Natural Resources for Human Health

eISSN: 2583-1194

Original Research

Received 09 March 2026
Revised 12 April 2026
Accepted 25 May 2026

Natural Resources for Human Health 2026, Vol. 6 (3s): 491–507

KEYWORDS:
Deep Learning, EEG (Electroencephalogram), Time-Frequency Representations, Data Augmentation

A Hybrid Deep Learning Approach for Classifying EEG Time-Frequency Representation based on Data Augmentation and Enhancement

Laith Khalid Younis Al-Bahadli[1], Reza Mollaei Karami[2]

[1]Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Technology,
Babol, Iran (E-mail: laith_khalid93@yahoo.com)
[2]Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran (E-mail: mollaeikarami@gmail.com)

ABSTRACT: EEG (Electroencephalography) is a signal that records the electrical activity of the brain. This experimental signal is used to evaluate the electrical activity of the brain, which is used to detect possible problems in communication between brain cells. In addition, EEG can be used for Human-Computer Interaction (HCI), medical diagnostics, and psychological research. Despite its important applications, EEG signals are difficult to process and classify due to their non-stationary nature, high noise, and high variability in recording conditions. In the meantime, EEG Time-Frequency Representations have many advantages for better representation of the signal as input to a classification system based on deep learning models. However, the gap in these methods is the lack of sufficient data for training and also the lack of attention to various features of the EEG signal. To fill these gaps, a hybrid deep learning method is proposed in this research for classifying EEG time-frequency representations based on data augmentation and data improvement. The first innovation of this research is the use of a CNN+LSTM hybrid network. The use of a hybrid network allows for the simultaneous use of both sequence and spatial features for classification. The second innovation is the use of data enhancement techniques using the Efficient Sub-Pixel Convolutional Neural Network (ESPCN) method. The third innovation is the use of data augmentation techniques. Simulation results on the BCI Competition IV dataset show that the three innovations of the proposed method lead to an improvement in the classification process of EEG time-frequency representations, achieving an accuracy of 88.23%.

INTRODUCTION

The human brain is a huge network of billions of nerve cells called neurons [1]. These neurons constantly communicate with each other and produce a weak electrical current. The total electrical activity of these neurons forms the brain's electrical signal, or EEG [2]. EEG is one of the common and painless methods for recording and examining the brain's electrical activity [3, 4]. This test is often performed to diagnose and follow up on diseases such as seizures, epilepsy, migraines, brain infection or inflammation, and other neurological disorders [5]. In an EEG, brain waves are recorded by electrodes placed on the head and displayed as a graph so that the doctor can assess the health of the brain [6]. This method does not pose any risk to the patient or pregnant women and can help in more accurate diagnosis and selection of appropriate treatment [7]. In short, this signal can be used for human-computer interaction, medical diagnostics [8] and psychological research as shown in **Figure 1** [9].

EEG signals are difficult to process and classify due to their non-stationary nature, high noise and high variability in recording conditions. Therefore, extracting features and identifying patterns related to mental and motor activities is an important process for preparing the signal for subsequent applications [10]. This processing leads to providing a suitable representation for the classification section and subsequent high-level operations [11]. The EEG signal is a function of time in its raw form, but it can be represented as a time-frequency function with transformations. For example, this function can be transferred to the frequency domain using the Fourier transform. In the frequency domain, the signal is separated into different frequency components and the distribution of signal energy in different frequency bands can be examined. This helps in better understanding brain activity and detecting hidden patterns in the EEG signal.

