

Article

Classification of Motor Imagery EEG Signals Based on Data Augmentation and Convolutional Neural Networks

Yu Xie ^{1,*}  and Stefan Oniga ^{1,2,*} ¹ Faculty of Informatics, University of Debrecen, 4032 Debrecen, Hungary² North University Center of Baia Mare, Technical University of Cluj-Napoca, 400114 Cluj-Napoca, Romania

* Correspondence: yu.xie@inf.unideb.hu (Y.X.); oniga.istvan@inf.unideb.hu (S.O.)

Abstract: In brain–computer interface (BCI) systems, motor imagery electroencephalography (MI-EEG) signals are commonly used to detect participant intent. Many factors, including low signal-to-noise ratios and few high-quality samples, make MI classification difficult. In order for BCI systems to function, MI-EEG signals must be studied. In pattern recognition and other fields, deep learning approaches have recently been successfully applied. In contrast, few effective deep learning algorithms have been applied to BCI systems, especially MI-based systems. In this paper, we address these problems from two aspects based on the characteristics of EEG signals: first, we proposed a combined time–frequency domain data enhancement method. This method guarantees that the size of the training data is effectively increased while maintaining the intrinsic composition of the data. Second, our design consists of a parallel CNN that takes both raw EEG images and images transformed through continuous wavelet transform (CWT) as inputs. We conducted classification experiments on a public data set to verify the effectiveness of the algorithm. According to experimental results based on the BCI Competition IV Dataset2a, the average classification accuracy is 97.61%. A comparison of the proposed algorithm with other algorithms shows that it performs better in classification. The algorithm can be used to improve the classification performance of MI-based BCIs and BCI systems created for people with disabilities.

Keywords: motor imagery (MI); electroencephalogram (EEG); data augmentation (DA); convolutional neural network (CNN)

**Citation:** Xie, Y.; Oniga, S.Classification of Motor Imagery EEG Signals Based on Data Augmentation and Convolutional Neural Networks. *Sensors* **2023**, *23*, 1932. <https://doi.org/10.3390/s23041932>

Academic Editors: Andrea Bizzego, Annabel Chen and Gianluca Esposito

Received: 20 January 2023

Revised: 6 February 2023

Accepted: 7 February 2023

Published: 9 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A brain–computer interface (BCI) is a system that facilitates communication between the human brain and external devices (such as a computer or other electronic devices) without the need for any intermediaries. It allows the user to control the computer or smart device directly through signals generated by the brain without the involvement of peripheral organs and muscles. BCI research has significant theoretical, military, medical, and recreational value. BCI research has considerable value across a wide range of fields [1–4]. Theoretically, it has the potential to offer unprecedented insight into the functioning of the human mind. For military purposes, it can enhance soldiers’ efficiency in the field. In the medical realm, it could potentially be used to treat certain conditions, such as paralysis, that have previously been considered untreatable. Finally, it enables the creation of captivating leisure experiences, allowing users to delve into virtual settings and interact with them in unprecedented ways. Electroencephalography (EEG) is a widely utilized signal for building BCI. It offers many advantages, such as being cost-effective, non-invasive, and portable [5]. At the same time, we face the challenge of the high dimensions EEG signal being too weak and the signal-to-noise ratio being low [6]. Furthermore, MI-EEG is a non-linear and unstable signal, which means that its parameters (e.g., mean and variance) change over time [7].

Several scholars have recently used convolutional neural networks to feature the extraction of EEG signals. This technique significantly decreases the number of connections

and parameters in a deep network while using the pertinent geographical or temporal links between nearby data points to infer helpful features for a particular machine learning job. Ref. [8] proposed a frequency complementary map selection (FCMS) scheme based on augmented CSP (ACSP) for convolutional neural networks (CNNs) to reduce the dependence of feature mapping across frequency bands. However, this method fails to take full advantage of the temporal information of EEG. Ref. [9] used short-time Fourier transformation (STFT) to map the EEG signals to image signals and then used CNN and automatic stacking encoding (SAE) to extract features and classify them. In [10], a CNN is used to learn the temporal information of EEG based on envelope representations filtered by Hilbert transforms. Since the envelope signal is a low-frequency signal, it can be down-sampled, thus reducing the dimensionality of the data. CNNs and other deep learning algorithms need a large amount of data in order to achieve competitive performance on the target task. If we lack sufficient data to train, training deep learning models can overfit certain features to a specific training set, thus limiting the generalizability of the model. The model's limitations lie in its inability to correctly predict or comprehend new data, due to only being trained on a particular set of data. In order to increase the accuracy of the model, we need more data and ensure that it is representative of the data we expect to encounter in the future. This will help the model to generalize better and provide more accurate predictions.

Due to the experiments' intricate procedures and rigid guidelines to acquire MI-EEG signals, it is challenging to obtain enough training data to classify MI-EEG signals [11]. For instance, blinking and body trembling can both produce sudden spikes in the EEG signal; these are known as eye movement artifacts and muscle movement artifacts, respectively [12]. Additionally, acquiring a valid sample requires several preparation processes, following the pointing arrows on the screen and imagining the movement followed by a short pause. All this requires the subject's attention during this period, but as the experiment time increases, it becomes challenging to maintain attention all the time. As a result, it is incredibly difficult to collect enough good-quality MI-EEG signals for the training process. Researchers usually employ data augmentation (DA) technology to address general overfitting issues. Through data transformation or oversampling, DA techniques enhance the volume of existing data sets [13]. It has been shown that DA is successful in a variety of fields, including target recognition and image processing [14]. Typically, enhanced data are produced using two techniques. The first method involves employing geometric transformations, while the second entails adding noise to the existing training data. Ref. [15] proposed a multi-input CNN to classify multi-channel MI-EEG Signals. They rotated and flipped the MI-EEG image to achieve the purpose of increasing the data volume. Ref. [11] proposes a DA method that combines rotating images and adding noise. The EEG, in contrast to images, is a collection of boisterous, particularly relevant (in time and space), non-stationary time series from various electrodes. Geometric transformations do not directly apply to EEG data due to the risk of distorting time-domain features [16]. Ref. [17] applied the empirical mode decomposition (EMD) to create new artificial EEG frames, followed by transforming all EEG data into tensors as inputs of the neural network by complex Morlet wavelets. EMD separates the EEG signal into its intrinsic mode functions (IMFs), which are usually narrowband, and the resulting IMFs can be analyzed individually to extract information about specific frequency bands, which are thought to be related to different cognitive processes. DA methods [13,18,19] appropriate for EEG signals are crucial to be investigated at this point. Overall, designing a CNN suitable for feature extraction and classification for a segment of highly complex and unstable EEG signals is a great challenge, while using CNNs for signal classification has the potential demand for a large amount of data and system robustness. In this paper, we address these problems from two aspects based on the characteristics of EEG signals: first, we proposed a combined time–frequency domain data enhancement method. Our proposed DA method differs from current approaches in that it operates in both the time and frequency domains to transform the data, which ensures the effectiveness of increasing the size of the training data, while still preserving the inherent

structure of the data. Second, we designed a parallel inputs CNN followed by raw image and continuous wavelet transform (CWT) transforming EEG image as inputs and conduct classification experiments on a public data set to verify the effectiveness of the algorithm. This model not only emphasizes the main features of the original data, but also preserves other valuable features of the original data.

