabilitation," *Nature Reviews Neurology*, vol. 12, no. 9, pp. 513-525, 2016.
- [2] S. N. Abdulkader, A. Atia, and M.-S. M. Mostafa, "Brain computer interfacing: Applications and challenges," *Egyptian Informatics Journal*, vol. 16, no. 2, pp. 213-230, 2015.
- [3] S. P. van den Broek, F. Reinders, M. Donderwinkel, and M. Peters, "Volume conduction effects in EEG and MEG," *Electroencephalography and clinical neurophysiology*, vol. 106, no. 6, pp. 522-534, 1998.
- [4] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *Journal of neural engineering*, vol. 15, no. 3, p. 031005, 2018.
- [5] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *sensors*, vol. 12, no. 2, pp. 1211-1279, 2012.
- [6] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: a review," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1753-1762, 2016.
- [7] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of neural engineering*, vol. 16, no. 3, p. 031001, 2019.
- [8] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of neural engineering*, vol. 16, no. 5, p. 051001, 2019.
- [9] A. Al-Saegh, S. A. Dawwd, and J. M. Abdul-Jabbar, "Deep learning for motor imagery EEG-based classification: A review," *Biomedical Signal Processing and Control*, vol. 63, p. 102172, 2021.
- [10] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 11, pp. 5619-5629, 2018.
- [11] Z. Wang, L. Cao, Z. Zhang, X. Gong, Y. Sun, and H. Wang, "Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 23, p. e4413, 2018.



- [12] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 156-165.
- [13] R. T. Schirmermeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391-5420, 2017.
- [14] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.
- [15] E. Santamaría-Vázquez, V. Martínez-Cagigal, F. Vaquerizo-Villar, and R. Hornero, "EEG-Inception: A Novel Deep Convolutional Neural Network for Assistive ERP-Based Brain-Computer Interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 12, pp. 2773-2782, 2020.
- [16] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
- [17] M. Riyad, M. Khalil, and A. Adib, "MI-EEGNET: A novel convolutional neural network for motor imagery classification," *Journal of Neuroscience Methods*, vol. 353, p. 109037, 2021.
- [18] J. Liu, F. Ye, and H. Xiong, "Multi-class motor imagery EEG classification method with high accuracy and low individual differences based on hybrid neural network," *Journal of Neural Engineering*, vol. 18, no. 4, p. 0460f1, 2021.
- [19] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "LSTM-based EEG classification in motor imagery tasks," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 26, no. 11, pp. 2086-2095, 2018.
- [20] T.-j. Luo, C.-l. Zhou, and F. Chao, "Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network," *BMC bioinformatics*, vol. 19, no. 1, pp. 1-18, 2018.
- [21] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial-temporal recurrent neural network for emotion recognition," *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 839-847, 2018.
- [22] A. Sundaresan, B. Penchina, S. Cheong, V. Grace, A. Valero-Cabré, and A. Martel, "Evaluating deep learning EEG-based mental stress classification in adolescents with autism for breathing entrainment BCI," *Brain Informatics*, vol. 8, no. 1, pp. 1-12, 2021.
- [23] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, "EEG-TCNet: An Accurate Temporal Convolutional Network for Embedded Motor-Imagery Brain-Machine Interfaces," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020: IEEE, pp. 2958-2965.
- [24] Y. K. Musallam *et al.*, "Electroencephalography-based motor imagery classification using temporal convolutional network fusion," *Biomedical Signal Processing and Control*, vol. 69, p. 102826, 2021.
- [25] P. Zhang, X. Wang, J. Chen, W. You, and W. Zhang, "Spectral and temporal feature learning with two-stream neural networks for mental workload assessment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 6, pp. 1149-1159, 2019.
- [26] L. A. Gemein *et al.*, "Machine-learning-based diagnostics of EEG pathology," *NeuroImage*, vol. 220, p. 117021, 2020.
- [27] S. Hashempour, R. Boostani, M. Mohammadi, and S. Sanei, "Continuous Scoring of Depression from EEG Signals via a Hybrid of Convolutional Neural Networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022.
- [28] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.
- [29] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical chemistry*, vol. 36, no. 8, pp. 1627-1639, 1964.
- [30] R. W. Schafer, "What is a Savitzky-Golay filter?[lecture notes]," *IEEE Signal processing magazine*, vol. 28, no. 4, pp. 111-117, 2011.
- [31] M.-H. Lee *et al.*, "EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, p. giz002, 2019.
- [32] E. Formaggio, S. F. Storti, R. Cerini, A. Fiaschi, and P. Manganotti, "Brain oscillatory activity during motor imagery in EEG-fMRI coregistration," *Magnetic Resonance Imaging*, vol. 28, no. 10, pp. 1403-1412, 2010.
- [33] M. M. Müller, T. Gruber, and A. Keil, "Modulation of induced gamma band activity in the human EEG by attention and visual information processing," *International Journal of Psychophysiology*, vol. 38, no. 3, pp. 283-299, 2000.
- [34] R. Woolson, "Wilcoxon signed-rank test," *Wiley encyclopedia of clinical trials*, pp. 1-3, 2007.
- [35] Y. Y. Chen, K. J. Lambert, C. R. Madan, and A. Singhal, "Mu oscillations and motor imagery performance: A reflection of intra-individual success, not inter-individual ability," *Human Movement Science*, vol. 78, p. 102819, 2021.
- [36] X. Zhang, Y. Guo, B. Gao, and J. Long, "Alpha frequency intervention by electrical stimulation to improve performance in mu-based bci," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 6, pp. 1262-1270, 2020.
- [37] A. Bollimunta, J. Mo, C. E. Schroeder, and M. Ding, "Neuronal mechanisms and attentional modulation of corticothalamic alpha oscillations," *Journal of Neuroscience*, vol. 31, no. 13, pp. 4935-4943, 2011.
- [38] Y. Bagherzadeh, D. Baldauf, D. Pantazis, and R. Desimone, "Alpha synchrony and the neurofeedback control of spatial attention," *Neuron*, vol. 105, no. 3, pp. 577-587. e5, 2020.
- [39] S. M. Doesburg, J. J. Green, J. J. McDonald, and L. M. Ward, "From local inhibition to long-range integration: a functional dissociation of alpha-band synchronization across cortical scales in visuospatial attention," *Brain research*, vol. 1303, pp. 97-110, 2009.



