Karadeniz Fen Bilimleri Dergisi, 15(1), 152-170, 2025. DOI: 10.31466/kfbd.1509850

KFBD

Karadeniz Fen Bilimleri Dergisi
The Black Sea Journal of Sciences
ISSN (Online): 2564-7377 https://dergipark.org.tr/tr/pub/kfbd

Araştırma Makalesi / Research Article

A Novel Approach to Motor Imagery EEG Signal Transformation and Classification Using Stockwell Transform and Deep Learning Models

Çağatay Murat YILMAZ[1*]

Abstract

Motor imagery (MI) classification using EEG signals has gained popularity, playing an essential role in developing technologies such as brain-computer interfaces (BCIs). This paper proposes novel approaches using the Stockwell transform (S-transform) to encode signals into images in time-frequency space and classify them by feeding them to pre-trained Inception-ResNet-V2, AlexNet, and SqueezeNet CNNs. High subject-to-subject and session-to-session signal variability hinder the recognition of MI tasks. Most literature has studied within-subject performance. This study conducted experiments using a leave-one-subject-out cross-validation strategy, investigated inter-subject variation's effect and contributed by evaluating the model's performance and generalization ability. At the same time, different sessions and the presence or absence of feedback were assessed, and the results were analyzed. The results are encouraging, considering the difficulty of classifying MI and inter-subject differences. For a cue-based paradigm and non-feedback signals, the results are between 62.1-80.8%; for signals with smiley feedback, the results are between 57.1-96.3%; and for signals with and without feedback are between 56.8-91.4%. These findings highlight the potential of combining the S-transform with CNNs, offering valuable insights into inter-subject variability in EEG-based BCI applications.

Keywords: Stockwell Transform, Convolutional Neural Networks, Transfer Learning, Motor Imagery, Brain-Computer Interfaces, Artificial Intelligence Software.

Stockwell Dönüşümü ve Derin Öğrenme Modelleri Kullanarak Motor Hareket Hayali EEG Sinyallerinin Dönüştürülmesi ve Sınıflandırılması için Yeni Bir Yaklaşım

Öz

EEG sinyalleri kullanılarak motor hareket hayali (MHH) görevlerinin sınıflandırılması, beyin-bilgisayar arayüzleri (BBA) gibi teknolojilerinin gelişiminde önemli rol oynayarak popülerlik kazanmıştır. Bu çalışmada, EEG sinyallerini zaman-frekans uzayında görüntülere kodlamak için Stockwell dönüşümünü kullanan ve görüntüleri önceden eğitilmiş Inception-ResNet-V2, AlexNet ve SqueezeNet evrişimli sinir ağlarına (ESA) vererek sınıflandıran yaklaşımlar önerilmiştir. Denekten-deneğe ve oturumdan-oturuma değişkenliğin fazla olması MHH görevlerinin tanınmasını zorlaştırmaktadır. Literatür çalışmalarının çoğu denek içi performansı incelemiştir. Bu çalışmada ise bir katılımcıyı dışarıda bırak çapraz doğrulama stratejisi kullanılmış, denekler arası MHH varyasyonun etkisi araştırılmış, modellerin performansı ve genelleme yeteneğini değerlendirerek literatüre katkıda bulunulmaya çalışılmıştır. Aynı zamanda farklı oturumlar ve geri besleme olup olmama durumları da değerlendirilmiştir. MHH görevlerini sınıflandırmanın zorluğu ve denekler arası farklılıklar göz önüne alındığında sonuçlar ümit vericidir. İpucu tabanlı gösterim paradigması ve geri bildirimsiz sinyaller için sonuçlar %62,1-%80,8 arasında; gülen yüz geri bildirimi içeren sinyaller için %57,1-%96,3 arasında; geri bildirim içeren ve içermeyen sinyaller için ise %56,8-%91,4 arasındadır. Bu bulgular, MHH görevleri için Stockwell dönüşümü ile ESA'larla birleştirmenin potansiyelini vurgulamakta ve EEG tabanlı BBA uygulamalarında denekler arası değişkenlik hakkında bilgi sunmaktadır.

Anahtar Kelimeler: Stockwell Dönüşümü, Evrişimli Sinir Ağları, Transfer Öğrenme, Motor Hareket Hayali, Beyin Bilgisayar Arayüzleri, Yapay Zeka Yazılımı.

[1]Karadeniz Technical University, Software Engineering Department, Faculty of Engineering, Trabzon, Türkiye, cmyilmaz@ktu.edu.tr

[*]Sorumlu Yazar/Corresponding Author

Geliş/Received: 04.07.2024 **Kabul/Accepted:** 16.12.2024 **Yayın/Published:** 15.03.2025

Yılmaz, Ç.M. (2025). A Novel Approach to Motor Imagery EEG Signal Transformation and Classification Using Stockwell Transform and Deep Learning Models. *Karadeniz Fen Bilimleri Dergisi*, 15(1), 152-170.

1. Introduction

Electroencephalography (EEG) is a neuroimaging method that enables the measurement, recording, and analysis of brain signals. In motor imagery (MI), performed using EEG signals, actions are executed without using muscles and the motor system but only by mentally imagining the actions. For example, when people imagine moving their left arm, the signals recorded from the brain's C3, Cz, and C4 locations can be considered MI-EEG signals, and the imagination of left-hand movement can be associated with these signals. MI has the potential for designing different systems and problem-solving, especially for BCIs, in problems such as wheelchairs, virtual reality environments, and exoskeletons that control limbs. MI-EEG signal classification has been performed using machine learning for decades. They generally include preprocessing, feature extraction, and classification stages. During preprocessing, filtering for noise and artifact removal is performed. In feature extraction, descriptive information is extracted in a summarized form to characterize the signals. In the classification, the features are given as input, and models that can classify the features are built. The success of the models is then measured using the test signals (Yilmaz, 2021). Deep learning (DL) methods have become very popular compared to classical machine learning with automatic feature extraction and classification capabilities. Some studies have used one-dimensional input to deep neural networks (DNNs). Some one-dimensional inputs were created by treating MI-EEG signals as time series, whereas others were built by computing feature vectors. For example, Wang et al. (2024) proposed a novel end-to-end network, a fusion multi-branch one-dimensional convolutional neural network (CNN), to decode MI-EEG signals without pre-processing, addressing the challenges of low signal-to-noise ratio and inter-subject variability. The results obtained in subject-dependent and subject-independent modes are promising. As it is known, noise and signal sources other than EEG can adversely affect the brain signals of interest and negatively affect the classification performance of EEG systems. To address these problems, Ferdi et al. (2024) used one-dimensional CNNs to classify MI tasks such as right/hand, feet, and sedentary. They have shown that BCIs can be built with CNNs for people with disabilities who are deprived of security measures.