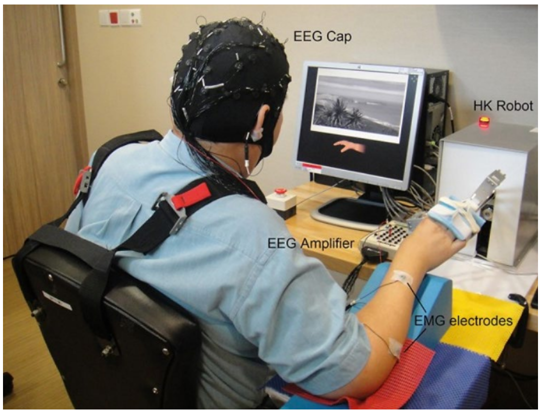


Figure 1 EEG-based stroke rehabilitation in an HCI framework [8]

There are various methods to achieve this goal. In the meantime, EEG time-frequency representation has many advantages. The aim of this research is to first investigate the methods of EEG time-frequency representation. After this transformation, deep learning-based classifiers will be used. The importance of the problem of feature extraction and finding the appropriate EEG representation is because this process is important in the classification of these signals and plays a key role in improving the performance of assistive, diagnostic, and rehabilitation systems [12].

Deep learning is an excellent solution for classifying EEG time-frequency representations [13, 14]. However, the gap in these methods is the lack of sufficient data for training and also the lack of attention to the various features of the EEG signal. To fill these gaps, in this research, a hybrid deep learning method for classifying EEG time-frequency representations based on data augmentation [15] and data improvement is presented. The first innovation of this research is the use of a CNN+LSTM hybrid network. The use of a hybrid network makes it possible to use both sequence features and spatial features for classification at the same time. The second innovation is the use of data enhancement techniques using the Efficient Subpixel Convolutional Neural Network (ESPCN) method. The third innovation is the use of data augmentation techniques. Data augmentation leads to improvement of the learning process. The use of this framework based on three innovations makes it possible to increase the accuracy and speed of classification of EEG time-frequency representations. Therefore, the three contributions of this research are as follows:

- The first innovation is the use of a CNN+LSTM hybrid network. The use of a hybrid network allows both sequence and spatial features to be used for classification at the same time.
- The second innovation is the use of a data enhancement technique using the ESPCN method. Data enhancement leads to an improvement in the learning process and ultimately increases classification accuracy.
- The third innovation is the use of a data augmentation technique. Data augmentation leads to an improvement in the learning process and ultimately increases classification accuracy.

The rest of the paper is as follows. In the second section, a literature review is conducted. In the third section, the proposed method for classifying the time-frequency representation of the EEG signal is presented, and the numerical simulation results are presented in the fourth section.

LITERATURE REVIEW

Various studies have been conducted in the field of preprocessing for EEG classification. For example, a comprehensive study in this area has been conducted in [11]. In this study, three main transformation methods were used as input for classification. Also, two different CNN models were used in the classification stage. Two designs were also used for the dataset. First, a synthetic dataset and then a real dataset was used to train and test the performance of the models. One of the important advantages of this study method is the examination of several transformation methods, several classification methods and datasets. However, there are several disadvantages. First, no preprocessing was used. Noise reduction techniques and adaptive filters can lead to increased accuracy. The accuracy of the method is very low. For some transformations, the accuracy is about 57%, which is not a good percentage. The neural network used is single and only suitable for image data. While RNN and LSTM networks can also examine the temporal dimension of the EEG signal in a hybrid structure. The diversity of the signals examined is also low and belongs to a specific group.

In [16], a study titled "Deep Time-Frequency Domain Convolutional Neural Network for Classification of Focal and Non-Focal EEG Signals" was conducted. [16] presents a new method for automatic classification of focal and non-focal EEG signals. This method is based on the use of synchronization transform (SST) and deep convolutional neural network for classification. The time-frequency matrices of EEG signals are evaluated using Fourier SST (FSST) and wavelet SST (WSST). A two-dimensional (2D) deep CNN is used for classification using the time-frequency matrix of EEG signals. The FSST module is located at the beginning of this architecture. The main disadvantage of FSST is that the time accuracy and frequency accuracy cannot be high at the same time. In [17], an image transformation method for EEG signals was presented. The innovation of [17] consists in investigating several new approaches for preprocessing brainwave signal data, in which statistical features are extracted and then formatted as visual images based on the order in which dimensionality reduction algorithms select them. These data are then considered as visual input for 2D and 3D CNNs, which then extract "features of features". The problem of this study is that the functions used in [17] are not real-time and are suitable for specific applications.