**Abbas Salami** received the B.Sc. degree in Electrical Engineering from Ferdowsi University Mashhad, Iran, in 2015, and the M.Sc. degree in Biomedical Engineering from Amirkabir University of Technology, Iran, in 2018. He is currently pursuing a PhD degree in Computer Science at the University of Essex, UK. His current research interests include brain-computer interfaces, biological signals processing, and deep learning.



**Dr. Andreu-Perez** (PhD '12, SMIEEE' 19) is Senior Lecturer and Chair of the Smart Health Technologies Group within the Centre for Computational Intelligence, University of Essex (UK). His main expertise is in application-driven methodological research in Artificial Intelligence for neuro/bio-engineering. Javier's methods interest is in deep learning, fuzzy systems, and human-centred artificial intelligence, particularly developing novel methods focusing on accountability and explainability. Javier has published more than 50 articles, including several prestigious IEEE Transactions and other Top journals in artificial intelligence and neuroscience. In total, Javier's work in Artificial Intelligence and Biomedical engineering has attracted more than 2500 citations. Javier also contributes as associate Editor-in-Chief of the Journal Neurocomputing (Elsevier) and other prestigious journals in computational intelligence and emerging technologies. Javier has also received remunerated research awards and has been the primary investigator of projects funded by the United Kingdom research councils, international organisations, and corporations.



**Helge Gillmeister** received a BSc in Psychology with Cognitive Science from University College London (UK) in 2000 and a PhD in Psychology from Birkbeck, University of London in 2005. She is currently a Reader at the Department of Psychology and Centre for Brain Science at the University of Essex (UK). She worked as a Research Assistant from 2000 to 2001 and again from 2005 to 2009 at the MRC Cognition and Brain Sciences Unit in Cambridge (UK), Birkbeck, UCL and City University London. Her research in cognitive neuroscience

is centred around the bodily self in health and disease. Dr Gillmeister is a member of the British Association for Cognitive Neuroscience, the Experimental Psychology Society, the European Brain and Behaviour Society and the European Society for Cognitive and Affective Neuroscience.