Two-dimensional or higher input is another way to feed DNNs with MI-EEG signals. Unlike conventional EEG features, DL approaches use 2D-coded representations as the network input. In addition, preserving the spatial information of the EEG, convolution operations, and feature hierarchies are other essential factors. Therefore, another critical part of the literature feeds DNNs with 2D-encoded inputs (Yilmaz, 2023). Time-frequency methods such as continuous wavelet transform (CWT) and short-time Fourier transform (STFT) are often used for 2D encoding. These methods are advantageous because they can observe the frequency components of different time intervals and how they change with time. For example, Reddy et al. (2024) preprocessed EEG signals using common average reference and Laplace filtering. A sliding window technique was utilized to increase the number of time segments and avoid overfitting. The signals were converted into spectrograms with STFT, the parts associated with mu and beta bands were extracted, and the outputs of the C3, Cz, and C4 channels were fused to form the inputs for the CNN model with self-attention. The results obtained are promising compared to the literature. Generating network inputs with diverse time-frequency transformations is frequently used in the state of the art. However, the number of original methods other than time-frequency transforms is limited. Studies on converting EEG signals into images using various transformations can be discovered in (Yilmaz, 2023).

The S-transform is another time-frequency transformation technique. It has been used in many fields, such as BCIs, for classifying EEG signals. Some of these studies are as follows: Ortiz et al. (2020) designed an MI mental task-based BCI for controlling the lower-limb exoskeleton. They extracted features from each epoch using the S-transform. Although the transformation was applied to all 1-s epochs, a 0.25–0.75 s range was used to avoid border effects. All training and testing experiments were conducted in a real-time environment. The proposed design allows for real-time closed-loop control of the Rex exoskeletons. Qian et al. (2020) employed the S-transform after applying a common spatial pattern (CSP) and gave image features to CNNs. The study was conducted on BCI competition IV dataset I calibration data and used 2-s data after displaying the visual cue. After preprocessing the multichannel signals, spatial filters were calculated with CSP, and four rows of filtered time series were calculated. Then, for each time-series signal, time-frequency information was computed with the S-transform, and the power spectrum of the μ and β bands was extracted and combined by equalizing the spectrum sizes. Finally, the total spectrums for the four-time series were combined to form the final inputs. These images were fed to CNNs for learning and testing and good results were achieved. Alwasiti et al. (2020) attempted to recognize MI tasks using S-transform time-frequency representation and deep metric learning (DML). The goal is to achieve good performance in MI problems with a small amount of limited training data per subject and to reduce the problem of large inter-individual variability. In this method, the signals of each channel were first converted into spectrograms. Then, multi-spectral topographical plots were generated for each trial by combining the images of all channels according to the 10-10 layout. The plotted images were then used in the training/testing of DML. The method showed the best results in time-frequency representation in the 2–78 Hz (y-log scaled) frequency band. The study was conducted on the PhysioNet dataset. The S-transform yielded better results than the STFT. The proposed method could generate successful models even with a small amount of data, approximately 120 trials per subject. Chacon-Murguia et al. (2020) used an 8–28 Hz frequency band and applied the S-transform for a 1 Hz resolution. Then, for each channel, they computed the average of the natural logarithms over the sum of the absolute values. Finally, the features from the C3 and C4 channels were combined and given to the classifiers.

In another method of CSP+ S-transform, CSP filters were used to choose channels, and the transform was used for time-frequency features. Support vector machines (SVM) + CSP obtained more successful results than SVM+ S-transform and SVM+CWT. Salimpour et al. (2022) classified left- and right-hand MI tasks. The approach initially filters the raw signals between 7 and 30 Hz because event-related synchronization and desynchronization patterns are seen in the α and β ranges. Subsequently, the C3/C4 channels were subjected to the S-transform, and absolute time-frequency maps were generated. Finally, the 7–30 Hz time-frequency maps of the C3/C4 channels were vertically combined and fed into the CNNs. Deep feature extraction was performed using two- and three-layer CNNs and pre-trained models, including AlexNet and VGG19. Semi-supervised discriminant analysis was applied to reduce the size of the deep feature vectors. Finally, the k-nearest neighbor (KNN), discriminant analysis, decision tree, random forest, SVM, and ensemble of these classifiers were used for classification. The best results were obtained using CNN-based features from S-transform time-frequency maps, feature selection by semi-supervised discriminant analysis, and SVMs. However, most of these studies did not evaluate the potential of the S-transform for time-frequency analysis of MI-EEG signals. In addition to EEG, the S-transform has been used to classify weak signals such as Electrocorticography (ECoG). Like EEG, these signals contain enormous amounts of data for processing and existing irrelevant features that limit the performance of the classifier systems. Chang and Yang (2018) used S-transform to extract features, Bayesian linear discriminant analysis (LDA) for classification, and a genetic algorithm for feature selection. Almost half of the features were selected and good results were obtained regarding computational cost and performance.

This study aimed to classify MI tasks for BCIs. For this purpose, in the first part of the study, MI signals were encoded as images in time-frequency space with S-transform, and the obtained inputs were given to pre-trained CNNs. No such studies are in the literature, especially using Inception-ResNet-V2 and S-transform. Therefore, it is essential to examine the performance of the S-transform with pre-trained CNNs. The signal processing techniques used in the visualization of MI-EEG signals and their integration with DL models are the most important contributions of the study's first phase. By comparing the performance of different pre-trained CNN models, this study can contribute to evaluating the most efficient methods. It addresses data scarcity problems and training time length in classifying MI-EEG signals using transfer learning techniques. The results were obtained using the leave-one-subject-out cross-validation (LOSO-CV). This LOSO-CV approach also contributes to testing the performance of models in different individuals and observing their validity and generalizability. This approach is critical for understanding the subject characteristics on model performance and BCIs because MI can vary widely among individuals. Results were obtained by considering different sessions with/without feedback, and the contributions of differences were also analyzed. Considering the differences in MI performance between subjects, encouraging results were obtained. In Sec. 2, the dataset is presented. Next, the computation approach of 2D time-frequency images is explained. This section also briefly explains AlexNet, SqueezeNet, and Inception-ResNet-V2 CNNs. Sec. 3 presents the experimental studies, results, comparisons with the literature and discussions. This article's last section outlines the challenges, suggestions for future research, and conclusions.

2. Materials and Methods

2.1. BCI Competition IV Data Set 2b

The BCI Competition IV was designed to process and classify signals from BCIs. Data Set 2b is a clue-based dataset published in this competition and can be accessed at (URL-1). It contains two different challenges. The first is recording signals in the presence of eye artifacts, and the second is the session-to-session transfer using different paradigms (Tangermann et al., 2012). The dataset includes recordings from nine participants, each performing two tasks: MI of the left and right hands. EEG data from the MI-associated C3, Cz, and C4 locations were sampled at 250 Hz. A notch filter was used at 50 Hz, and the signals were bandpass filtered between 0.5-100 Hz. EOG recordings were also collected for artifact processing. The BioSig toolbox was utilized to read data in the GDF format, following the steps provided in Appendix A.2 of (Tangermann et al., 2012).

