Journal of Neuroscience Methods 345 (2020) 108886

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

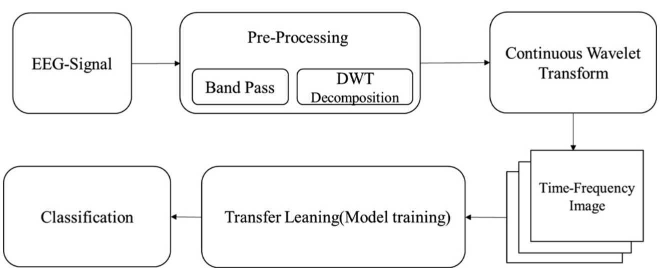
journal homepage: www.elsevier.com/locate/jneumeth

CWT Based Transfer Learning for Motor Imagery Classification for Brain computer Interfaces

Piyush Kant*, Shahedul Haque Laskar, Jupitara Hazarika, Rupesh Mahamune

Department of Electronics and Instrumentation Engineering, National Institute of Technology, Silchar 788010, Assam, India

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
EEG signal processing
cwt filter-bank
deep learning
short-time Fourier transform
convolutional neural network
Transfer Learning

ABSTRACT

Background: The processing of brain signals for Motor imagery (MI) classification to have better accuracy is a key issue in the Brain-Computer Interface (BCI). While conventional methods like Artificial neural network (ANN), Linear discernment analysis (LDA), K-Nearest Neighbor (KNN), Support vector machine (SVM), etc. have made significant progress in terms of classification accuracy, deep transfer learning-based systems have shown the potential to outperform them. BCI can play a vital role in enabling communication with the external world for persons with motor disabilities.
New Methods: Deep learning has been a success in many fields. However, for Electroencephalogram (EEG) signals, relatively minimal work has been carried out using deep learning. This paper proposes a combination of Continuous Wavelet Transform (CWT) along with deep learning-based transfer learning to solve the problem. CWT transforms one dimensional EEG signals into two-dimensional time-frequency-amplitude representation enabling us to exploit available deep networks through transfer learning.
Results: The effectiveness of the proposed approach is evaluated in this study using an openly available BCI competition data-set. The results of the approach have been compared to earlier works on the same dataset, and a promising validation accuracy of 97.06% is achieved in our investigation.
Comparison with existing methods and Conclusion: Our approach has shown significant improvement over other studies, which is 5.71% improvement over earlier reported algorithm (Tabar and Halici, 2017) using the same dataset. Results show the validity of the proposed Deep Transfer-Learning based technique as a state of the art technique for MI classification in BCI.

* Corresponding author.
E-mail address: piyushtwri@gmail.com (P. Kant).

https://doi.org/10.1016/j.jneumeth.2020.108886
Received 25 February 2020; Received in revised form 25 July 2020; Accepted 25 July 2020
Available online 28 July 2020

1. Introduction

Brain-computer interface (BCI), also referred to as Human-Machine Interface (HMI), incorporates recording and decoding brain signals to control an external device (Hsu and Sun, 2009). The communication between the human brain and the mechanical equipment is established here through the software technology incorporated in between. One of the common goals of developing the BCI technique is to help physically challenged but mentally sound people to sustain on their own by controlling a machine with the brain signals (Khasnobish and Bhattacharyya, 2010). It focusses on developing neuro-prosthetics that aims to restore movement, hearing, and sight of impaired individuals. Although multiple BCI studies have shown its successful applications in transferring the neural command to control external tools, e.g., computer cursors, prosthetic limbs, robotic arms, or wheelchairs, still there is scope to improve the classification accuracy for better implementation.

The human brain produces electrical signals which can be detected employing electroencephalography (EEG). It is a non-invasive technique that collects the brain's oscillatory activation pattern from the scalp. Thus, it is very much suitable and reliable method to receive the control command for BCI (Hsu and Sun, 2009; Khasnobish and Bhattacharyya, 2010). Studies using EEG signals while imagining finger or limb movement, commonly known as motor imagery (MI), to operate an artificial intelligence system have been observed in the literature (Hsu and Sun, 2009). The goal of such BCI studies is to recognize the MI-task induced EEG pattern. The non-stationary EEG signal has high-temporal resolution (Khasnobish and Bhattacharyya, 2010), and the amplitude ranges from 0.5 to 100μV. The frequency range of an EEG signal generally lies between 0.5 to 100 Hz, and it is sub-divided into various sub-frequency bands. For MI-based BCI systems, the activity in Mu (8-13 Hz) and Beta (13-30 Hz) frequency bands recorded from the sensorimotor cortex are found to be efficient in detecting event-related synchronization (ERS) (Klimesch et al., 1996) and de-synchronization (ERD) pattern (Göksu, 2018). Mu band is a feature characteristic of EEG signal in motor events as EEG patterns in the brain rhythm decreases (desynchronizes) during motor activities. Mu rhythm got its name from suppressed EEG patterns in the sensory-motor cortex during MI. These patterns are generated due to ERD in brain potentials, which resembles the Greek letter Mu when inverted in polarity. Some studies have shown that Beta band along with mu rhythms is related closely to event-related potential (Kim et al., 2016; McFarland et al., 2000). It also has been pointed out that many times frequency bands are subject-specific and can slightly vary subject to subject (Zhang et al., 2015).

Information extraction from the recorded EEG signals to recognize the specific activation pattern plays a vital role in BCI based research. Although the concept of BCI was proposed decades ago, the reliable translation of the user's intention to the control device is still a significant challenge. A successful BCI system has two basic requirements, which include an effective set of EEG features that should be able to differentiate task-induced brain activity and an efficient machine learning tool for classifying the extracted features. Since both time and frequency information of EEG is essential, a multiresolution based wavelet transform is found to be more suitable in EEG analysis compared to the frequency-based Fourier domain (Hsu and Sun, 2009). Multiple studies have incorporated wavelet transform based feature extraction techniques in BCI applications (Göksu, 2018; Bhattacharyya et al., 2010). Another popular feature extraction method used in MI-applications involves common spatial patterns (CSP) (Tabar and Halici, 2017; Zhang et al., 2015; Christensen et al., 2019). CSP constructs additive sub-windows of the signal with a maximum difference invariance. These extracted features represent the statistical characteristics of the time domain signal and reduce the computational complexity.