2. Methods

CWT and CNNs are both used in EEG analysis to extract essential features and patterns from EEG signals. The CWT provides a time–frequency representation of EEG signals, capturing both frequency content and temporal evolution. This is important in EEG analysis because EEG signals are non-stationary and transient in nature, and the CWT can effectively capture these features. CNN, on the other hand, is a type of deep learning model that is well-suited for image and signal analysis, making them a useful tool in EEG analysis. By learning complex patterns and relationships in EEG signals, CNNs can identify important EEG features, such as spikes and oscillations, that are indicative of brain activity and neurological conditions. By leveraging the advantages of both CWT and CNNs, EEG analysis can take advantage of the time–frequency representations provided by CWT, as well as CNNs’ ability to identify intricate patterns in EEG signals. This leads to improved EEG analysis and a better understanding of brain activity and neurological conditions.

2.1. Continuous Wavelet Transformation (CWT)

CWT is a popularly used method for the time–frequency analysis of signals. It was proposed by Morlet and Grossman in 1987. CWT technique provides a time–frequency representation of signals by decomposing a signal into wavelets of different scales, which are functions that are used to analyze the different frequency components of the signal. This allows for a more complete analysis of signals compared to traditional frequency analysis methods, which only provide a single frequency representation of the signal at a specific point in time. The CWT has proven to be an effective tool for analyzing EEG signals with respect to their frequency and temporal characteristics, which is crucial for understanding brain activity and diagnosing various neurological conditions [20–22].

The expression of the continuous wavelet transform is shown in Equation (1) [23].

$$w_s(a, \tau) = a^{\frac{1}{2}} \int s(t) \phi^* \left(\frac{t - \tau}{a} \right) dt, \quad (1)$$

where $s(t)$ is the input signal, a is the scaling of the wavelet transform, ϕ is the wavelet basis function, and τ is the time offset. There are five commonly used wavelet basis functions: Morlet wavelet, Mexican Hat wavelet, Haar wavelet, Daubechies wavelet and SymN wavelet clusters. We choose the Morlet wavelet as the wavelet basis function. Its time-domain expression is as follows:

$$\phi(t) = \left(\frac{2}{\pi T^2} \right)^{\frac{1}{4}} \exp \left(-\frac{t^2}{T^2} + jw_c t \right), \quad (2)$$

The expression of frequency is:

$$\phi(w) = \left(\frac{T^2}{2\pi} \right)^{\frac{1}{4}} \exp \left(-\frac{(w - w_c)^2}{4T^2} \right), \quad (3)$$

By analyzing the data, T and w_c of the wavelet function are determined. We use CWT to preprocess the original signal and then use the mapped time–frequency domain image as one of the inputs to the proposed CNN.

2.2. CNN

Deep learning is an essential aspect of machine learning—an emerging field that keeps moving forward. In many fields, deep learning has gradually become a pioneer, attracting the attention of numerous scholars [24–27]. In deep learning, CNNs are a widely used neural network model with multiple applications across various fields [28–30]. Meanwhile, there is also relevant research progress in BCI systems [31,32]. Multiple convolutional and pooling layers can be combined in the network structure's middle, followed by fully connected layers.

The heart of CNN is the convolutional layer and its primary function of the convolutional layer is to extract features from the input signal. The ability of the convolutional layer to perform the relevant convolutional operations on the input signal is crucial. Convolutional kernels, also known as filters, can exist in more than one in the framework of a single convolutional layer. The weight parameters and bias of the convolution kernels can be changed during the training of the neural network. Matrix multiplication principles can be used by convolutional operations to generate feature mappings from the input to the output. The position of the neural element in the feature map output by the k th convolution kernel is assumed as (m, n) . The output is shown in Equation (4)

$$y_{m,n} = f(w(i)_k * I(m, n) + b), \quad (4)$$

where $I(m, n)$ is the input data, b is the bias, $w(i)_k$ is the k th convolution kernel of the i th layer, f is the activation function of the nerve element, and its common form contains $\tanh$, sigmoid, and rectified linear unit (ReLU) [33].

$$\tanh : f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}, \quad (5)$$

$$\text{sigmoid} : f(x) = \frac{1}{1 + e^{-x}}, \quad (6)$$

$$\text{ReLU} : f(x) = \max(0, x), \quad (7)$$

On the side, subsequent connections to tighter layers enable the extraction of more distinct features. The features of high-dimensional input data can be continuously extracted by combining and superimposing multiple convolutional layers so that more advanced abstract features can be obtained from the signal to the greatest extent. In addition to the convolution operation, the convolution layer also includes the processing of operations such as padding and stride, and the associated computational process is more complex.

Pooling, also known as the sampling layer, is the process of reducing the input image's length and width sizes. Following the convolutional layer, the pooling layer can perform the operation of downsampling to extract local features. The pooling layer allows reducing the number of parameters of the network, which means that the computational effort and complexity of the model are also decreased. At the same time, this can be very robust to small errors in the data and improve the overfitting problem of the network model. The pooling layer has two primary pooling operations. By maximum pooling, the target region's feature value is determined by the largest element value in the region. The maximum pooling operation emphasizes more local features and optimizes the small errors generated by the convolution layer. By averaging pooling, similarly, the target region's feature value is determined by the average value of elements in it. In addition to retaining more adjacency information, the averaging pooling operation improves the region error caused by the size of the convolution kernel. The maximum and average pooling expression is shown in Equations (8) and (9). We assume that the size of the pooling kernel is $(N * N)$.

$$\text{Max pooling} : f(x) = \max(x_{[m,m+N],[n,n+N]}), \quad (8)$$

$$\text{Average pooling : } f(x) = \frac{1}{N * N} \sum_{m=m_1, n=n_1}^{m=m_1+N, n=n_1+N} x_{m,n}, \quad (9)$$

The fully connected layer is generally treated as the concluding component of the CNN structure. The feature maps produced by intermediate layers are converted into vector format by the feature space transformation when it passes through the fully connected layer, which combines the previously extracted features for use in matrix multiplication. The fully connected layer is responsible for transforming the spatial high-dimensional features that have already been extracted by the CNN and concluding the overall learning process with non-linear mapping.

The better robustness of the CNN network model and its strong generalization performance are mainly attributed to the use of design ideas such as sparse connectivity of convolutional layers, weight sharing, sampling of pooling layers and non-linear mapping of fully connected layers. In traditional recognition studies, a large number of samples are analyzed and processed as a requirement for effective feature extraction and classification recognition. In short, CNN models are automatically trained on the intrinsic features of the signal through convolutional operations and other related operations [34]. CNN has not only excellent feature extraction effects but also has strong interpretability in its network model structure.