In the data set, five different sessions were recorded for each subject. The first two sessions were prepared with a cue-based screening paradigm, including non-feedback signals. Each session contains six runs, and ten trials were recorded for each MI task in each run (60 left and 60 right-hand trials in total for one session). The timing paradigms used to record the signals are shown in Figure 1 (a). The other three online feedback sessions were recorded with smiley feedback. The sessions consisted of four runs, each containing 20 trials for each type of MI (80 left and 80 right-hand trials for one session). The timing paradigms used to record the signals are shown in Figure 1 (b). Details of this open access benchmark dataset can be accessed via (URL-1; Tangermann et al., 2012).

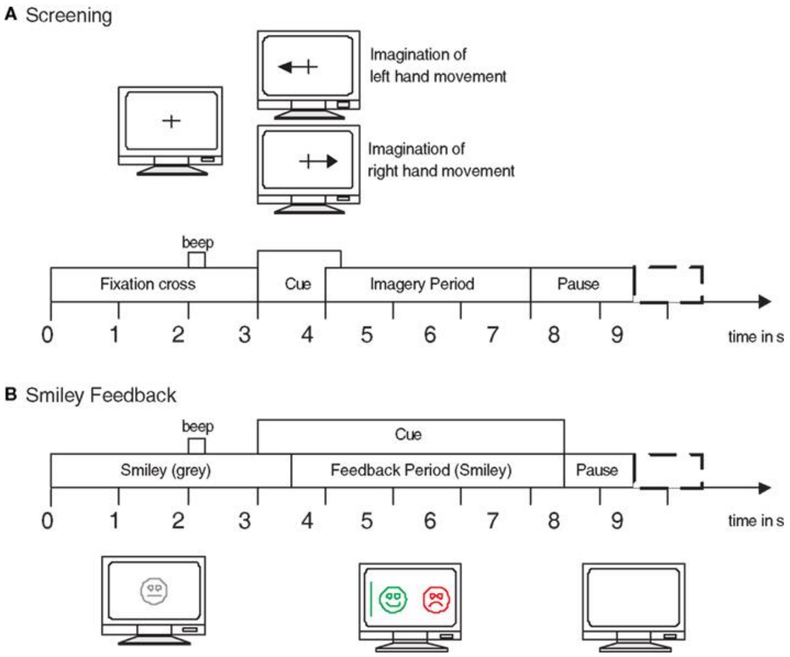


Figure 1. Timing scheme of the paradigm: (a) for the first two sessions and (b) for the last three sessions.

2.2. Stockwell Transform (S-transform)

The S-transform, proposed by R.G. Stockwell et al. (1996), is an extension of the CWT established on the principle of a moving and scalable localizing Gaussian window. Although it directly relates to the Fourier spectrum, it also provides frequency-dependent resolution (Stockwell et al., 1996). It is also similar to the STFT, but it allows the width and height of the analyzing window to vary with frequency, resembling the CWT (Das, et al., 2013). This transform is frequently used in signal processing and data analysis. It is used in the computation of the time-frequency transform of biomedical signals. In this study, the source code given by Sundar (2024) was used to compute the S-transform. The step-by-step procedure for expressing any EEG signal in the time-frequency domain and then transforming it into an image is as follows. Assume a single-channel EEG signal as a time series, denoted by $x(t)$. The continuous time S-Transform $S(\tau, f)$ of this signal is computed as given in Eq. 1, where $w(t)$ denotes the time window centered in $t = 0$ (Zidelmal et al., 2014).

$$S(\tau, f) = \int_{-\infty}^{\infty} x(t)w(t-\tau)\, e^{-i2\pi ft}dt \qquad (1)$$

In this study, a normalized Gaussian window is used as in Eq. 2. In this equation, σ represents the standard deviation and controls the width or spread of the window, determining how much of the signal around a given center t is included in the analysis.

$$w(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} \qquad (2)$$

Since the signal is processed with windows, a Gaussian function with translation τ and dilatation (or window width) σ is used. To obtain a frequency-dependent window width function, σ is written as a function of frequency of $\sigma(f) = 1 \,/\, |f|$ (Zidelmal et al., 2014). The consequent frequency-dependent window function is as follows:

$$w(t-\tau) = \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(t-\tau)^2 f^2}{2}} \tag{3}$$

After adding frequency-dependent window function, the continuous-time S-Transform is derived as in Eq.3. In this equation, τ is the time of the spectral localization, and f is the Fourier frequency. The varying window width with frequency allows for multi-resolution analysis of the signal (Zidelmal et al., 2014). A voice $\mathrm{S}(\tau, \mathrm{f}_0)$ is a one-dimensional function of a time variable τ and a constant frequency f_0 (Das et al., 2013; Zidelmal et al., 2014).

$$S(\tau, f) = \frac{|f|}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t)\, e^{-\frac{(t-\tau)^2 f^2}{2}} e^{-i2\pi f t} dt \tag{4}$$

This study applies the discrete S-transform by utilizing the fast Fourier transform. Consider a discrete-time signal $x[kT]$ generated by sampling $x(t)$ at T intervals for $k = 0, 1, \dots (N-1)$. In the discrete case, the S-transform is the projection of the vector defined onto a spanning set of non-orthogonal vectors. Each basis vector of transform is divided into N localized vectors by an element-by-element product with N-shifted Gaussians (Stockwell et al., 1996; Kalbkhani et al., 2017; Das et al., 2013). The discrete Fourier transform is calculated as in Eq. 5.

$$X(\frac{n}{NT}) = \frac{1}{N} \sum_{k=0}^{N-1} x(kT) e^{\frac{-i2\pi\, nk}{N}} \tag{5}$$

The S-Transform of $x(kT)$ is computed on EEG signals in discrete form as in Eq. 6, where $n = 0, \dots, M - 1$, and M is the length of the transform and $W(m,n) = e^{(-2\pi^2 m^2)/(n^2)}$. $X(.)$ is acquired by shifting the discrete Fourier transform of $x(kT)$ and $(n/NT) \rightarrow f$ and $jT \rightarrow \tau$ (Das et al., 2013; Zidelmal et al., 2014).

$$S(j, \mathrm{n}) = \sum_{m=0}^{N-1} X(m+n)\, W(m,n) e^{\frac{i2\pi\, mj}{N}}; \quad n \neq 0 \tag{6}$$

After the transformation, a $N \times M$ two-dimensional matrix is obtained, where the time component is represented by the columns and the frequency by the rows. The output of the transform is two-dimensional, and it can be used to produce a time-frequency image or spectrogram. The matrix obtained using the transform contains complex values. For this purpose, the study first took the magnitude (using Matlab's abs function) of the matrix obtained by transformation. Then, the complement of this matrix (using Matlab's imcomplement function) was calculated. Thus, the magnitude information of the transformation was represented as images. A sample image acquired by applying the S-transform to a 3-second signal is shown in Figure 2. The S-transform was calculated using the Matlab functions given in (Sundar, 2024).