[18] proposed a new automatic EEG signal recognition method to help in the diagnosis of epilepsy. Specifically, the multiwavelet basis function expansion (MWBF) method is first adopted to build a time-varying autocorrelation (TVAR) model of EEG signals, and then a robust hyper-orthogonal least squares (UOLS) algorithm is used to detect the model structure with the help of information derived from EEG signals. In addition, the power spectral density (PSD) estimation method is applied to extract high-resolution time-frequency features. In particular, to fully exploit the spatio-temporal information of the extracted features, the features were fed to spike neural networks (SNN) for classification. It should be noted that although spike networks perform well, they do not seem to be suitable for use with transformed EEG signals.

[19] shows that in all cases, classification of raw data leads to superior performance compared to spectral EEG features. Interestingly, [19] shows that the neural network tailored to process EEG spectral features has increased performance when applied to classification of raw data. [19] shows that the same convolutional networks used to process EEG spectral features perform better when applied to raw EEG data. Although the results presented in [19] are impressive, the results presented in this paper cannot be generalized to all cases.

The transformation of one-dimensional (1D) biomedical signals into two-dimensional (2D) images allows the use of CNNs for classification tasks. [20] investigated the effectiveness of different 1D-to-2D transformation methods for classifying ECG and EEG signals. For this purpose, [20] selected five transformation methods: CWT, FFT, STFT, signal transformation (SR), and recurrence plots (RP). One of the weaknesses of the method [20] is the use of a single network to evaluate the results. However, the use of five transformation models requires a more comprehensive study. In [21], SPNN and FFT method were also used for time-frequency analysis of EEG signal. As it is clear, one of the main innovations of this method is in the coding stage, where the pulse density demodulation (PDM) method is used. There are several other studies [15, 22-25] in this field that have used similar ideas.

In general, previous studies face gaps such as lack of proper preprocessing, low accuracy in some methods, use of single neural networks, and insufficient attention to the temporal nature of the EEG signal. Also, limited data diversity and neglect of data augmentation and improvement methods are other important weaknesses of these studies, which can reduce the performance of deep learning models. To fill these gaps, in this research, a new method for classifying EEG time-frequency representations based on data augmentation and improvement is proposed.

Natural Resources for Human Health
eISSN: 2583-1194

PROPOSED METHOD

1-1 Flowchart of the Proposed Method

The flowchart of the proposed method is shown in **Figure 2**. The raw EEG signal contains various noises such as environmental electrical noise and noise caused by muscle activity. EEG signal preprocessing refers to a set of techniques used to remove the initial noise and normalize the signal. Therefore, there is a preprocessing step after data collection in the general framework. Then, the transformation is performed to obtain the spectrum of the EEG signal. Using spectral analysis, the power of the EEG signal in different frequency bands can be obtained. The signal strength indicates the amount of electrical activity in that frequency band. Spectral analysis helps to examine changes in brain activity over time and identify specific patterns in the EEG signal.