2.3. Proposed CNN Structure

Event-related desynchronization (ERD)/event-related synchronization (ERS) phenomena typically occur between 8 and 32 Hz in MI-EEG signal tasks, according to related studies [35,36]. In [37], the researchers designed a CNN model with a parallel structure using three different frequency bands as input. They also designed a mixed convolutional scale in their model by assigning each frequency band to the convolutional layer corresponding to three different scales of the convolutional kernel. This adoption of a parallel approach to extracting features at multiple scales improves the accuracy of MI EEG signal classification. In [38], the authors proposed a parallel multiscale filter bank convolutional neural network for MI classification. They used time-domain images as input and then used four kinds of different time-scale convolutional kernels for feature extraction to improve the performance, robustness, and migration learning of the model. The most popular application of CNN is to use either time domain signals or frequency domain signals [9] as input. Since EEG models have both continuous and complex variations, it is hardly possible for the model to extract enough features from only frequency or time domain dimension, a multi-input time-frequency CNN structure is proposed in this paper.

As the EEG signal has temporal, frequency, and spatial information, the proposed CNN is divided into two parts in order to extract more comprehensive temporal and spatial features. As shown in Figure 1, the input to the left half of the model is a time-domain image. This section focuses on modifying the size of the convolution kernel of the original model [31]. The first and second convolutional layers of the network are 1-D convolutional layers with kernel size 1×3 and 10×1 , and stride step size 2×1 and 2×1 , which are used to learn the spatial information and temporal features between each channel, respectively. The right half of the model input is the image after CWT mapping. The overall and local features of the EEG signal may be more clearly evident following the wavelet transform of the original signal, allowing the network to extract features more effectively [34]. The two convolutional layers consist of 32 and 64, convolutional kernels size 5×5 , and stride step size 2×2 and 2×2 , respectively. The 2-D feature maps extracted from the left and right parts are expanded into a 1-D vector by a fully connected layer. Then, we concatenate all the 1D features from the two branches into a 1D vector and use the vector as the input of the classifier. In this work, we use the SoftMax function as the classifier. ReLUs are used as the activation function in the proposed model because it increases classification accuracy while speeding up convolutional neural network learning. To prevent overfitting of the neural network, L2 regularization is used. In the training, the L2 regularization parameter is set to 0.01, and the Adam optimizer is employed as the optimization method. The initial

learning rate is set to 0.1, and the learning rate is automatically optimized. Batch size and number of epochs are set to 64 and 100, respectively. The early stop mechanism is also applied in the training phase. The training process is stopped when the validation loss stops decreasing within 5 epochs. This is performed to prevent the model from continuing to fit the training data and instead keep the best-performing weights that generalize well to new data.

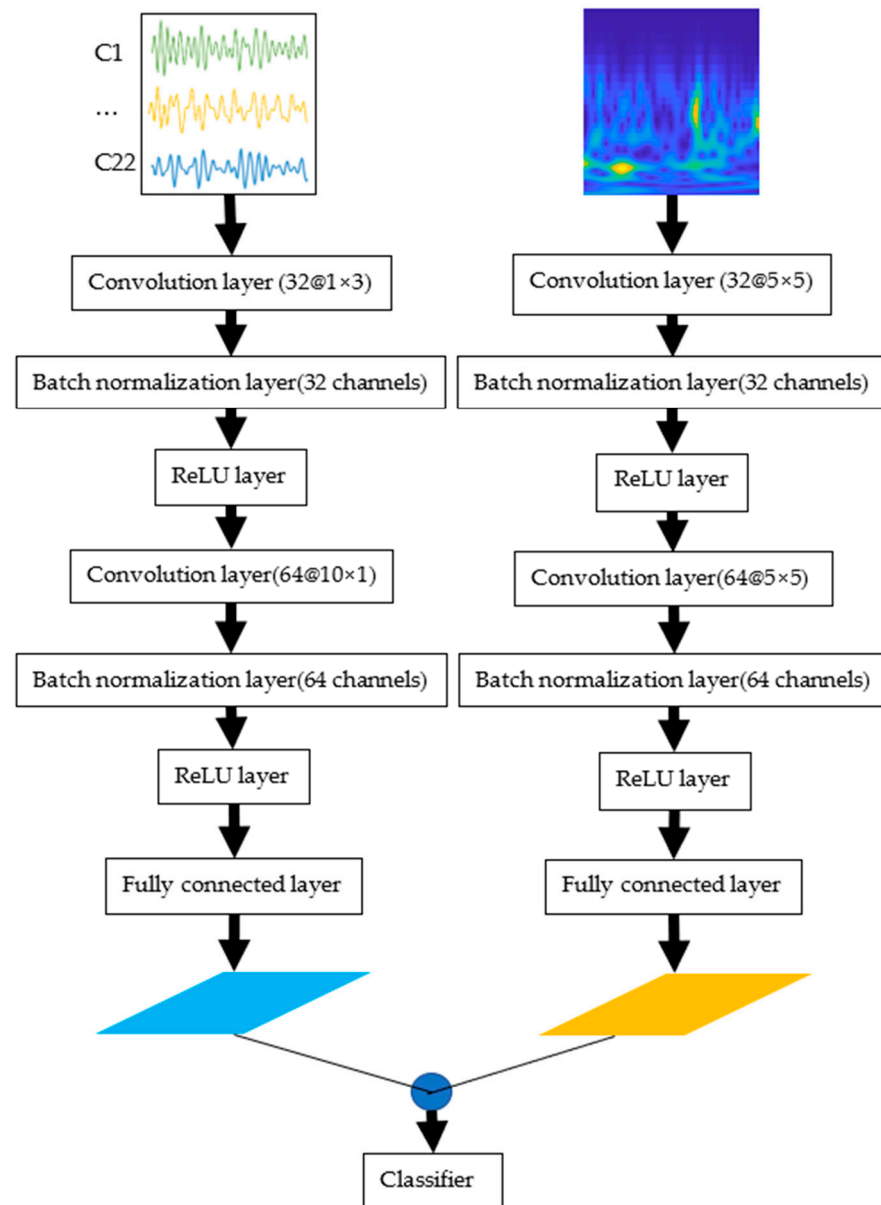


Figure 1. The proposed CNN framework.

CNN's pooling layer can simplify the network parameters, but some useful features might be lost. In [39], a method to classify MI-EEG signals using a simplified CNN was proposed. In this paper, the pooling layer is removed from the standard CNN to optimize the network structure and prevent the loss of effective features.

2.4. Data Augmentation

The EEG signals collected by each electrode are commonly considered as 1-D data, in contrast to the computer vision field, where images are often rotated, cropped, deformed, scaled, and subjected to other types of DA methods that frequently use 2-D data. The

two main categories of the current EEG signal data enhancement techniques are data transformation [18,40,41] and noise addition [16,42].