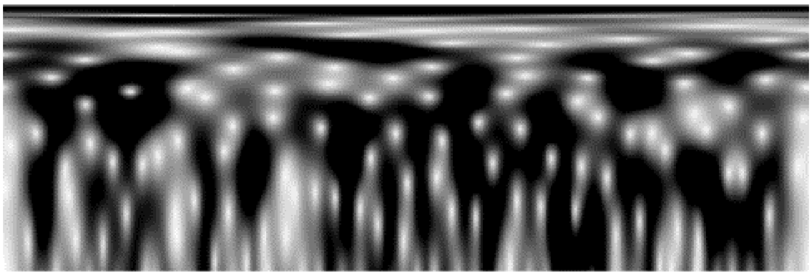


Figure 2. Sample image encoding of EEG signals in 2D time-frequency space using the S-transform.

2.3. Transfer Learning with Convolutional Neural Networks

Due to challenges such as strict requirements for experiments (moving as little as possible to reduce motion artifact contamination, EEG device constantly placed on the head, etc.), inadequate participants, variations from subject to subject, and between sessions, it is challenging to record enough EEG signals to meet the needs of DNNs. Furthermore, data are typically not labeled, which poses difficulties (Yilmaz, 2023). Transfer learning can be utilized to solve such data scarcity problems. Transfer learning transfers attributes and learning models prepared for a particular purpose between problems. For example, in BCIs, personalized features and classifiers must be built for new users. Instead of creating a model from scratch for each new user, models can be built using existing information with transfer learning (Yilmaz, 2021). When transfer learning is used, models can be built with fewer training data than usual. This approach is faster and easier than training CNNs with random initial weight values and layers. However, as the models and the network configuration are built on a separate problem set, the network's hyperparameters for the problem to which the transfer is made must be well-tuned. However, finding the best configurations of hyperparameters, such as the number, type, and order of layers, the size of filters, the number of filters in the convolutional layer, and activation functions, is complex. (Stockwell et al., 1996; Krizhevsky et al., 2017).

This study investigated the performance of S-transform spectrograms with the LOSO strategy in pre-trained CNNs. Therefore, fundamental pre-trained AlexNet, GoogLeNet, and Inception-ResNet-V2 CNNs were used. Each has unique characteristics and architectures, offering distinct advantages for MI. Inception-ResNet-V2 combines the residual connections with the Inception architecture's ability and proposes a costlier hybrid Inception version with significantly enhanced recognition performance. AlexNet, a pioneering DL model, provides a baseline with a simpler architecture, allowing us to examine its performance compared to more complex models. SqueezeNet, known for its lightweight architecture, offers an efficient alternative with a significantly reduced parameter count, which is beneficial in scenarios with limited computational resources. Three models were selected to represent a range of architectures from complex to lightweight, covering a diverse set of feature extraction capabilities and computational costs. A combination of Inception-ResNet-V2 and S-transform has yet to be studied in the literature. Specifically, the spectrograms were fed to pre-trained CNNs. These networks were trained to categorize over a million images from the ImageNet database into 1000 categories and learn rich feature representations for various images.

2.3.1. AlexNet

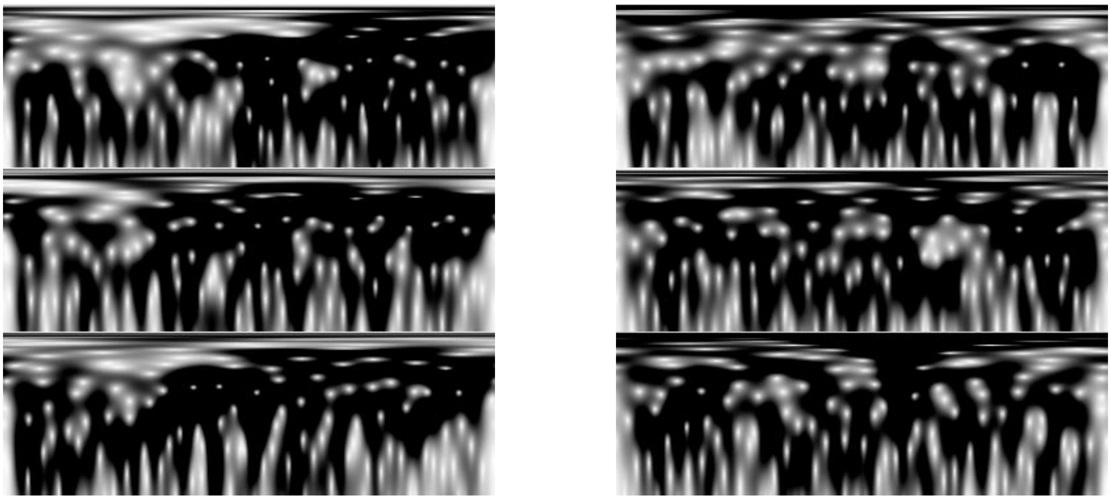
AlexNet contains eight layers: the first five are convolutional, and the last three are fully connected. It uses a regularization technique called dropout to prevent overlearning in fully connected layers. It also offers a very efficient 2D convolution implementation that processes the image’s top and bottom layer parts on different GPUs for faster training. It makes CNNs run faster using a smoothed linear unit function (ReLU) instead of a hyperbolic tangent activation (Stockwell et al., 1996; Krizhevsky et al., 2017). AlexNet was implemented using MATLAB’s neural network toolbox.

2.3.2. SqueezeNet

SqueezeNet is designed to implement CNNs on low-memory hardware such as FPGAs. Despite having 50 times fewer parameters, it performed similarly to AlexNet. Some modifications were made to the network architecture design to achieve acceptable performance with fewer parameters. 1×1 layers replaced most 3×3 convolutional layers with nine times fewer parameters. The number of input channels was reduced to 3×3 filters with squeeze layers, and the number of fully connected layers was reduced. Fire uses a 50% dropout technique after nine modules to avoid overlearning. It uses ReLU as an activation function (Iandola et al., 2016). SqueezeNet was implemented using MATLAB’s neural network toolbox.