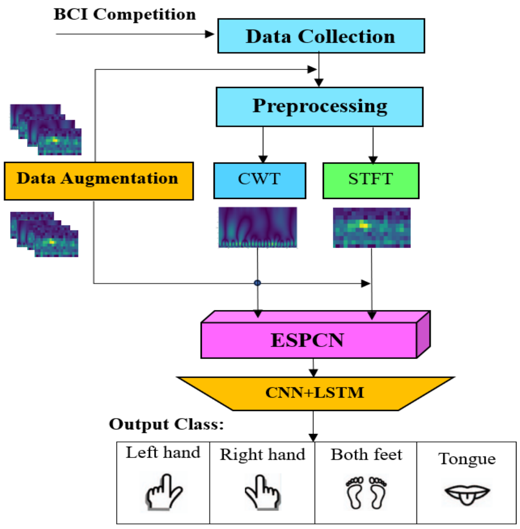


Figure 2 Flowchart of the proposed method

In this research, the short-time Fourier transform (STFT) and continuous wavelet transform (CWT) methods are used to transform the signal. The reason for using two different methods is that their combination leads to increased classification accuracy. STFT is one of the oldest and most widely used methods for analyzing non-stationary signals. In STFT, the signal is divided into short time segments. Then, the Fourier transform is applied to each segment. The result of this process is a time-frequency spectrum (spectrogram) that shows which frequencies are dominant in each time interval. The important advantage of STFT is its simplicity and computational speed. The main disadvantage of STFT is that the time accuracy and frequency accuracy cannot be high at the same time. For this reason, CWT was also used. CWT is a more flexible method than STFT. The most important advantage of CWT is its ability to display local details of the signal. In this method, instead of using a fixed window, a wavelet function with different scales is used to analyze the signal. This feature creates good time accuracy at high frequencies and better frequency accuracy at low frequencies. That is, two important features are created in different frequency ranges. The output of this method is displayed as a scalogram, which is an image.

In the proposed method, four techniques will be used to increase data: time shifting and adding noise (before converting EEG to image) and frequency rotation and masking (after converting EEG to image). The combination of these four techniques is one of the most effective methods for increasing data in the classifier training process. In this method, time shifting simulates natural changes in the time of occurrence of brain patterns and makes the model resistant to the exact location of the patterns. Adding noise simulates the real conditions of EEG recording and makes the model resistant to environmental disturbances and interfering signals. Rotation creates spatial variation in the image and the latent patterns in it are better learned. Finally, frequency masking on parts of the image in the frequency axis causes the model to reduce its dependence on a specific region of the spectrum and learn more general features. This combination increases data diversity, improves model generalization, and reduces overfitting, without fundamentally changing the physiological structure of the signal. ESPCN will also be used to improve data quality. ESPCN is a lightweight and very fast superresolution method that uses pixel shuffling instead of complex layers to improve image quality without significantly increasing the computational load. This feature makes it ideal for time-

eISSN: 2583-1194

frequency EEG images, as it enhances clarity and can also be used for real-time processing or low-resource systems. Then, in the last step, a hybrid deep learning classifier consisting of LSTM and CNN networks determines the output class and evaluates the results. Using a hybrid network makes it possible to use both sequence and spatial features for classification at the same time.

1-1-1 Data Collection

Data collection is an important part of studies related to EEG signal classification. In this study, the BCI Competition IV – Dataset 2a dataset was used, which has 22 channels. Of course, not all 22 channels were used in this study, because in this case the data would have been too large and difficult to work with. Therefore, instead of using all channel sensors, only 8 sensors were selected. This not only reduces noise and reduces the complexity of the model but also allows for a better focus on the motor area.

The right side of the body can be measured with channels C3, CP3, and FC3. For the left side, there are also channels C4, CP4, and FC4. The Cz, CPz channels are also used for analyzing Tongue-Foot movements. The channels used in the proposed method are shown in **Figure 3.**

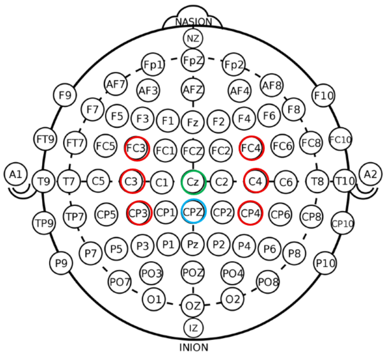


Figure 3 Channels used in the proposed method

1-1-2 Preprocessing

The raw EEG signal may be accompanied by unwanted components such as sounds and waves from the energy source, environmental electrical noise and noise from muscle activity, eye blinking, and heart rate, which is inevitable. In other words, an artifact is anything that is recorded in the EEG but does not originate from the brain. These unwanted components affect the EEG analysis and provide incorrect information, and therefore, EEG signal preprocessing is required to eliminate them. In the proposed method, EEG signal preprocessing refers to a set of techniques that are used to remove elementary noise and normalize the signal. According to the type of problem, a Bandpass Filter between frequencies of 8 and 30 Hz has been used in this research. A 50 Hz notch filter has also been used to remove city power noise.