The application of the classification techniques on these statistical features helps in determining the BCI-based EEG activation pattern. As such, different machine learning tools have been evaluated to obtain the best optimum solution for MI task recognition. Significant improvement classification accuracy has been observed with traditional classification algorithms such as artificial neural network (ANN) (Zhou et al., 2008; Lee et al., 2017), Bayesian classifier (Kang and Choi, 2014), K-nearest neighbor (KNN) (Bhattacharyya et al., 2010), quadratic discriminant analysis (QDA) (Bhattacharyya et al., 2010), linear discriminant analysis (LDA) (Bhattacharyya et al., 2010; Zhou et al., 2008; Murugappan et al., 2010) and support vector machine (SVM) (Zhou et al., 2008; Liu et al., 2010).

Traditional methods such as ANN can efficiently train the network but are restricted to certain limitations such as a limited number of hidden layers. These methods could not exploit multi-dimensional information to its full. In BCI, traditional methods needs spatial filtering to take electrode's location into account for the spatial orientation of electrodes. Deep Neural Networks (Sturm et al., 2016) turn out to be a viable solution to the classification problems in practical applications as it uses two-dimensional images which also includes the special orientation of the electrodes which is evident in Fig. 3.

Despite all these impressive results obtained, there is still room for improvement in terms of accuracy, interpretability, and usability for real-time applications. Automatic detection of MI-based EEG requires a sophisticated learning algorithm. Deep learning with a convolutional neural network (CNN) is a recent development in the machine learning field, which has performed incredibly well in EEG data analysis (Schirrmeister et al., 2017; Lawhern et al., 2018). Recently, these learning algorithms had worked well in the detection of emotion (Akalya Devi et al., 2018), epilepsy (Antoniades et al., 2016), Parkinson's disease (Acharya et al., 2018), and seizure (Acharya et al., 2018) using EEG.

2. Related Work

Substantial studies have been carried out to improve the classification accuracy of MI data. The performance of these systems mostly depends upon features selected and the algorithm employed for classification purposes. Fourier transform has been used for the extraction of elements in the different frequency spectrum and features in the EEG data. With higher-order statistics of Power Spectral Density (PSD) at different orders, features such as Logarithmic amplitudes, Spectral moment for first and second-order yielded. A minimal misclassification rate of 10 percent was observed (Zhou et al., 2008) using SVM and Linear Discriminant Analysis (LDA). Wavelet transform among tools has garnered much trust to decompose EEG signals and thus has been used for feature extraction by researchers (Murugappan et al., 2010; Hazarika et al., 1997; Jahankhani et al., 2006; Subasi, 2007). Brain's spatial patterns contain significant features, and common spatial patterns (CSP) features have been used widely for EEG signal processing and mainly for BCI purposes. CSP along with SVM is used in (Liu et al., 2010), which yields 82.6% classification accuracy for BCI competition II data set III. Different classification techniques such as KNN (Khasnobish and Bhattacharyya, 2010) which yields 84.29 percent classification accuracy for two-class motor imagery, with average band power of Alpha (Mu) frequency band and Beta band. In a recent study, Shannon entropy as a feature provided 86.4% classification accuracy using the SVM classifier (Kant et al., 2019).

Recently, Deep learning is a pivotal platform that has attracted the researcher's attention. Deep learning is mainly available in image and video classification domain. It is new to biomedical signals, and in recent years, many public reviews (Wainberg et al., 2018), deep learning in the biomedical field have been discussed. Classification problem using CNN and stacked autoencoders (SAE) along with Short Time Fourier Transform (STFT) has been employed for deep learning for MI data, and the results produced is a 9 percent improvement over the algorithm qualified for the EEG dataset by (Tabar and Halici, 2017). Although STFT is a competing tool for time-frequency analysis yet scope for improvement in time-frequency tradeoff is there. In our approach, CNN learns the activation patterns from the input data of MI signals. Convolution operations are subjected to the application on the time axis only, not on the location and frequency. Thus, the activation patterns shape (i.e., power values at various frequencies), and their occurrence (i.e., EEG channel location) are learned by the convolutional layer. Then, a vibrational autoencoder (VAE) with five hidden layers improves the classification through a deep network. Transfer Learning is enabling task implementation for training and testing models for multi-domain studies (Weiss et al., 2016).

In BCI, field researchers reviewed implementation strategies of transfer learning (Hajinoroozi et al., 2017). Implementation of transfer learning via CNN across subjects (Zheng and Lu, 2016) was demonstrated, and kernel principle analysis for identification of the parameter relationship of classifier (Lin and Jung, 2017) is evaluated in earlier studies. Reliable classification accuracy is always desired in BCI application as human subjects are involved. Improvement in accuracy with new algorithms and techniques promises reliable results and better response. In the paper, we intend to use transfer learning on top of existing deep networks to improve the classification accuracy using our approach.

3. Methodology

3.1. Input data

The dataset used is from BCI competition 2003, dataset III (Lemm et al., 2004), which is a well-established dataset for Motor Imagery (MI) classification tasks (Tabar and Halici, 2017). The dataset contains three channels EEG data from a healthy female, 25 old participant for the imagination of the left, and right-hand movement. The imaginary task was to move a block based on the given cue in the left or right direction. The data in the study contains recordings from the motor cortex region of the brain using three electrodes ($C3$, Cz and $C4$) during the motor imagery of both left/right hand movement. As Cz electrode has very little information related left or right hand motor imagery, we have considered C3 and C4 electrodes for our study.

These electrodes are located in the motor cortex area of the brain, which is shown in Fig. 1. The data contains 40 trials per session with half left and half right hand MI signals placed randomly. Seven sessions of such 40 trials have been recorded with their labels. The total duration for each trial was 9 seconds with a sampling frequency (Fs) 128 Hz. The recording scheme is shown in Fig. 2. Timing scheme of the task cortex. For the first two seconds, subject relaxes, in 2-3 seconds, a '+' sign is shown, and the cue (Left/Right) is shown starting from 3rd second to 9th second.