According to the analysis of the essential features of EEG signals, it can be seen that EEG signals have relatively obvious characteristics in both time and frequency domains. The DA algorithm is designed in the time and frequency domains while maintaining the original characteristics of the EEG signal as much as possible in this work.

The proposed method will be divided into two steps: time domain transformation and frequency domain transformation, as shown in Figures 2 and 3.

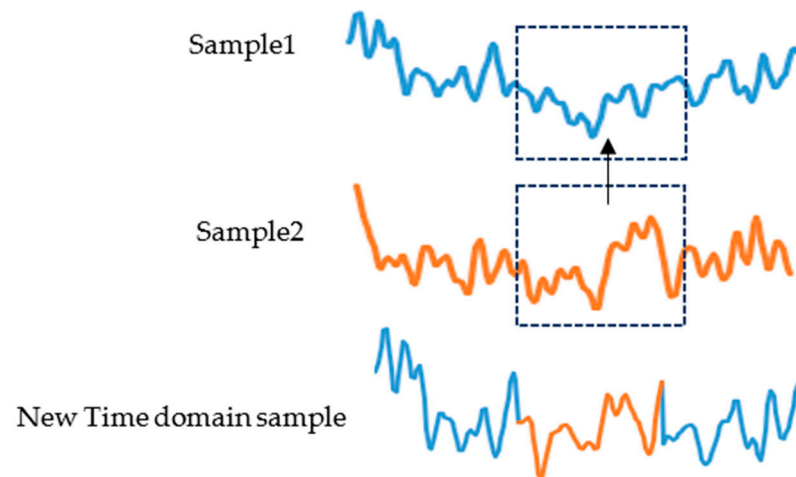


Figure 2. An example of time domain sample generation.

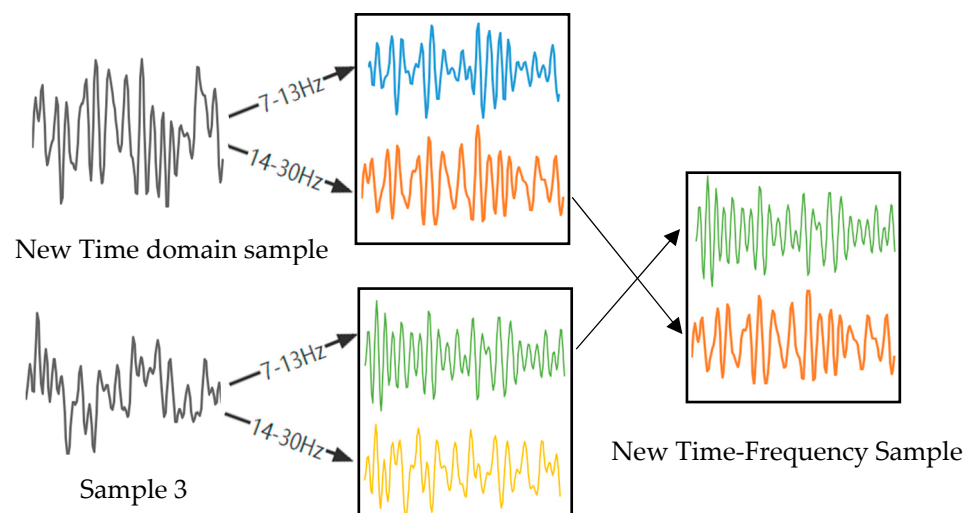


Figure 3. An example of time–frequency domain sample generation.

- Sample 1, sample 2, and sample 3 are three samples of the same class that were randomly selected. We randomly capture a period of 1 s of data from sample 2 to replace the data at the same time position in sample 1. Figure 1 shows a time domain sample generation. Do the same for the test set;
- The artificial time-domain EEG sample and sample 3 are divided into two frequency bands, 7–13 Hz and 14–30 Hz, after band-pass filtering, and then a frequency band of sample 3 is exchanged with the corresponding frequency band of the artificial time-domain EEG sample to reconstruct the time–frequency EEG signal.

3. Results

All the experimental results are obtained using a computer equipped with an Intel Core i5-7300HQ CPU and a GeForce GTX 1050Ti GPU.

3.1. Database

Since 2001, several international BCI competitions have been held to deliver scientists working in this area with reliable data sources and uniform standards of detection algorithms. To test the performance of the proposed methodology, we selected a public data set (2008 BCI Competition IV Data Set 2a [43]) for this purpose.

A diagram of the acquisition process for this data set can be seen in Figure 4. The complete acquisition process of a single EEG signal is divided into three parts. The first 3 s of a trial is the preparation phase. The directional arrow prompts appear on the screen and are accompanied by an audible alert. Subjects need to follow the on-screen prompts to start the motor imagery from the 3rd to the 6th second. After that, the subjects rested and waited for the cue to start the next trial. We intercepted from 3.5 s to the end of the 6th second, considering that the subject might not be able to respond when the cue first appeared, which means that 625 data points were obtained as an original sample in our experiment. The data format for each raw sample is 625×22 , where 625 is 2.5 s of data points (250 Hz) and 22 means there are 22 electrodes.

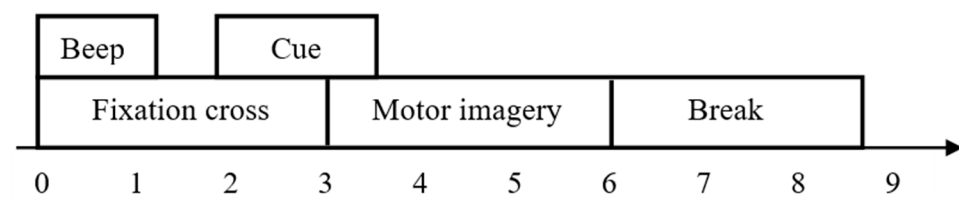


Figure 4. Timing scheme of the paradigm.

Data are collected from each subject in a total of 576 trials. We repeated the proposed DA process 1000 times to obtain 1000 new samples. The final number of samples for each subject increased from the original 576 to 1576.

3.2. Analysis of DA Methods

Figures 5 and 6 show the spectral power comparison plots before and after the signal augmentation, respectively. The red line is the spectrum of the original data, and the blue line is the spectrum after the augmentation. Further, the spectrum of C3, Cz, and C4, which are more closely imagined with the motion, are shown as examples. By comparison, it was found that the characteristics of the original EEG signal were largely retained before and after our augmentation method. The proposed DA method focuses on increasing the energy of the replaced band while keeping the spectral ratio of the signal to its replaced band relatively constant. This process can be used to improve the signal-to-noise ratio and to gain a better explanation of the signal's characteristics.

3.3. Performance Evaluation Metrics

In this section, we introduce several practical metrics to evaluate the performance of the proposed method. The accuracy Acc is calculated as:

$$Acc = \frac{T_p + T_N}{T_p + T_N + F_p + F_N}, Acc \in [0, 1], \quad (10)$$

where T_p , T_N , F_p , and F_N represent true positives, true negatives, false positives, and false negatives, respectively.