2.3.3. Inception-ResNet-V2

Inception architecture achieves very good performance at a relatively low computational cost. Simultaneously, residual connections with more traditional architecture have also shown promising performance. Szegedy et al. (2017) researched the advantage of combining Inception with residual connections and found that training with residual connections significantly accelerates the training of Inception networks. Inception-ResNet-v2 is a costlier hybrid Inception version with significantly enhanced recognition. For example, it achieved better performance with fewer computations than inceptionResNetv1. In this network, residual connections link different-sized convolution filters, avoiding degradation and shortening training (Talukder et al., 2023). InceptionResNetV2 was implemented using MATLAB's neural network toolbox.



(a) (b)

Figure 3. Timing 375×375-sized final spectrogram images for (a) left- and (b) right-hand MI tasks

3. Experiments, Findings and Discussion

This study used 3-s EEG signals between 4 and 7 s. In the first two sessions without feedback, this period refers to the last moments of cue presentation and the time between the start (4 s) and end (7 s). The last three sessions with feedback refer to the MI process in which the cue and feedback were given continuously. Signals were sampled at 250 Hz and contained 750 samples in total. Data reading in gdf format, regression coefficients for correcting EOG artefacts in EEG recordings, and data correction and cutting continuous sequences into segments were performed using the BioSig toolbox. This paper investigated a new method to input MI signals into CNNs. The technique combines spectrograms from different channels to incorporate the information of different electrodes into time-frequency-based image features. This method is based on the technique described briefly in Section 2.2, which is based on computing the S-transform of EEG signals and then transforming the magnitude information into an image. The method is as follows: First, the MI-related parts of each of the signals of channels C3, Cz, and C4 in the range of 4-7 s were extracted. The spectrograms of the signals were computed and converted into images. Then, the images obtained from each channel were scaled to 125×375 using cubic interpolation. Finally, the spectrograms of all channels were vertically merged to preserve the neighborhood, and the final images of size 375×375 were created. These images were used as inputs to the CNNs. Figure 3 shows sample spectrogram images of the left/right-hand MI signals recorded from subject B01 in the first session.

This study first investigated the effect of inter-subject variation. Therefore, the results were obtained using a LOSO-CV. In this strategy, if there are N subjects, N-1 is used for training, and the remaining is for testing. For example, the results for B01 were calculated by treating B02-B08 as training and B01 as test data. This process was repeated with all subjects. This study used Inception-ResNet-V2, SqueezeNet, and AlexNet pre-trained CNNs to classify spectrogram images. These networks were pre-trained using ImageNet. In transfer learning, hyperparameters such as initial learning rate, mini-batch size, number of learning steps, and optimization algorithm should be determined using fine-tuning. If these parameters are well-tuned, new classification tasks can be transferred with less training data, and the network is trained faster and easier than random initial weights. For fine-tuning, the following parameters were used. The maximum number of epochs was set at 50. The training used an early stopping strategy to avoid overfitting, which was half the maximum number of epochs. Initial learning rates were set to 1e-3, e-4, and 1e-5. Adam and stochastic gradient descent (SGD) with momentum optimizers were employed. The methods were implemented on a personal computer with an Intel Core i7-8700 CPU, a Nvidia GeForce GTX 1050 Ti 4 GB GDDR5 graphics card, and 16 GB RAM. It suffered from memory insufficiency at 16 mini-batch sizes when using the Inception-ResNet-V2 network in the Matlab R2022b environment (this problem has not occurred for other CNNs). Due to this, mini-batch sizes of 4 and 8 were used, and the results were calculated for both separately.

The studies were conducted using three different experiments. These experiments are as follows, along with the session information they contain. Expt 1 uses the first two sessions' data recorded without feedback, referred to as 01T and 02T. Expt 2 uses the last three sessions' data recorded with smiley feedback, referred to as 03T, 04E, and 05E). Expt 3 uses all sessions' data with and without feedback, referred to as 01T, 02T, 03T, 04E, and 05E). The evaluations were performed using the three experiments above and LOSO-CV for the nine subjects. The results are shown in Tables 1–3. Table 1 illustrates the results of Expt 1, which uses a cue-based screening paradigm and non-feedback signals. On average, the best results across subjects were 66.1±5.5% with AlexNet, whereas slightly lower results were obtained with SqueezeNet and Inception-ResNet-V2. The most successful results for all subjects were between 62.1% - 80.8%. Low results were observed in almost all subjects except B04. The fact that EEG recordings were taken without feedback and subject-to-subject variation cannot be captured by transfer learning-based CNN approaches can be seen as a possible reason for these outcomes. The results for Expt 2, which uses data recorded with smiley feedback, are shown in Table 2. The average best results across subjects are 76.9±12.7% with AlexNet, whereas slightly lower results were obtained using SqueezeNet and Inception-ResNet-V2. The best results for all subjects were between 57.1% and 96.3%. The classification performance of B02–B03 is relatively low. It may be due to the low model-building performance of the other subjects' information used in training for the test subject or the failure of these subjects to execute MI. The results for subjects other than B02–B03 were good. For subject B04, the classification performance was 96.3%. A possible reason for these results is that acquiring feedback enhances MI performance. Figure 4 illustrates the training and validation accuracy and loss of Expt 2 for subject B04 and the first 4.5k iteration. The results shown in the accuracy and loss graphs are reasonable and consistent with the nature of the LOSO-CV strategy, particularly for MI classification tasks using EEG data. In the accuracy graph, the training accuracy gradually increases and stabilizes around 92%, indicating that the model is effectively learning from the training data. The validation accuracy, while slightly lower than the training accuracy, follows a similar trend, with expected fluctuations due to inter-subject variability. Each subject's EEG signals can vary significantly, which inherently affects the validation performance when different subjects are used as the validation set in LOSO-CV. Similarly, the loss graph shows a consistent decrease in training and validation loss, confirming that the model optimizes effectively. The absence of significant divergence between training and validation performance suggests that the model is balanced and can generalize reasonably well despite the challenges posed by subject-to-subject differences. These observations support the approach's validity and the results' reliability.