1-1-3 STFT and CWT Transforms

The input to the deep hybrid model is a two-dimensional image. This image is obtained from the transformation of the EEG signal, which shows the spectrum of the signal. Using spectral analysis, the power of the EEG signal in different frequency bands can be obtained. The signal strength indicates the amount of electrical activity in that frequency band. Spectral analysis helps to examine changes in brain activity over time and identify specific patterns in the EEG signal. In this research, the Short-Time Fourier Transform (STFT) and Continuous Wavelet Transform (CWT) methods were used to transform the signal. The reason for using two different methods is that their combination leads to increased classification accuracy.

The result of this process is a time-frequency spectrum (spectrogram) that shows which frequencies were dominant in each time interval. The important advantage of STFT is its simplicity and computational speed. The main disadvantage of STFT is that both time and frequency accuracy cannot be high at the same time. For this reason, CWT is also used. STFT transform is defined as follows:

$$X_{stft}(a,b) = \int_{-\infty}^{\infty} x(\tau)\, w(\tau - b) e^{-j2\pi a\tau} d\tau \quad (1)$$

In this relation, $x(t)$ is the EEG signal, $w(\tau - b)$ is the window function, a is the frequency scaling parameter, b is the time scaling parameter, and $X_{stft}(a,b)$ is the output transform. To represent this transform as a Spectrogram, we have:

$$Spectrogram\ (a,b) = \left|X_{stft}(a,b)\right|^2 \quad (2)$$

CWT is a more flexible method than STFT. In this method, instead of using a fixed window, a wavelet function with different scales is used to analyze the signal. This feature allows for good time accuracy at high frequencies and better frequency accuracy at low frequencies. That is, two important features are created in different frequency ranges. The output of this method is displayed as a scalogram. The most important advantage of CWT as a transformation method is its ability to display local details of the signal. The CWT transform is defined as follows:

$$X_{cwt}(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\, \psi^* \left(\frac{t-b}{a}\right) dt \quad (3)$$

In this relation, $x(t)$ is the EEG signal, ψ is the parent wavelet, a is the frequency scaling parameter, b is the time scaling parameter, and $X_{cwt}(a,b)$ is the output transform. To represent it as a Scalogram, we also have:

$$Scalogram\ (a,b) = |X_{cwt}(a,b)|^2 \quad (4)$$

The output of each transformation is expressed as a two-dimensional image with dimensions $F \times T$, so that we have:

$$F \times T \equiv X\,(a,b) \quad (5)$$

In this relation, F is the number of frequency samples and T is the number of time samples, where the intersection of each of these samples is a pixel that determines the energy of the signal at a specific time at a specific frequency. For example, these dimensions can be expressed as 64×128, where the number of time samples is 128 and the number of frequency samples is 64.

1-1-4 Data Augmentation

Four techniques have been used to augment data: time shifting and adding Gaussian noise (before converting EEG to image) and frequency rotation and masking (after converting EEG to image). The combination of these four techniques is one of the most effective methods for augmenting data in the classifier training process.

Time shifting simulates natural changes in the time of occurrence of brain patterns and makes the model robust to the exact position of the patterns. In this case, if the initial signal is defined as $EEG\,(t)$, then the new signal $EEG_N(t)$ is generated by a shift of τ milliseconds from the initial signal. It should be noted that the $EEG\ (t)$ signal is non-stationary and hence time shifting creates appropriate variation in the data augmentation process. The formula for this technique is expressed as follows:

$$EEG_N(t) = EEG(t + \tau) \quad (6)$$

Also, as a second technique, adding noise simulates the real conditions of EEG recording and makes the model robust against environmental disturbances and interfering signals. The formula for this technique is as follows:

$$EEG_N(t) = EEG(t) + \frac{N(0,1)}{S} \quad (7)$$

The above two techniques were at the signal level. At the image level, two techniques are used: rotation and frequency masking. The rotation technique for increasing data leads to more spatial variation in the data. It should be noted that this technique preserves the original frequency-time structure despite creating spatial variation for better understanding of latent patterns. The formula for this technique is as follows:

$$\begin{bmatrix} a_N \\ b_N \end{bmatrix} = M_D \left(\begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} \frac{T}{2} \\ \frac{F}{2} \end{bmatrix} \right) + \begin{bmatrix} \frac{T}{2} \\ \frac{F}{2} \end{bmatrix} \quad (8)$$