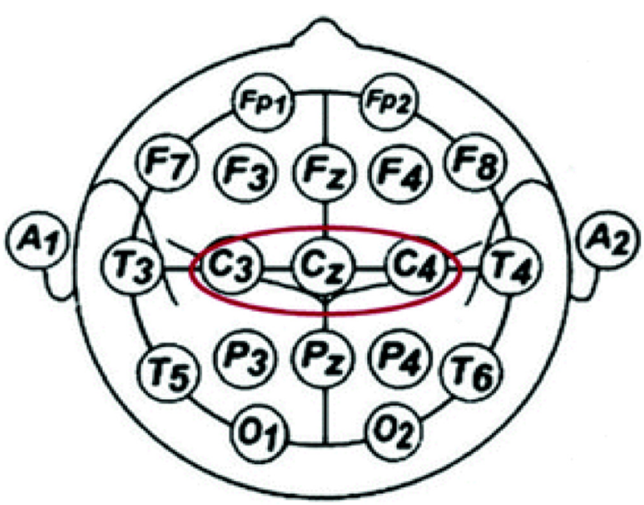


Fig. 1. Electrode position in the motor cortex.

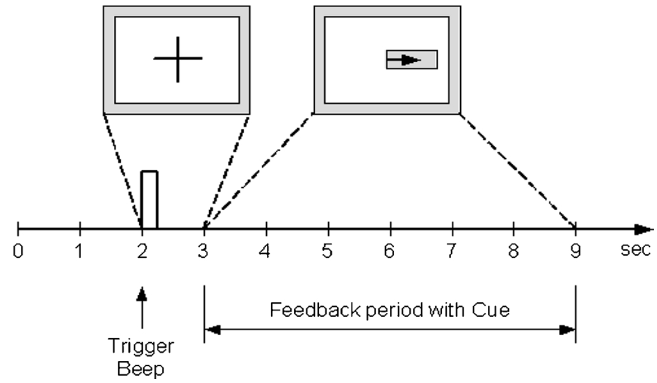


Fig. 2. Timing scheme of the task cortex.

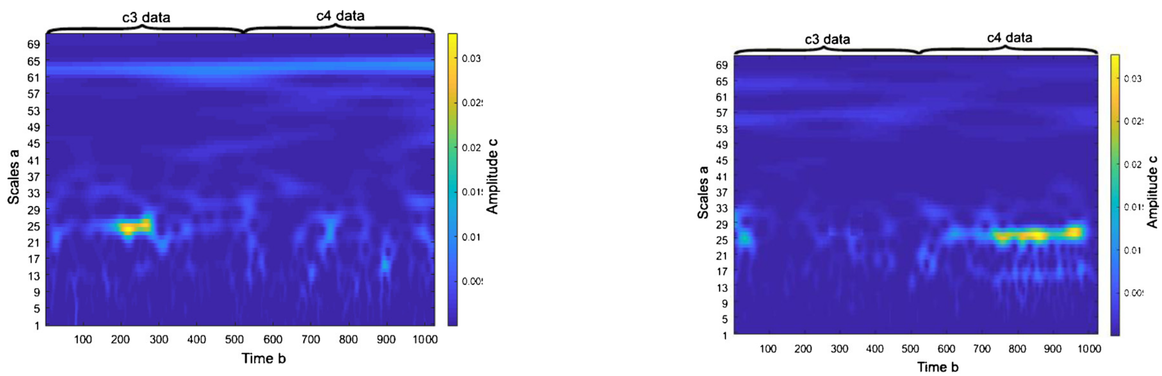
The data technically contains 280 out of which 140 trials were made available with their labels, and the other 140 samples were kept for evaluation. Each trial has a 9-second duration of data for each channel $C3$, Cz, and $C4$ per trial with all the labels available.

3.2. Preprocessing

The EEG data was bandpass filtered 8-30 Hz. Moreover, the signal is further decomposed into two separate bands Mu (8-14 Hz) and Beta (15-30 Hz) frequency bands (Scherer and Vidaurre, 2018) using wavelet packet transform (WPT), as these bands have been reported to have the information related to MI Wavelet transform has proven to be instrumental for analyzing a signal into the time-frequency domain. Wavelet transform methods such as WPT don't just work on a single scale, ie either time or frequency. It works as a multiscale transformation, that decomposes the signal into multiple scales (representing coarseness of the signal) (Kiymik et al., 2005).WPT allows us to deal with the time-frequency tradeoff, in contrast to Short-Time Fourier Transform (STFT) or spectrogram (Hajinoroozi et al., 2017; Zheng and Lu, 2016), that utilizes a single scale only.

In this study, we have used a scalogram that is represented by absolute values of Continuous Wavelet Transform (CWT) of the signal and can be plotted as a function of both frequency and time (Lee and Choi, 2018). MI signals are slowly varying events peppered by abrupt transients with features occurring at different scales, thus providing superior time localization for high frequency, short-duration events, and superior frequency localization for longer-duration, low-frequency events can be achieved using scalogram. One-dimensional EEG data is transformed via CWT through filter-bank (Christensen et al., 2019) which contains time-frequency and amplitude data in a single image. The filter-bank is a set of CWT parameters to be implemented on the given signal. In the experiment, Filterbank parameters were kept constant for CWT. The wavelet used for CWT is the analytic Morse wavelet (Chaudhary et al., 2019) as it has better time-frequency localization. For Morse wavelet symmetry parameter (gamma) and time bandwidth product were kept 3 and 60 respectively. Voices per octave were kept 10. Data from electrode C3 and C4 are stacked together, C4 after C3 in order to represent all the data into a single representation of one event (Left or Right) hand imagery as shown in Fig. 3(a) and (b). Processed data is further used to train the model using deep neural networks.

Deep neural networks employ Convolutional Neural Network (CNN) for feature recognition. Pre-trained CNN architectures used in the study require 2-dimensional image data as an input which makes 1-dimensional EEG signals non-compatible with those networks. One of the main advantages of CWT is that it transforms EEG signals into equivalent images thus making it compatible for training with deep networks. The image after the CWT contains the time, scale, and amplitude information. Time-amplitude values at different scales are the



(a) Left hand (b) Right Hand

Fig. 3. Scalogram representation for both hand Motor imagery.

features used in the study for motor imagery classification. In the last preprocessing step the generated 2-D image is rescaled according to the input requirement of the particular deep network, for instance, alexnet has an input size of 224 × 224 pixel. Therefore the image is resized to 224 × 224 pixels before feeding in for the transfer learning using alexnet.