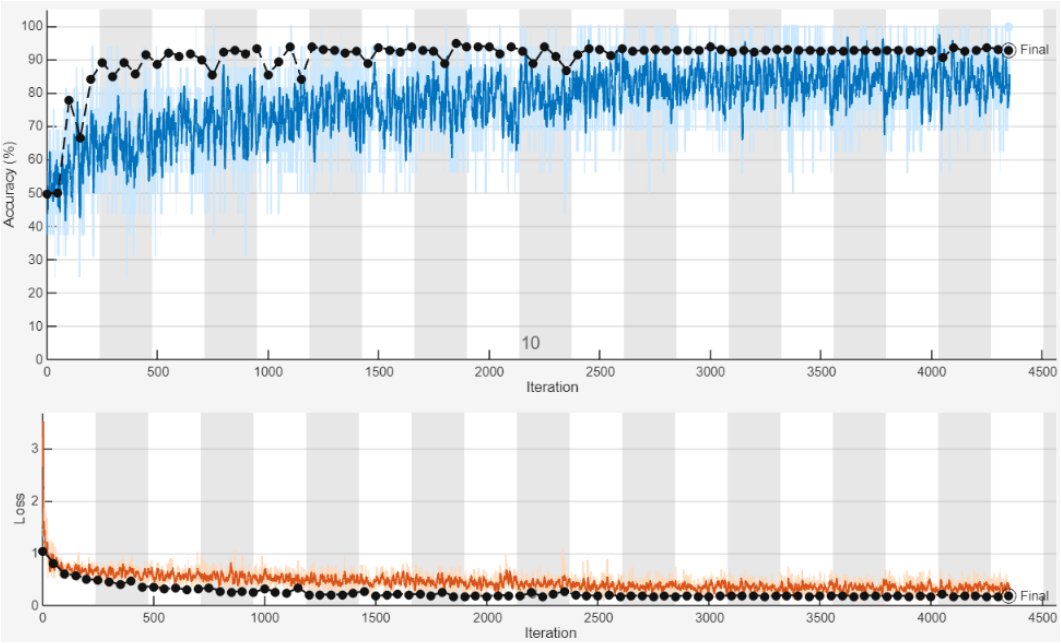


Figure 4. Training and validation accuracy and loss of pre-trained AlexNet for LOSO strategy

The results of the Expt 3 using data from all sessions with and without feedback are given in Table 3. On average, the best results across subjects were obtained using AlexNet with an accuracy of 72.5±10.1%, whereas the other CNNs showed slightly lower results. The best results for all subjects were between 56.8% - 91.4%. The results for B02 and B03 are pretty low. These results may be due to poor information transfer from other subjects or failure to perform the MI tasks. The results from subjects other than B02 and B03 were better, and the success rate for B04 was as high as 91.4%. Combining non-feedback data with feedback data enhanced the performance of the non-feedback system. However, the results were lower than those of the system with feedback.

Table 1. Subject-to-subject results for cue-based screening paradigm without feedback (Expt 1 - first two sessions of Data Set 2b)

Method	B01	B02	B03	B04	B05	B06	B07	B08	B09	Avg±SD
Inception-ResNet-V2	**68.3**	**63.8**	60.0	**80.8**	62.7	**67.1**	**63.3**	61.8	64.6	65.8±5.8
SqueezeNet	65.0	62.1	61.3	72.7	61.5	62.1	60.0	56.8	62.5	62.7±4.1
AlexNet	65.8	60.4	**68.8**	79.2	**66.5**	65.4	59.6	**62.1**	**67.1**	**66.1±5.5**

Table 2. Subject-to-subject results for cue-based screening paradigm with smiley feedback (Expt 2 - last three sessions of Data Set 2b)

Method	B01	B02	B03	B04	B05	B06	B07	B08	B09	Avg±SD
Inception-ResNet-V2	72.3	**58.9**	**57.1**	94.6	81.0	75.6	74.0	**88.3**	86.3	76.5±12.0
SqueezeNet	69.8	58.2	55.0	94.8	78.8	74.4	74.0	85.0	86.5	75.2±12.2
AlexNet	**72.7**	57.3	55.8	**96.3**	**82.1**	**78.5**	**75.4**	86.7	**87.3**	**76.9±12.7**

Table 3. Subject-to-subject results for cue-based screening paradigm with and without feedback (Expt 3 – all five sessions of Data Set 2b)

Method	B01	B02	B03	B04	B05	B06	B07	B08	B09	Avg±SD
Inception-ResNet-V2	71.9	**58.5**	55.8	89.1	74.6	72.9	70.3	**76.2**	78.2	71.9±9.4
SqueezeNet	72.2	56.8	55.4	90.8	74.5	73.1	70.3	75.4	**80.0**	72.1±10.3
AlexNet	**73.3**	56.5	**56.8**	**91.4**	**74.9**	**73.3**	**71.8**	74.9	79.3	**72.5±10.1**

Most of the literature has studied within-subject performance. In these studies, the data for each subject were divided into training, test, and validation sets, and the results were calculated. In this study, the LOSO strategy was used, which is closer to real-time applications. The number of studies conducted using the LOSO is quite limited. Table 4 gives literature methods with feature extraction, DNN input formation, classification methods, and the session information used in the experiments. Zhu et al. (2019) studied training-free MI systems and examined how well information from other subjects can be used in new subjects. A separated channel CNN (SCCN) was suggested to encode multichannel EEG data, and this study employed multichannel series in the CSP space to preserve temporal information. The encoded features are concatenated to classify MI tasks and sent into the recognition network. In addition, results were obtained for the encoder that covers all channels with CNN architecture (OCNN), logistic regression (LR), LDA, SVM, and KNN. In the study, after the results are obtained for each subject using the LOSO, the averages of all subjects are given as the result. SCNN achieved the best average accuracy of 0.64 across all subjects. Furthermore, the baseline methods achieved a maximum accuracy of 0.5, while OCNN achieved 0.62. With the best results, the proposed method enhanced the performance between 2-15%. It outperformed traditional methods regarding learning transfer accuracy, indicating the possibility of free MI interfaces. Compared with SCCN, S-transform with AlexNet achieved 8.5% better results in all sessions. Simultaneously, the implementation of S-transform with Inception-ResNet-V2 and SqueezeNet outperformed SCCN. The results were also better than those of the baseline methods. The S-transform and CNN combinations showed an average standard deviation of 10% for all subjects. The main reason is that the results were obtained using the LOSO strategy. Inter-subject distinctions significantly affect MI tasks, and there are differences between the performance of the feedback and non-feedback sessions.

Table 4. Classification accuracies obtained with LOSO, cross-subject, and MTS strategies in Data Set 2b

Study	Feature extraction and DNN input forming method(s)	Classification Method(s)	Sessions Utilised	Average Acc.
Zhu et al. (2019)	CSP features	KNN	Sessions with and without feedback + LOSO strategy	49%
		LR		49%
		LDA		49%
		SVM		50%
	CSP + Encoding the multi-channel data with SCNN	Encoder that covers all channels with OCNN		62%
		SCCN		64%
Zanini et al. (2017)	Spatial covariance matrices + Riemannian geometry + Symmetric positive definite matrices + Affine transform	Standard minimum distance to mean classifier + Probabilistic classifiers based on a density function defined on the symmetric positive definite manifold	Two sessions without feedback and one session with feedback + cross-subject classification	69.69±9.70%
Luo (2022)	Covariance matrix alignment + FWR-CSP + JPDA	Regularized minimum Mahalanobis distance	Two sessions without feedback and one session with feedback + cross-subject classification	70.14±8.89%
Luo et al. (2023)		DS-KTL	Two sessions without feedback and one session with feedback + MTS strategy	**70.53±9.7%**
Proposed methods	S-transform	Inception-ResNet-V2	Sessions without feedback + LOSO-CV	65.8±5.8%
		SqueezeNet		62.7±4.1%
		AlexNet		66.1±5.5%
	S-transform	Inception-ResNet-V2	Sessions with feedback + LOSO-CV	76.5±12.0%
		SqueezeNet		75.2±12.2%
		AlexNet		**76.9±12.7%**
	S-transform	Inception-ResNet-V2	Sessions with and without feedback + LOSO-CV	71.9±9.4%
		SqueezeNet		72.1±10.3%
		AlexNet		**72.5±10.1%**