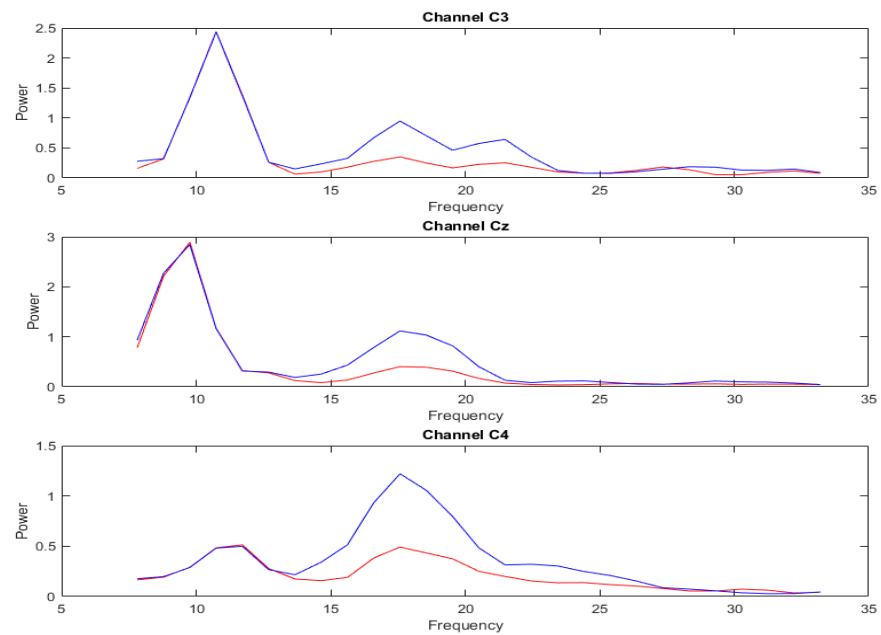


Figure 5. The spectrum of subject 1 imagined the left-hand movement before and after enhancement, in which the 14–32 Hz band of the original signal was replaced. The signal depicted by the red line represents the pristine data, while the signal represented by the blue line corresponds to the output obtained following DA.

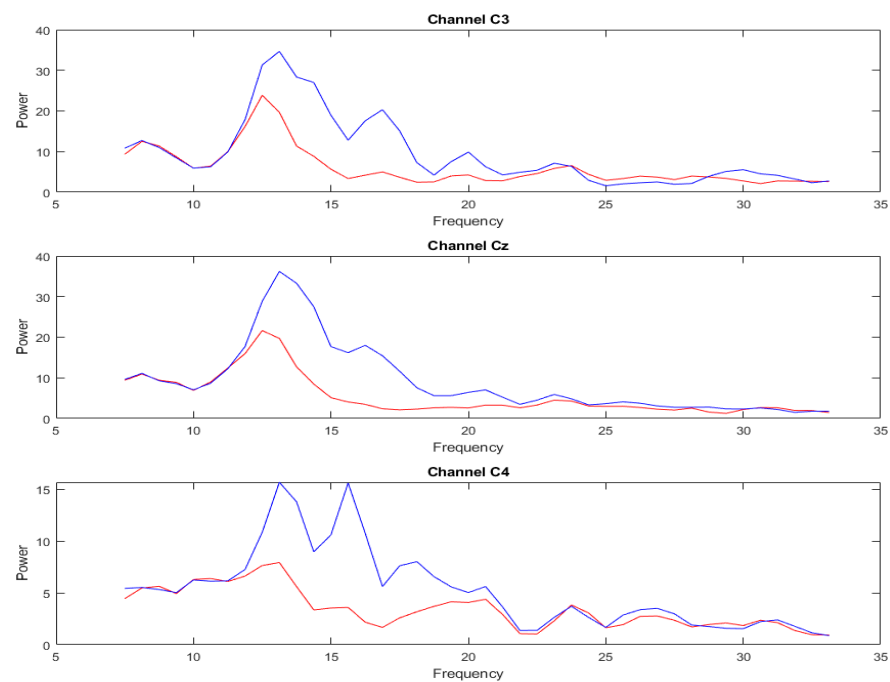


Figure 6. The spectrum of subject 1 imagined the left-hand movement before and after enhancement, in which the 7–14 Hz band of the original signal was replaced. The signal depicted by the red line represents the pristine data, while the signal represented by the blue line corresponds to the output obtained following DA.

The other three important metrics, *Precision*, *Recall*, and *F1 score*, can be expressed as:

$$Precision = \frac{T_p}{T_p + F_p}, \quad (11)$$

$$Recall = \frac{T_p}{T_p + F_N}, a \in R[0, 1] \quad (12)$$

$$F1 = 2 * \frac{Recall + Precision}{Recall * Precision}, a \in R[0, 1] \quad (13)$$

3.4. Determining the Length of Data Segments

Data segmentation and overlap are common ways to increase the amount of data, which can significantly improve classification accuracy [44]. It can be extremely beneficial in terms of reducing resource consumption during network calculational operations. Not only does it decrease the size of the input, but it also reduces the amount of data that needs to be transmitted and stored during the training and inference process, resulting in improved performance and reduced costs. Consequently, these methods can be valuable optimization tools for us. We used a 50% overlap window for the sample, which means that 50% of the data are the same as in the previous segment. The next thing we determined was the scale of the segmentation. Figure 7 shows the accuracy comparison of different size segments, where the 0.5 s window has the worst classification accuracy, and the 2 s and 3 s windows have more similar accuracy. This means that the data of the 2 s fragment already contain a sufficient amount of MI information. To reduce the resources used, subsequent experiments all use a 2 s time window.

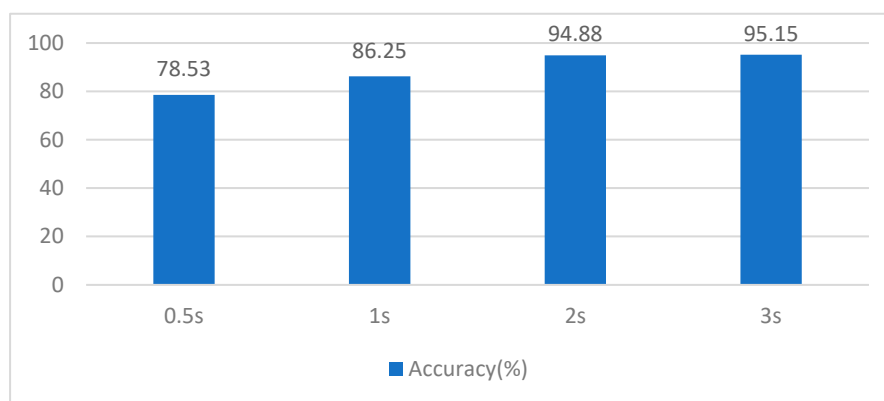


Figure 7. Comparison of different length segments.

3.5. Performance of the Proposed Model

To demonstrate the superiority of the proposed method, we compared our results with some state-of-the-art models. Table 1 introduces the presentation of comparative literature.