3.3. Convolutional Neural Network

Convolutional Neural Network (CNN) is a multiple layer feedforward neural network, which includes different types of layers, which are convolution layer, ReLU layer, pooling layers, and fully connected output layers. CNN is designed to recognize features in images such as edges and shapes.

Convolutional Layer: The layer which comes first in CNN architecture is always a Convolutional Layer. Typically an input layer to a CNN accepts $MxNx3$. Here MxN is the two-dimensional size of image with three layers to represent a colored image with intensities of Red, Green, and Blue (RGB), respectively. CNN uses a filter with particular parameters which are having the same depth as the input image, and the filter is convolved with the image. The filter represents a curve or shape to which the input image is convolved with. The shape that resembles the curve in the input image that is represented by the filter ends up in higher values as a result of convolution. Convolution operation can be represented by equation (1).

$$s(t) = (x^*w)(t) \tag{1}$$

Pooling layer: The pooling layer is used to reduce the data size. Pooling involves arranging the matrix data in different segments and replacing the whole segment with a single value hence reducing the metrics data dimension. Some of the popular pooling functions are Max-pooling and Average-pooling, in which the segmented matrices are replaced by the maximum or average of all the values in the current segment as shown in Fig. 4.

3.3.1. Fully Connected layer

To fit the network layer architecture, dimensionality of layers is altered in a fully connected layer. A fully connected layer is a function operation that is between $\mathbb{R}m$ and $\mathbb{R}n$ each dimension of input and output are connected to each other. A fully connected layer connects all the activations from the previous layer to the next layer of the network, just as it is usually in a conventional artificial neural network.

3.3.2. Softmax layer

The Softmax function translates input from earlier layers into a probability for the classes that sum to one. Thus this layer plays a critical role in output as the predicted output is the class which has maximum probability for the given input data.

Several deep neural networks that have been used to classify images are available. There are two types of deep networks used in this study firstly, linear networks which as the name suggests have simple linear architecture and other, directed acyclic graph (DAG) networks. DAG networks are more complex and generally contain more than one parallel layer. Although these networks are pre-trained to classify other images we can modify them to suit our classification problem using transfer learning by tweaking required parameters. Fig. 5 shows some of those network architecture (Yu et al., 2016). Each network has its own set of convolutional layers based on the architecture. Several other deep learning DAG networks such as googlenet (Tang et al., 2017), resnet (s) (He et al., 2016), squeezenet (Gaikwad and El-Sharkawy, 2018) were used in the study to evaluate them to find out their applicability for the same data set.

3.4. Data Training

Traditional training methods, in comparison to deep networks, require a smaller sample size for training, whereas deep networks need a large sample size for training. To address this issue transfer learning uses pre-trained models as the starting point. Transfer learning repurposes the task on second related tasks and limits the amount of training data required. Deep learning methods have initially been designed to classify images. While time-series data such as EEG is not compatible with deep neural networks (Gaikwad and El-Sharkawy, 2018) as they require image as input. In our approach, we solved the issue with CWT. CWT transforms the time series data into equivalent image representation that can be trained in deep networks. In this study, we have taken several pre-trained and well established trained different deep networks, which are: *vgg19, alexnet, vgg16, googlenet, squeezenet, resnet50, googlenet, densenet201, resnet18, resnet101*. Here vgg16, vgg19, and alexnet are series networks while remaining networks are multi-chain Directed Acyclic Graph (DAG) networks.

Each network given above has its own specific size for input data. Each input image is resized according to the respective network's input layers size requirement. And for transfer learning, there were some layers (final fully connected layer and output layer) that needed to be replaced with a new one in order to make the network compatible with the EEG data and classes. We have used three sets of frequency spectrums first Mu + Beta (8-30 Hz), Mu (8-14 Hz), and Beta (14-30 Hz) frequency bands to determine which frequency band performs best with our approach. Data was split into a training set and testing set 50% of the 280 samples, which were then used to train the model and for verification with true data labels to test the performance of the trained model, respectively. The dataset has 140 training sets and 140 test sets. Thus the model was trained and tested using 140 datasets each. Fig. 6 demonstrates typical training progress for deep networks on Matlab software.

For all the networks have been kept training hyperparameters for

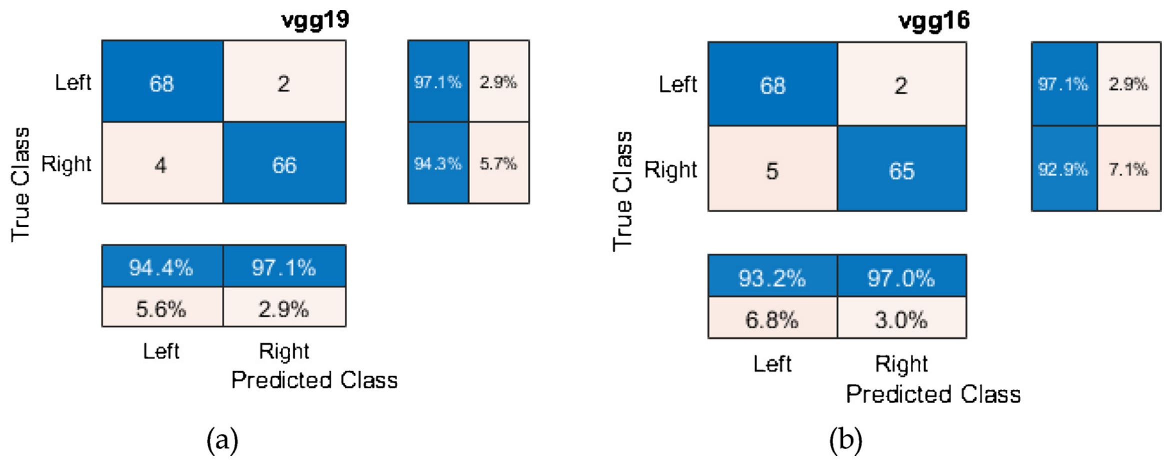


Fig. 4. Sample pooling of pixels (a) Pooling pixel for first four pixels (b) Subsequent pixel pooling.

constant. We have divided data in multiple epochs witch are allowed to go maximum 25. Mini-batch size is the number that represents the number of samples after which the internal parameters of the model are updated. Mini-batch size for training in our experiment was kept 7, and the Initial learning rate for each training was kept 0.0001.