Luo et al. (2023) presented a new method called Dual Selections-based Knowledge Transfer Learning (DS-KTL) for cross-subject MI signals. This method selects discriminative features in the source domain while correcting pseudo-labels from the target domain. The first three sessions were employed, and a signal range of 3.5–7 s was used. The study used a multi-source to single-target (MTS) strategy (as in LOSO), which selects one participant as the target and the remaining eight as the source domain for nine subtasks. In Luo et al. (2023), results were also obtained for the methods proposed by Zanini et al. (2017) and Luo (2022), which are successful methods used in the classification of EEG signals. Zanini et al. (2017) introduced a methodology utilizing Riemannian geometry to address cross-session and cross-subject classification in BCIs. To ensure data comparability, they applied an affine transformation to the spatial covariance matrices of the EEG signals from each session/subject. They established an appropriate reference state and proposed a method for online estimating the reference matrix, enabling the technique to be applicable for real-time use. The results were obtained using the minimum distance of the Riemannian means with Riemannian geometry (RA-MDM). Luo (2022) proposed an efficient dual regularization-based MI-EEG feature learning framework for cross-subject tasks. The framework includes feature weighting regularized (FWR) CSP features and joint probability distribution adaptation (JPDA), and minimum Mahalanobis distance. In these studies, DS-KTL showed an average accuracy of 70.53±9.7% on the two sessions without feedback and one with feedback (first three sessions). The results were 59.75% (B02) - 92.75% (B04) range for all subjects. In the same sessions, an average accuracy of 69.69 ± 9.70% was obtained with RA-MDM and 70.14 ± 8.89% with FWR-JPDA across subjects (Luo et al., 2023). Compared with these, in Expt 1, where the first two sessions without feedback were used, RA-MDM, FWR-JPDA, and DS-KTL showed better results across subjects. On the other hand, the standard deviation of these methods is higher. The success of these methods is probably due to their use of feedback session signals, which are more representative. In Expt 3, where all sessions were used, S-transform with AlexNet + S-transform performed 72.5±10.1%, about 2% better than RA-MDM, FWR-JPDA, and DS-KTL on average. The feedback-driven data contributed to this outcome. The other CNNs inputted with S-transform similarly achieved better results. The study also observed that when only feedback data was utilized, the average success across subjects increased to 76.9±12.7%. While low standard deviations were observed in the sessions without feedback, the standard deviations between subjects increased considerably for the signals with feedback.

4. Conclusion and Recommendations

Despite the successes of machine/deep learning techniques in various fields, the desired success and widespread use of MI-EEG systems has yet to be achieved. The most crucial challenges are the high subject-to-subject and session-to-session signal variability and the calibration stage. MI signals differ not only between subjects but also within the same subject. Even MI signals obtained from the same subject on the same day can differ, primarily because of the nonlinear/non-stationary dynamic structure. Therefore, poor transfer learning is a challenging problem. This problem may not be solved even with calibration steps, which can be time-consuming and uncomfortable for users. DNNs can address these high signal variability issues and reduce the time-consuming calibration phase. However, these approaches must overcome challenges such as computational complexity and the necessity of extensive, labelled data.

This paper proposes a new method for inputting MI signals into DNNs. This method combines the spectrograms obtained from the S-transform with different channel data. The classification of MI tasks with spectrogram images was performed using pre-trained Inception-ResNet-V2, SqueezeNet, and AlexNet CNNs. No study exists in the literature where S-transform and these pre-trained CNNs are used together and applied to MI signals. In this part of the study, a new approach was presented, which combined the strengths of the S-transform in the time-frequency domain with pre-trained CNNs. Furthermore, the performances of different CNNs were compared, providing an in-depth analysis of the literature. This part of the study also contributes to the literature on developing MI classification models with small amounts of data using transfer learning. Another noteworthy aim of the study was to investigate the transfer of information among different subjects. Therefore, the results were obtained using the LOSO strategy and compared with literature approaches that used the same strategy. This LOSO-CV strategy is also valuable for evaluating how well the models perform across individuals, thereby assessing their validity and generalizability. This method is essential for analyzing how individual subject characteristics affect model performance and their applications in BCI systems. The results were derived by evaluating sessions with/without feedback, and the impact of these variations on MI-EEG systems was analyzed. Considering the differences in MI performance among subjects, numerous studies have yielded promising outcomes.

The results are promising, considering the MI signal differences between the subjects and sessions. Therefore, it is clear that the S-transform has the potential for MI classification problems. The S-transform is a special case of the CWT that uses a Morlet-type wavelet. It requires minor phase and amplitude adjustments (Gibson et al., 2006). The S-transform is similar to the STFT, but it allows the width and height of the analyzing window to vary with frequency, resembling the CWT (Das, et al., 2013). While maintaining a direct relationship with the Fourier spectrum, it provides frequency-dependent resolution (Stockwell et al., 1996). Therefore, the S-transform can be considered computationally more costly. However, it is thought that there will be a similar average time load across the entire classification system, especially in the field of MI, since it works with short-term, low sampling rates and a small number of signals (very low compared to signals such as audio). For real-time and large systems, steps to reduce the computational cost should be taken either methodologically, as in (Phan, 2024), or in the machine learning operations stages.

The best results for the subjects were obtained using different hyperparameters. It would be better to determine these parameters to cover all subjects. Additional models could further enrich the analysis. Therefore, additional architectures will be included to provide a more comprehensive evaluation. The success of the S-transform will increase even more with CNNs learned from scratch. However, the training set should be enriched by increasing the data using techniques such as sliding windows. The problems become even more challenging, and different problems are added to the pool when real-time subject-independent MI-based BCIs are targeted. Therefore, it is essential to investigate systems that can successfully, reliably, and robustly model MI signals. Many areas still require improvement, from feature extraction to the design of DNNs. With new techniques, if successful results are obtained, BCIs can be created even for subjects without MI data, bypassing the calibration. Therefore, inter-subject studies rather than intra-subject ones should be performed. In addition, machine learning workflows should be automated using technologies such as MLOps.

Acknowledgments

The author would like to thank TUBITAK for providing project hardware resources for this paper's GPU computational operations (Project No: 119E397) and the Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology, for sharing the dataset.