Table 1. Contribution to the comparative literature.

Literature	Classifier	Contribution
VGGNet [45]	CNN	- Using VGGNet for MI dataset training. - Better efficiency and accuracy compared to SVM, ANN, and traditional CNN.
EEGNet [26]	CNN	- Introduce EEGNet, a compact convolutional neural network for EEG-based BCIs. - Demonstrate three different approaches to visualize the contents of a trained EEGNet model to enable interpretation of the learned features.

Table 1. Cont.

Literature	Classifier	Contribution
HS-CNN [37]	CNN	- Different convolution kernel scales are designed for EEG signals in different frequency bands, and convolution operations are performed in the temporal and spatial dimensions using one-dimensional convolution kernels. - Identically labeled samples are segmented and then swapped and recombined, then divided by frequency bands, and different segments of the same frequency band are swapped and recombined.
PSO-CNN [46]	CNN	- The signal is input to the CSP filter, while the PSD of different frequency bands is calculated, and the particle swarm optimization algorithm is introduced for feature optimization. - CNN has only one convolutional layer.
CNN-LSTM [47]	CNN+LSTM	- Using the Filter Bank Common Spatial Pattern technique - Hybrid network to extract multimodal features.
DWT-CNN [48]	CNN	- Sliding window for data enhancement, calculating PSD for optimal frequency bands. - Decompose 3×3 convolution kernels into 4×1 or 2×1.

We retrained two high-cited models (VGGNet and EEGNet) to evaluate their performance using the same dataset. The comparison results are shown in Table 2. In addition, research methods with the suffix (_A) indicate the use of DA techniques.

Table 2. Compared results with EEGNet and VGGNet.

Subject No.	Acc (%)					
	VGGNet	VGGNet_A	EEGNet	EEGNet_A	Our	Our_A
1	79.32	85.36	88.21	91.33	97.23	98.61
2	69.01	77.22	70.35	72.58	94.77	98.29
3	91.11	92.67	92.40	93.14	94.12	97.99
4	68.32	66.12	81.26	85.65	95.22	97.1
5	59.00	70.15	85.77	83.85	93.25	96.72
6	70.32	71.39	73.12	76.98	92.36	96.37
7	82.55	85.73	93.10	98.54	94.8	97.69
8	79.33	88.52	90.73	89.32	95.75	97.22
9	70.74	72.29	91.05	89.99	96.38	98.51
Mean	74.41	78.82	85.11	86.82	94.88	97.61

According to the proposed method, the accuracy is improved to 97.61%, the highest among the compared methods. In addition, Our DA method can achieve varying degrees

of accuracy improvement, with average accuracy improvements of up to 4.41% and 11.15% for the individual subject.

The compared results with other articles [37,46–48] are shown in Table 3. It can be seen that our method has a relatively high classification accuracy among the state-of-the-art methods. The proposed method shows a large improvement in the accuracy of subject 2 (up to 18.01% increase) and subject 6 (up to 19.45% increase). Most specifically, compared with Tables 2 and 3, there is no significant difference in classification performance between subjects (the maximum difference between each subject is 5%, which reaches 32% in VGGNet and 12% in DWT-CNN). It indicates that the method in this study overcomes the effect of individual EEG signal variability and can automatically learn the unique EEG signal activity patterns of different subjects in four imaginary states.

Table 3. Compare the table of the proposed method with state-of-the-art methods.

Subject No.	Acc (%)					
	HS-CNN	PSO-CNN	CNN-LSTM	DWT-CNN	Our	Our_A
1	90.07	93.30	98.82	98	97.23	98.61
2	80.28	84.59	98.64	98	94.77	98.29
3	97.08	91.68	96.92	95	94.12	97.99
4	89.66	84.55	96.50	96	95.22	97.1
5	97.04	86.54	92.75	86	93.25	96.72
6	87.04	76.92	91.84	92	92.36	96.37
7	92.14	94.03	95.07	94	94.8	97.69
8	98.51	93.20	95.25	93	95.75	97.22
9	92.31	92.24	99.23	98	96.38	98.51
Mean	91.57	85.56	96.13	94	94.88	97.61

Alternatively, there are various other metrics that can be employed to evaluate the performance of our proposed model. The confusion matrices for the proposed model are given in Figures 8 and 9. Table 4 provides a comprehensive analysis of the precision and recall values of our proposed model on the BCI Competition IV dataset 2a.

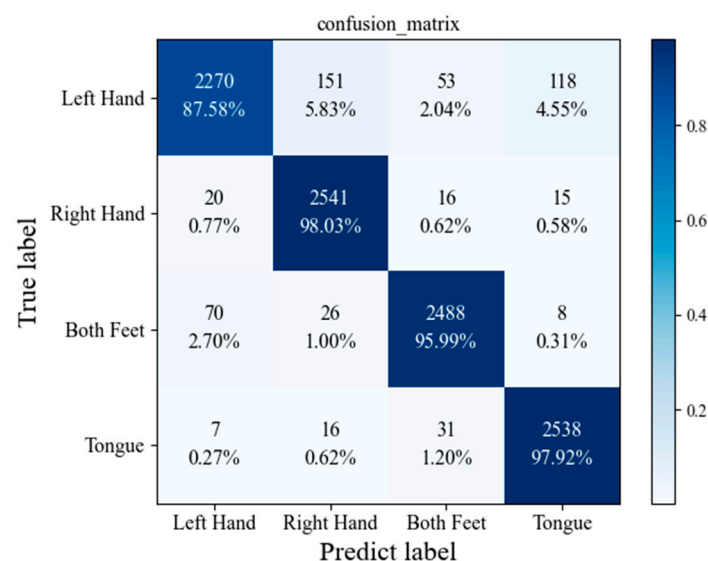


Figure 8. Confusion matrix for the proposed model.

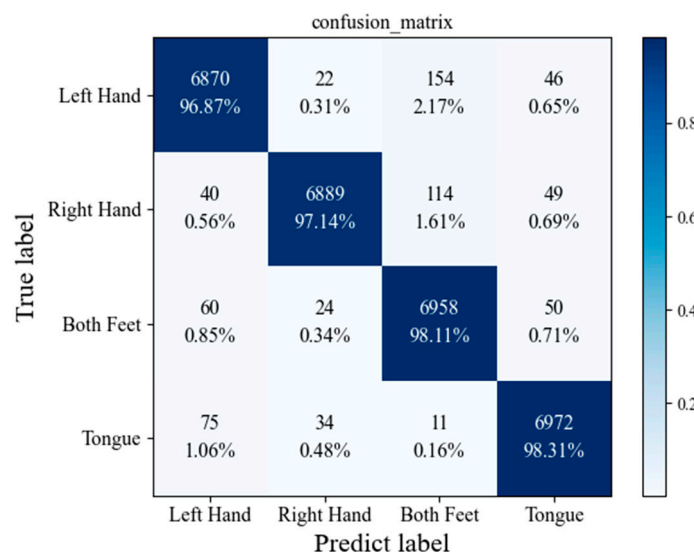


Figure 9. Confusion matrix for the proposed model with DA.