4. Results

For performance evaluation of our approach, we have used classification accuracy, and the Kappa score is calculated. Both evaluation methods were applied on three decomposed sets (Full-spectrum (8-

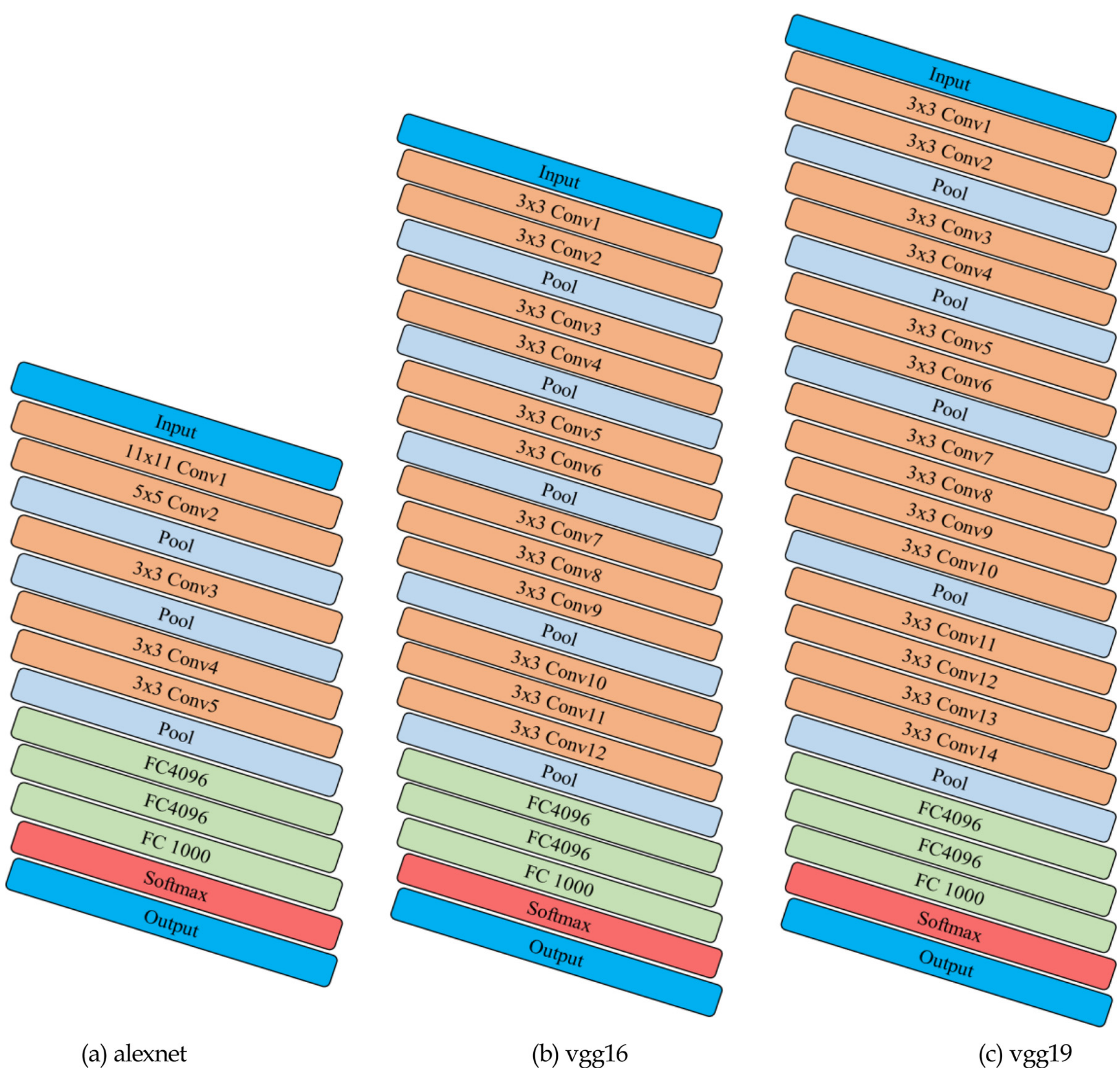
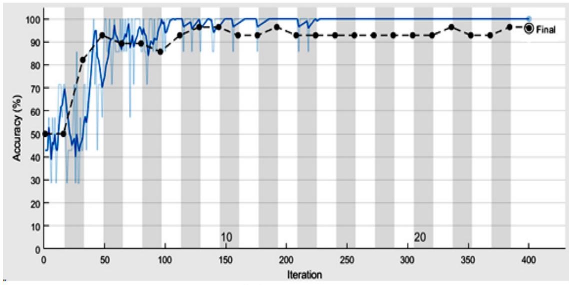
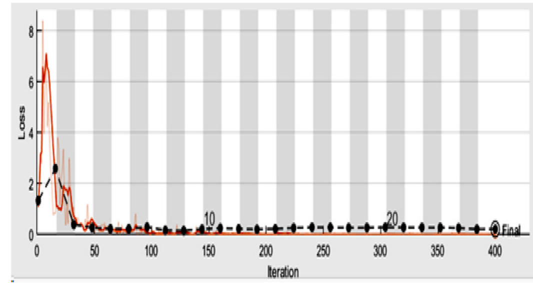


Fig. 5. Typical architecture of some of the pre-trained deep networks used in the study.



(a)



(b)

Fig. 6. Training progress plots for vgg19 (a) Training Accuracy (b) Training Loss.

30 Hz), Mu (8-14 Hz), and Beta (14-30 Hz)). This can help us determine the most relevant frequency band or spectrum that goes along with transfer learning.

4.1. Classification Accuracy

To evaluate the performance of the approach, we have held out the testing data, and that with labels are tested for percentage of true predictions.

$$Accuracy = \left(\frac{N_{Correct}}{N_{total}}\right) \times 100\,\% \tag{2}$$

Here $N_{Correct}$ are the number of correct classification, and the total number of trials in test dataset is represented by N_{total}.