Authors’ Contributions

Çağatay Murat Yılmaz conducted all the research and wrote the manuscript.

Statement of Conflicts of Interest

There is no conflict of interest.

Statement of Research and Publication Ethics

The author declares that this study complies with Research and Publication Ethics.

References

Alwasiti, H., Yusoff, M. Z., Raza, K. (2020). Motor imagery classification for brain computer interface using deep metric learning. IEEE Access, 8, 109949-109963. doi.org/10.1109/ACCESS.2020.3002459

Chacon-Murguia, M. I., Rivas-Posada, E. (2020, July). Feature extraction evaluation for two motor imagery recognition based on common spatial patterns, time-frequency transformations and SVM. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE. https://doi.org/10.1109/IJCNN48605.2020.9206638

Chang, H., Yang, J. (2018). Genetic-based feature selection for efficient motion imaging of a brain–computer interface framework. Journal of Neural Engineering, 15(5), 056020. https://doi.org/10.1088/1741-2552/aad567

Das, M. K., Ari, S. (2013). Analysis of ECG signal denoising method based on S-transform. IRBM, 34(6), 362-370. https://doi.org/10.1016/j.irbm.2013.07.012

Ferdi, A. Y., Ghazli, A. (2024). Authentication with a one-dimensional CNN model using EEG-based brain-computer interface. Computer Methods in Biomechanics and Biomedical Engineering, 1-12. https://doi.org/10.1080/10255842.2024.2355490

Gibson, P., Lamoureux, M. Margrave, G. (2006) Letter to the Editor: Stockwell and Wavelet Transforms. J Fourier Anal Appl 12, 713–721. https://doi.org/10.1007/s00041-006-6087-9

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360. https://arxiv.org/abs/1602.07360

Kalbkhani, H., Shayesteh, M. G. (2017). Stockwell transform for epileptic seizure detection from EEG signals. Biomedical Signal Processing and Control, 38, 108-118. https://doi.org/10.1016/j.bspc.2017.05.008

Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90. https://doi.org/10.1145/3065386

Luo, T. -J. (2022) Dual regularized feature extraction and adaptation for cross-subject motor imagery EEG classification. IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1092-1099) IEEE. https://doi.org/10.1109/BIBM55620.2022.9995282

Luo, T. J. (2023). Dual selections based knowledge transfer learning for cross-subject motor imagery EEG classification. Frontiers in Neuroscience, 17, 1274320. https://doi.org/10.3389/fnins.2023.1274320

Ortiz, M., Ferrero, L., Iáñez, E., Azorín, J. M., Contreras-Vidal, J. L. (2020). Sensory integration in human movement: A new brain-machine interface based on gamma band and attention level for controlling a lower-limb exoskeleton. Frontiers in Bioengineering and Biotechnology, 8, 735. https://doi.org/10.3389/fbioe.2020.00735

Phan, D. T. (2024). Reduce Computational Complexity for Continuous Wavelet Transform in Acoustic Recognition Using Hop Size. arXiv preprint arXiv:2408.14302.

Reddy, A. K. G., Sharma, R. (2024). Enhancing motor imagery classification: a novel CNN with self-attention using local and global features of filtered EEG data. *Connection Science*, *36*(1). https://doi.org/10.1080/09540091.2024.2426812

Qian, L., Feng, Z., Hu, H., & Sun, Y. (2020, October). A novel scheme for classification of motor imagery signal using Stockwell transform of CSP and CNN model. IEEE International Conference on Systems, Man, and Cybernetics (pp. 3673-3677). IEEE. https://doi.org/10.1109/SMC42975.2020.9282917

Salimpour, S., Kalbkhani, H., Seyyedi, S., Solouk, V. (2022). Stockwell transform and semi-supervised feature selection from deep features for classification of BCI signals. Scientific Reports, 12(1), 11773. https://doi.org/10.1038/s41598-022-15813-3

Stockwell, R. G., Mansinha, L., Lowe, R. P. (1996). Localization of the complex spectrum: the S transform. IEEE Transactions on Signal Processing, 44(4), 998-1001. https:/doi.org/10.1109/78.492555

Sundar, A. (2024). Time frequency distribution of a signal using S-transform (Stockwell transform). Retrieved from https://www.mathworks.com/matlabcentral/fileexchange/51808-time-frequency-distribution-of-a-signal-using-s-transform-stockwell-transform. MATLAB Central File Exchange.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1). https://doi.org/10.1609/aaai.v31i1.11231

URL-1: https://www.bbci.de/competition/iv/#datasets, (Date Accessed: 02 July 2024).

Talukder, M. A., Islam, M. M., Uddin, M. A., Akhter, A., Pramanik, M. A. J., Aryal, S., ... & Moni, M. A. (2023). An efficient deep learning model to categorize brain tumor using reconstruction and fine-tuning. Expert systems with applications, 230, 120534. https://doi.org/10.1016/j.eswa.2023.120534

Tangermann, M., Müller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., ... & Blankertz, B. (2012). Review of the BCI competition IV. Frontiers in neuroscience, 6, 55. https://doi.org/10.3389/fnins.2012.00055

Wang, W., Li, B., Wang, H., Wang, X., Qin, Y., Shi, X., Liu, S. (2024) EEG-FMCNN: A fusion multi-branch 1D convolutional neural network for EEG-based motor imagery classification. Med Biol Eng Comput 62, 107–120. https://doi.org/10.1007/s11517-023-02931-x

Yilmaz, C. M., (2021). Classification of EEG-based motor imagery tasks using 2-D features and quasi-probabilistic distribution models, Ph.D. Thesis, Karadeniz Technical University, Graduate Institute of Natural and Applied Sciences, Türkiye, 2021.

Yilmaz, C. M., Hatipoglu Yilmaz, B. (2023). Advancements in image feature-based classification of motor imagery EEG data: A comprehensive review. Traitement du Signal, 40(5). https://doi.org/10.18280/ts.400507

Zhu, X., Li, P., Li, C., Yao, D., Zhang, R., Xu, P. (2019). Separated channel convolutional neural network to realize the training free motor imagery BCI systems. Biomedical Signal Processing and Control, 49, 396-403. https://doi.org/10.1016/j.bspc.2018.12.027

Zanini, P., Congedo, M., Jutten, C., Said, S., Berthoumieu, Y. (2018) Transfer Learning: A Riemannian Geometry Framework with Applications to Brain–Computer Interfaces. IEEE Transactions on Biomedical Engineering, 65 (5), 1107-1116. https://doi.org/10.1109/TBME.2017.2742541

Zidelmal, Z., Amirou, A., Ould-Abdeslam, D., Moukadem, A., Dieterlen, A. (2014). QRS detection using S-Transform and Shannon energy. Computer methods and programs in biomedicine, 116(1), 1-9. https://doi.org/10.1016/j.cmpb.2014.04.008