Table 4. Precision and recall scores for the proposed model.

Metrics	Our	Our_A
Precision (%)	94.93	97.62
Recall (%)	94.88	97.61
F1-Score	0.95	0.98

4. Conclusions

As a result of the complexity of EEG signals, the automatic classification accuracy is usually not high and individual differences are significant in the case of limited training data. Another challenge is that existing CNN-based methods for EEG MI classification use a single domain to extract EEG features. This leads to limited classification since it requires simultaneous decoding of the time domain, frequency domain, and spatial information of EEG signals. We propose a hybrid CNN architecture with a data enhancement approach. The accuracy of EEG motor imagery classification is improved by using more different dimensional images as input to the CNN while increasing the amount of data.

The optimal window size may vary from dataset to dataset, so determining the size of the sliding window can improve the accuracy appropriately. For the 2008 BCI competition IV 2a dataset, the window size of 2 s meets our requirements, which not only reduces the consumption of computational resources but also has an acceptable accuracy rate. Our DA method has been proven to be highly effective in improving the accuracy of training VGGNet, EEGNet, and our proposed model. Overall, our proposed MI-EEG image classification method achieves an average accuracy of 97.61%. The improvement of this method is the design of two different scales of CNN for both time domain and CWT mapping maps, which makes the feature extraction more comprehensive. It also has the highest average classification accuracy compared to several other methods. This indicates the potential of our DA technique in helping to improve the performance of machine learning models. We believe this technique has the potential to be applied to other machine learning and MI-EEG analyzing tasks, and we look forward to further exploring its capabilities. In the future, we aim to design a specialized inference accelerator for this model that can be easily integrated into reconfigurable devices such as field programmable gate arrays (FPGAs). However, this goal presents several design challenges, particularly with the parallel CNN architecture. Both the CNN and CWT mapping operations require a significant amount of computational resources, which may result in insufficient memory or slow inference speed for the devices used. To achieve our goal, we need to overcome these

difficulties by finding innovative solutions that allow us to effectively utilize the available resources while ensuring optimal performance.

Author Contributions: Conceptualization, Y.X. and S.O.; methodology, Y.X.; software, Y.X.; validation, Y.X.; formal analysis, Y.X.; investigation, Y.X.; writing—original draft preparation, Y.X.; writing—review and editing, Y.X. and S.O.; supervision, S.O.; funding acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Curran, E.A.; Stokes, M.J. Learning to control brain activity: A review of the production and control of EEG components for driving brain–computer interface (BCI) systems. *Brain Cogn.* **2003**, *51*, 326–336. [[PubMed](#)]
- Zimmermann-Schlatter, A.; Schuster, C.; Puhan, M.A.; Siekierka, E.; Steurer, J. Efficacy of motor imagery in post-stroke rehabilitation: A systematic review. *J. Neuroeng. Rehabil.* **2008**, *5*, 8. [[PubMed](#)]
- Tong, Y.; Pendy Jr, J.T.; Li, W.A.; Du, H.; Zhang, T.; Geng, X.; Ding, Y. Motor imagery-based rehabilitation: Potential neural correlates and clinical application for functional recovery of motor deficits after stroke. *Aging Dis.* **2017**, *8*, 364–371. [[CrossRef](#)] [[PubMed](#)]
- Kumar, V.K.; Chakrapani, M.; Kedambadi, R. Motor imagery training on muscle strength and gait performance in ambulant stroke subjects—A randomized clinical trial. *J. Clin. Diagn. Res. JCDR* **2016**, *10*, YC01. [[PubMed](#)]
- Shih, J.J.; Krusienski, D.J.; Wolpaw, J.R. Brain-computer interfaces in medicine. In *Mayo Clinic Proceedings*; Elsevier: Amsterdam, The Netherlands, 2012; pp. 268–279.
- Ramoser, H.; Müller-Gerking, J.; Pfurtscheller, G. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* **2000**, *8*, 441–446. [[CrossRef](#)]
- Müller-Gerking, J.; Pfurtscheller, G.; Flyvbjerg, H. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin. Neurophysiol.* **1999**, *110*, 787–798.
- Yang, H.; Sakhavi, S.; Ang, K.K.; Guan, C. On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 2620–2623.
- Tabar, Y.R.; Halici, U. A novel deep learning approach for classification of EEG motor imagery signals. *J. Neural Eng.* **2016**, *14*, 016003. [[CrossRef](#)]
- Sakhavi, S.; Guan, C.; Yan, S. Learning temporal information for brain-computer interface using convolutional neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 5619–5629.
- Zhang, K.; Xu, G.; Han, Z.; Ma, K.; Zheng, X.; Chen, L.; Duan, N.; Zhang, S. Data Augmentation for Motor Imagery Signal Classification Based on a Hybrid Neural Network. *Sensors* **2020**, *20*, 4485. [[CrossRef](#)] [[PubMed](#)]
- Yang, B.; Duan, K.; Fan, C.; Hu, C.; Wang, J. Automatic ocular artifacts removal in EEG using deep learning. *Biomed. Signal Process. Control* **2018**, *43*, 148–158. [[CrossRef](#)]
- Lashgari, E.; Liang, D.; Maoz, U. Data augmentation for deep-learning-based electroencephalography. *J. Neurosci. Methods* **2020**, *346*, 108885. [[CrossRef](#)] [[PubMed](#)]
- Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv* **2017**, arXiv:1712.04621.
- Shovon, T.H.; Al Nazi, Z.; Dash, S.; Hossain, M.F. Classification of motor imagery EEG signals with multi-input convolutional neural network by augmenting STFT. In Proceedings of the 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), Dhaka, Bangladesh, 26–28 September 2019; pp. 398–403.
- Wang, F.; Zhong, S.-h.; Peng, J.; Jiang, J.; Liu, Y. Data augmentation for eeg-based emotion recognition with deep convolutional neural networks. In Proceedings of the International Conference on Multimedia Modeling, Bangkok, Thailand, 5–7 February 2018; pp. 82–93.
- Zhang, Z.; Duan, F.; Sole-Casals, J.; Dinares-Ferran, J.; Cichocki, A.; Yang, Z.; Sun, Z. A novel deep learning approach with data augmentation to classify motor imagery signals. *IEEE Access* **2019**, *7*, 15945–15954. [[CrossRef](#)]
- Pei, Y.; Luo, Z.; Yan, Y.; Yan, H.; Jiang, J.; Li, W.; Xie, L.; Yin, E. Data Augmentation: Using Channel-Level Recombination to Improve Classification Performance for Motor Imagery EEG. *Front Hum Neurosci* **2021**, *15*, 645952. [[CrossRef](#)] [[PubMed](#)]
- Luo, Y.; Zhu, L.-Z.; Wan, Z.-Y.; Lu, B.-L. Data augmentation for enhancing EEG-based emotion recognition with deep generative models. *J. Neural Eng.* **2020**, *17*, 056021.