4.2. Kappa Score

In order to evaluate the performance of our method, we have calculated the Kappa score. Kappa score is considered a statistically robust measure to verify the quality of the method (Ismail Fawaz et al., 2019). Generally, the values of the kappa score range between 1 to -1, where positive 1 indicates perfect classification. We found out kappa values as follows:

$$k = \frac{Acc - \mathrm{R_{Acc}}}{1 - \mathrm{R_{Acc}}} \tag{3}$$

Here Acc is the accuracy achieved, and $\mathrm{R_{Acc}}$ is random accuracy. Random accuracy is calculated as follows:

$$R_{Acc} = 1/N \tag{4}$$

Here N is the number of total classes. In this experiment, N is 2. Kappa values are calculated for each network according to equation 3 for the best classification accuracy achieved.

One important objective of our work was to implement Deep learning strategies using a limited amount of data. The use of transfer learning allows us to use pre-existing deep networks and modify them for our requirements with comparatively less amount of training data.

The results obtained are shown in terms of accuracy in Table 1. Table 2, which comprise of the best Kappa values for different transfer networks based on the respective deep network. As we saw in previous approaches based on dataset BCI competition 2003 data set III, most of the conventional methods achieved good accuracy up to 90.0%. It's important to point out deep learning, especially series networks such as vgg19, vgg16, and alexnet are have shown significant improvement in classifying motor imagery signals.

Table 2
Classification accuracy using the test dataset.

Network Name	Full-spectrum (8-30 Hz)	Mu (8-14 Hz)	Beta (14-30 Hz)
vgg19 (net-1)	95.71	94.29	72.14
Alexnet (net-2)	92.86	93.57	62.14
vgg16 (net-3)	95.00	94.29	69.29
Squeezenet (net-4)	72.86	78.57	64.29
resnet50 (net-5)	70.00	70.71	55.00
Googlenet (net-6)	72.14	70.00	65.71
densenet201 (net-7)	57.86	67.86	60.00
resnet18 (net-8)	57.86	64.29	42.14
resnet101 (net-9)	62.14	65.00	57.14

Training and test data have been used for transfer learning on nine different deep neural networks, out of which three network sets have achieved a maximum of 93.57% accuracy, which to the best of our knowledge has not been reported for the current dataset. Maximum accuracy achieved reached up to 95.71% in the case of vgg19, which is by far the best classification accuracy for the dataset in our experiments. Although the data were limited and, large data size can improve the results, as it can provide better variation for training data and a comprehensive testing dataset.

The results of the study indicated that Mu frequency band is good for the MI features as in vgg19, alexnet, and vgg16. Though the inclusion of beta-band (Full spectrum, 8-30 Hz) sometimes improves the model and gives better classification accuracy as in vgg19 and vgg16 it is evident that the Beta band alone did not contain substantial features for classification. Mu or alpha band is arguably most significant for classification and most of the time provides better results.

Series networks (*net 1-3*) outperformed branched DAG networks (*net 4-9*) both in terms of classification accuracy of the MI data and the Kappa score for networks (*net 1-3*). Some of the confusion matrices are shown in Fig. 7. There was a stark difference between Kappa score for linear networks is 87.14 and above, while branched DAG networks could attain 57.14 or even less as shown in Fig. 8. Even linear networks showed the highest accuracy of 93.57% or more, on the other side branched DAG networks could attain maximum 78.57% accuracy. The results evidently show the effectiveness of linear networks over branched DAG networks for our approach.

5. Discussion and Conclusion

In previous works study of the classification of MI signal have been discussed using various classification algorithms with conventional methods and with deep learning. Table 3 presents a comparison of

Table 1
Kappa values for top accuracy achieved by networks used.

Network	Vgg19	alexnet	vgg16	squeezenet	Resnet50	googlenet	Densenet201	Resnet18	Resnet101
Kappa	0.91	0.87	0.90	0.57	0.41	0.44	0.36	0.29	0.30

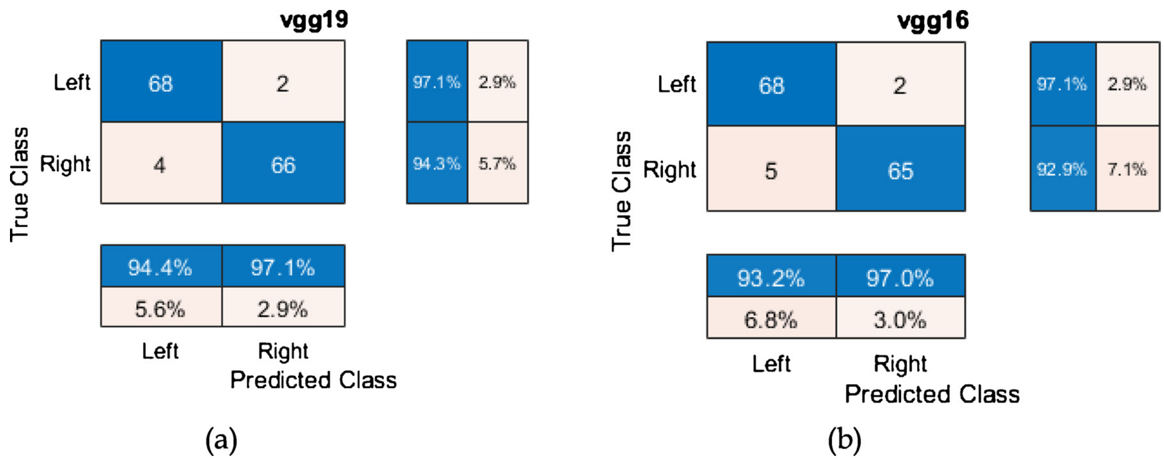


Fig. 7. Confusion Matrices for accuracy (a) 95.71% for test data using vgg19 (b)95% for test data using vgg16.

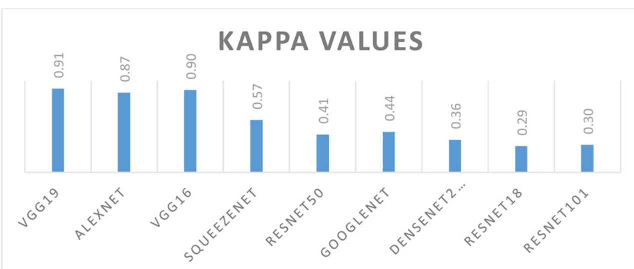


Fig. 8. Kappa values for different transfer learning models.

various methods used by researchers on the same dataset that is earlier discussed in the related work section.