20. Kant, P.; Laskar, S.H.; Hazarika, J.; Mahamune, R. CWT Based transfer learning for motor imagery classification for brain computer interfaces. *J. Neurosci. Methods* **2020**, *345*, 108886. [\[CrossRef\]](#)
21. Lee, H.K.; Choi, Y.-S. A convolution neural networks scheme for classification of motor imagery EEG based on wavelet time-frequency image. In Proceedings of the 2018 International Conference on Information Networking (ICOIN), Chiang Mai, Thailand, 10–12 January 2018; pp. 906–909.
22. Chaudhary, S.; Taran, S.; Bajaj, V.; Sengur, A. Convolutional neural network based approach towards motor imagery tasks EEG signals classification. *IEEE Sens. J.* **2019**, *19*, 4494–4500.
23. Sethi, S.; Upadhyay, R.; Singh, H.S. Stockwell-common spatial pattern technique for motor imagery-based Brain Computer Interface design. *Comput. Electr. Eng.* **2018**, *71*, 492–504. [\[CrossRef\]](#)
24. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. Deep learning-based electroencephalography analysis: A systematic review. *J. Neural Eng.* **2019**, *16*, 051001. [\[CrossRef\]](#)
25. León, J.; Escobar, J.J.; Ortiz, A.; Ortega, J.; González, J.; Martín-Smith, P.; Gan, J.Q.; Damas, M. Deep learning for EEG-based Motor Imagery classification: Accuracy-cost trade-off. *PLoS ONE* **2020**, *15*, e0234178. [\[CrossRef\]](#)
26. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *J. Neural Eng.* **2018**, *15*, 056013. [\[PubMed\]](#)
27. Xie, Y.; Oniga, S. A Review of Processing Methods and Classification Algorithm for EEG Signal. *Carpathian J. Electron. Comput. Eng.* **2020**, *13*, 23–29. [\[CrossRef\]](#)
28. Liu, Y.; Gross, L.; Li, Z.; Li, X.; Fan, X.; Qi, W. Automatic building extraction on high-resolution remote sensing imagery using deep convolutional encoder-decoder with spatial pyramid pooling. *IEEE Access* **2019**, *7*, 128774–128786.
29. Ouyang, Z.; Niu, J.; Liu, Y.; Guizani, M. Deep CNN-based real-time traffic light detector for self-driving vehicles. *IEEE Trans. Mob. Comput.* **2019**, *19*, 300–313.
30. Yu, Y.; Samali, B.; Rashidi, M.; Mohammadi, M.; Nguyen, T.N.; Zhang, G. Vision-based concrete crack detection using a hybrid framework considering noise effect. *J. Build. Eng.* **2022**, *61*, 105246. [\[CrossRef\]](#)
31. Xie, Y.; Oniga, S.; Majoros, T. Comparison of EEG Data Processing Using Feedforward and Convolutional Neural Network. In Proceedings of the Conference on Information Technology and Data Science 2020, Debrecen, Hungary, 6–8 November 2020; pp. 279–289.
32. Lotte, F.; Bougrain, L.; Cichocki, A.; Clerc, M.; Congedo, M.; Rakotomamonjy, A.; Yger, F. A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update. *J. Neural Eng.* **2018**, *15*, 031005.
33. Wang, Y.; Li, Y.; Song, Y.; Rong, X. The influence of the activation function in a convolution neural network model of facial expression recognition. *Appl. Sci.* **2020**, *10*, 1897.
34. Xie, Y.; Majoros, T.; Oniga, S. FPGA-Based Hardware Accelerator on Portable Equipment for EEG Signal Patterns Recognition. *Electronics* **2022**, *11*, 2410. [\[CrossRef\]](#)
35. McFarland, D.J.; Miner, L.A.; Vaughan, T.M.; Wolpaw, J.R. Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topogr.* **2000**, *12*, 177–186. [\[CrossRef\]](#)
36. Shahid, S.; Sinha, R.K.; Prasad, G. Mu and beta rhythm modulations in motor imagery related post-stroke EEG: A study under BCI framework for post-stroke rehabilitation. *Bmc Neurosci.* **2010**, *11*, 1.
37. Dai, G.; Zhou, J.; Huang, J.; Wang, N. HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification. *J. Neural Eng.* **2020**, *17*, 016025. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Wu, H.; Niu, Y.; Li, F.; Li, Y.; Fu, B.; Shi, G.; Dong, M. A parallel multiscale filter bank convolutional neural networks for motor imagery EEG classification. *Front. Neurosci.* **2019**, *13*, 1275. [\[PubMed\]](#)
39. Li, F.; He, F.; Wang, F.; Zhang, D.; Xia, Y.; Li, X. A novel simplified convolutional neural network classification algorithm of motor imagery EEG signals based on deep learning. *Appl. Sci.* **2020**, *10*, 1605.
40. Kim, S.K.; Kirchner, E.A.; Stefes, A.; Kirchner, F. Intrinsic interactive reinforcement learning—Using error-related potentials for real world human-robot interaction. *Sci. Rep.* **2017**, *7*, 17562.
41. Le Guennec, A.; Malinowski, S.; Tavenard, R. Data augmentation for time series classification using convolutional neural networks. In Proceedings of the ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data, Riva del Garda, Italy, 19 September 2016.
42. Li, Y.; Huang, J.; Zhou, H.; Zhong, N. Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks. *Appl. Sci.* **2017**, *7*, 1060. [\[CrossRef\]](#)
43. Tangermann, M.; Müller, K.-R.; Aertsen, A.; Birbaumer, N.; Braun, C.; Brunner, C.; Leeb, R.; Mehring, C.; Miller, K.J.; Mueller-Putz, G. Review of the BCI competition IV. *Front. Neurosci.* **2012**, *6*, 55.
44. Freeman, W.J. Origin, structure, and role of background EEG activity. Part 2. Analytic phase. *Clin. Neurophysiol.* **2004**, *115*, 2089–2107. [\[CrossRef\]](#)
45. Xu, G.; Shen, X.; Chen, S.; Zong, Y.; Zhang, C.; Yue, H.; Liu, M.; Chen, F.; Che, W. A deep transfer convolutional neural network framework for EEG signal classification. *IEEE Access* **2019**, *7*, 112767–112776. [\[CrossRef\]](#)
46. Majidov, I.; Whangbo, T. Efficient classification of motor imagery electroencephalography signals using deep learning methods. *Sensors* **2019**, *19*, 1736. [\[CrossRef\]](#)

47. Zhang, R.; Zong, Q.; Dou, L.; Zhao, X. A novel hybrid deep learning scheme for four-class motor imagery classification. *J. Neural Eng.* **2019**, *16*, 066004.
48. Ma, X.; Wang, D.; Liu, D.; Yang, J. DWT and CNN based multi-class motor imagery electroencephalographic signal recognition. *J. Neural Eng.* **2020**, *17*, 016073. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.