Different approaches to the classification of the motor imagery data have been used to classify the MI signals. Results in terms of accuracy are shown in Table 3. However several studies have shown minimal misclassification rate (MMR) in percentage. Ch. Schafer, et al. reported 10.71% MMR using the Bayesian classifier which happens to be the winner of the BCI competition for the same data. Several other studies using LDA classifier by X. GAO, et al., A. Narayana, et al., M. Sadashiv, et al. reported 13.57, 15.71, and 17.14 % MMR respectively (Blankertz et al., 2004). Earlier approaches mention used one-dimensional data with features extracted from it. However, when time-frequency-amplitude information combined along with the spatial features of the data, it provides promising results. Our approach has the advantage of having special features along with amplitude information which was further fed to different networks for transfer learning.

In this work, an approach to classify MI signals using transfer learning has been proposed. The authors used Transfer learning successfully to classify the two-class (Left/Right-hand motor imagery) EEG data. Out of three frequency bands (8-30 Hz, 8-14 Hz, and 14-30 Hz) evaluated Mu band proven to be most informative in terms of features, although in some cases, it enhances accuracy along with the Beta band.

Thus as results show although beta band alone is not very feature-rich for MI signals. It also carries some features which can be helpful when combined with mu band. The study showed the use of deep learning over conventional ANN could be instrumental for BCI applications. It has been observed that (a) Mu band features are most influential for the classification of MI signals, (b) Using the transfer learning method, a better classification rate can be achieved even with a limited amount of data.

The only drawback of the approach is that it takes long training hours to train the model in comparison to conventional methods but it considerably fast in classification once the model is fully trained. Training accuracy with limited data-set provided good classification, however, larger data-set may have improved the results.

In sensitive applications like medical diagnosis, our approach is promising and can be implemented where minimal errors are imitable. Once a model is trained, it takes substantially less time for a classification that makes this approach feasible to implement in a real BCI system.

Moreover, in further work, we plan to adapt more state-of-the-art methods for feature extraction and classification methods such as spiking neural networks (Kumarasinghe et al., 2020), joint time-frequency–space classification (Zhao et al., 2019), and enhance this technique to develop a network dedicated to BCI classification tasks.

CRediT authorship contribution statement

Piyush Kant: Conceptualization, Methodology, Software, Investigation, Writing - original draft, Writing - review & editing. **Shahedul Haque Laskar:** Supervision, Project administration. **Jupitara Hazarika:** Methodology, Resources. **Rupesh Mahamune:** Formal analysis, Visualization.

Table 3
Notable previous research work is done with the same dataset.

Authors	Year Published	Method used	Classification Accuracy
Khasnobish and Bhattacharyya (2010)	2010	Average band power of alpha and beta with KNN	84.29%
Liu et al. (2010)	2010	Common spatial pattern (CSP) with SVM	82.86%
Li et al. (2013)	2013	Adaptive power projection with Bayesian classifier	90%
Jang et al. (2016)	2016	STFT with K-Nearest Neighbor	83.57%
Tabar and Halici (2017)	2016	STFT with deep learning	90%
Jang et al. (2016)	2019	Genetic Algorithm based optimization with FKNN, LDA	84%
Eslahi et al. (2019)	2019	Wavelet transform with Shannon entropy using SVM and KNN	86.4%
Our approach		CWT Filter-bank, Transfer learning	95.71%

Declarations of Competing Interest

None.

References

Tabar, Y.R., Halici, U., 2017. A novel deep learning approach for classification of EEG motor imagery signals. J. Neural Eng. 14 (February (1)).
Hsu, W.Y., Sun, Y.N., 2009. EEG-based motor imagery analysis using weighted wavelet transform features. J. Neurosci. Methods 176 (2), 310–318.
Khasnobish, A., Bhattacharyya, S., 2010. K-Nearest Neighbor Classification of Left-Right Limb Movement Using EEG Data. In: International Conference on Systems in Medicine and Biology. ICSMB. pp. 1–6 May 2016.
Klimesch, W., Schimke, H., Doppelmayr, M., Ripper, B., Schwaiger, J., Pfurtscheller, G., 1996. Event-related desynchronization (ERD) and the Dm effect: Does alpha desynchronization during encoding predict later recall performance? Int. J. Psychophysiol. 24 (1–2), 47–60.
Göksu, H., 2018. BCI oriented EEG analysis using log energy entropy of wavelet packets. Biomed. Signal Process. Control 44, 101–109.
Kim, Y., Ryu, J., Kim, K.K., Took, C.C., Mandic, D.P., Park, C., 2016. Motor Imagery Classification Using Mu and Beta Rhythms of EEG with Strong Uncorrelating Transform Based Complex Common Spatial Patterns. Comput. Intell. Neurosci. 2016.
McFarland, D.J., Miner, L.A., Vaughan, T.M., Wolpaw, J.R., 2000. Mu and beta rhythm topographies during motor imagery and actual movements. Brain Topogr. 12 (3), 177–186.
Zhang, Y., Zhou, G., Jin, J., Wang, X., Cichocki, A., 2015. Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface. J. Neurosci. Methods 255 (November), 85–91.
Bhattacharyya, S., Khasnobish, A., Chatterjee, S., Konar, A., Tibarewala, D.N., 2010. Performance analysis of LDA, QDA and KNN algorithms in left-right limb movement classification from EEG data. In: in International Conference on Systems in Medicine and Biology. ICSMB 2010 - Proceedings. pp. 126–131.
Christensen, S.M., Holm, N.S., Puthusserypady, S., 2019. An improved five class MI based BCI Scheme for Drone Control Using Filter Bank CSP. In: 7th International Winter Conference on Brain-Computer Interface. BCI 2019.
Zhou, S.M., Gan, J.Q., Sepulveda, F., 2008. Classifying mental tasks based on features of higher-order statistics from EEG signals in brain-computer interface. Inf. Sci. (Ny). 178 (6), 1629–1640.
Lee, M., Bressler, S., Kozma, R., 2017. Advances in Cognitive Engineering Using Neural Networks. Neural Networks 92, 1–2.
Kang, H., Choi, S., 2014. Bayesian common spatial patterns for multi-subject EEG classification. Neural Networks.
Murugappan, M., Ramachandran, N., Sazali, Y., 2010. Classification of human emotion from EEG using discrete wavelet transform. J. Biomed. Sci. Eng. 03 (04), 390–396.
Liu, C., Bin Zhao, H., Li, C.S., Wang, H., 2010. CSP/SVM-based EEG classification of imagined hand movements. Dongbei Daxue Xuebao/Journal Northeast. Univ. 31 (8), 1098–1101.
Sturm, I., Lapuschkin, S., Samek, W., Müller, K.R., 2016. Interpretable deep neural networks for single-trial EEG classification. J. Neurosci. Methods 274 (December), 141–145.
Schirrmeister, R.T., et al., 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. Hum. Brain Mapp. 38 (November (11)), 5391–5420.
Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J., 2018. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. J. Neural Eng. 15 (5).
Akalya Devi, C., KarthikaRenuka, D., Soundarya, S., 2018. A survey based on human emotion identification using machine learning and deep learning. Journal of Computational and Theoretical Nanoscience 15 (5), 1662–1665.
Antoniades, A., Spyrou, L., Took, C.C., Sanei, S., 2016. Deep learning for epileptic intracranial EEG data. In: IEEE International Workshop on Machine Learning for Signal Processing. MLSP. vol. 2016-Novem.
Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adeli, H., 2018. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput. Biol. Med. 100, 270–278.
Hazarika, N., Chen, J.Z., Tsoi, A.C., Sergejew, A., 1997. Classification of EEG signals using the wavelet transform. In: International Conference on Digital Signal Processing. DSP. pp. 89–92 vol. 1.
Jahankhani, P., Kodogiannis, V., Revett, K., 2006. EEG signal classification using wavelet feature extraction and neural networks. In: Proceedings - IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing. JVA 2006. pp. 120–124.
Subasi, A., 2007. EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst. Appl. 32 (4), 1084–1093.
Kant, P., Hazarika, J., Laskar, S.H., 2019. Wavelet transform based approach for EEG feature selection of motor imagery data for braincomputer interfaces. In: Proceedings of the 3rd International Conference on Inventive Systems and Control. ICISC 2019. pp. 101–105.
Wainberg, M., Merico, D., Delong, A., B. F.-N. biotechnology, and undefined, 2018. Deep learning in biomedicine. nature.com.. .
Weiss, K., Khoshgoftaar, T.M., Wang, D.D., 2016. A survey of transfer learning. J. Big Data 3 (December (1)).
Hajinoroozi, M., Mao, Z., Lin, Y.P., Huang, Y., 2017. Deep transfer learning for cross-subject and cross-experiment prediction of image rapid serial visual presentation events from EEG data. in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10284. pp. 45–55.
Zheng, W.L., Lu, B.L., 2016. Personalizing EEG-based affective models with transfer learning. In: IJCAI International Joint Conference on Artificial Intelligence. vol. 2016-Janua. pp. 2732–2738.
Lin, Y.P., Jung, T.P., 2017. Improving EEG-based emotion classification using conditional transfer learning. Front. Hum. Neurosci. 11 (June).
Lemm, S., Schafer, C., G. C.-I. T. on, and undefined, 2004. BCI competition 2003-data set III: probabilistic modeling of sensorimotor/spl mu/rhythms for classification of imaginary hand movements. ieeexplore.ieee.org. .
Scherer, R., Vidaurre, C., 2018. Motor imagery based brain–computer interfaces. Smart Wheelchairs and Brain-Computer Interfaces. pp. 171–195.
Kiymik, M.K., Güler, I., Dizibüyük, A., Akin, M., 2005. Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application. Comput. Biol. Med. 35 (7), 603–616.
Lee, H.K., Choi, Y.S., 2018. A convolution neural networks scheme for classification of motor imagery EEG based on wavelet time-frequecy image. In: International Conference on Information Networking. vol. 2018-Janua. pp. 906–909.
Chaudhary, S., Taran, S., Bajaj, V., Sengur, A., 2019. Convolutional Neural Network Based Approach Towards Motor Imagery Tasks EEG Signals Classification. IEEE Sens. J. 19 (12), 4494–4500.
Yu, W., et al., 2016. Visualizing and Comparing AlexNet and VGG using Deconvolutional Layers. Icml 48, 1–7.
Tang, P., Wang, H., Kwong, S., 2017. G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition. Neurocomputing 225 (November 2016), 188–197.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 2016, December. pp. 770–778.
Gaikwad, A.S., El-Sharkawy, M., 2018. Pruning convolution neural network (squeezenet) using taylor expansion-based criterion. 2018 IEEE Int. Symp. Signal Process. Inf. Technol. ISSPIT 2018 2019 (Janua), 1–5.
Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A., 2019. Deep learning for time series classification: a review. Data Min. Knowl. Discov. 33 (July (4)), 917–963.
Li, C.Y., Liu, R., Wang, Y.Y., Wang, Y.X., Li, X., 2013. Adaptive power projection method for accumulative EEG classification. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. EMBS. pp. 7052–7055.
Jang, T.U., Kim, B.M., Yang, Y.M., Lim, W., Oh, D.H., 2016. Motor-imagery EEG signal classification using position matching and vector quantisation. Int. J. Telemed. Clin. Pract. 1 (4), 306.
Eslahi, S.V., Dabanloo, N.J., Maghooli, K., 2019. A GA-based feature selection of the EEG signals by classification evaluation: Application in BCI systems. Biomed. Signal Process. Control 32 (January (2017)), 69–75.
Blankertz, B., et al., 2004. The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials. IEEE Transactions on Biomedical Engineering 51 (6), 1044–1051.
Kumarasinghe, K., Kasabov, N., Taylor, D., 2020. Deep learning and deep knowledge representation in Spiking Neural Networks for Brain-Computer Interfaces. Neural Networks 121, 169–185.
Zhao, D., Tang, F., Si, B., Feng, X., 2019. Learning joint space–time–frequency features for EEG decoding on small labeled data. Neural Networks 114 (June), 67–77.