Where the rotation matrix is defined as follows:

$$M_D = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \tag{9}$$

By means of this technique, each point $\begin{bmatrix} a \\ b \end{bmatrix}$ is converted to a point $\begin{bmatrix} a_N \\ b_N \end{bmatrix}$ where $\begin{bmatrix} \frac{T}{2} \\ \frac{F}{2} \end{bmatrix}$ is the center of the image.

The last technique is frequency masking. In this technique, parts of the frequencies are masked or so-called filtered. This feature allows the model to receive diverse training in the learning process. The formula for this technique is as follows:

$$X_N(a,b) = \begin{cases} 0 & if\ b \in BW \\ X_N(a,b) & if \quad o.w \end{cases} \tag{10}$$

where BW is the desired frequency interval in the masking process. In summary, combining data augmentation techniques increases data diversity, improves model generalization, and reduces overfitting, without fundamentally changing the physiological structure of the signal.

1-1-5 Data Enhancement with ESPCN

The proposed method also uses the ESPCN method to improve data quality. ESPCN is a lightweight and very fast super-resolution method that improves image quality without a large increase in computational volume by using pixel shuffling instead of complex layers. This feature makes ESPCN ideal for time-frequency EEG images, as it enhances resolution and can be used for real-time processing or low-resource systems.

1-1-6 CNN+LSTM Hybrid Network

Then, in the last step, a hybrid deep learning classifier consisting of LSTM and CNN networks determines the output class and evaluates the results. Using a hybrid network allows both sequence and spatial features to be used for classification at the same time. The complete network architecture is shown in **Figure 4**.

The input to this hybrid network is modeled as a 3D tensor, where F is the number of frequency samples, T is the number of time samples, and C is the number of EEG channels. In fact, in this case, the transformed data from 8 channels are stacked together.

$$F \times T \times C \tag{11}$$

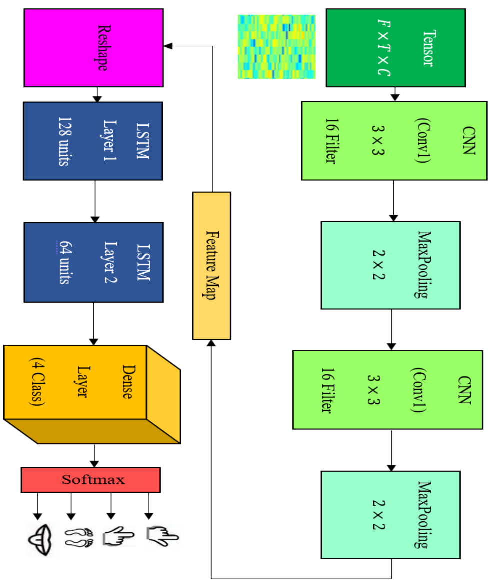


Figure 4 Hybrid Network Architecture

The first step of a CNN network is located. This network extracts local features in the time-frequency image. In this network, each of the 16 filters seeks to extract patterns such as local patterns in EEG images, energy changes, and frequency lines. Then, 16 feature maps are created. The next step is a 2×2 Max-Pooling step. The goal of this layer is to reduce the dimensionality while preserving the most important information.

Then there is another step of the convolutional layer structure and 2×2 Max-Pooling to further reduce the dimensionality. Of course, this dimensionality reduction should not be so great that important information is lost. Finally, a feature map is created that contains the important features extracted from the input tensor (including the C channel transformed EEG signal).

The goal of feature extraction is to represent the raw EEG signals by values with fewer and more relevant dimensions so that classifiers can identify which features represent each class. The question that arises in this section is why the raw EEG data or its image data are not used directly. The problem at hand here is the curse-of-dimensionality problem. The concept of this problem is that the number of samples for training the classifier increases exponentially with the dimension of the feature vector. The common recommendation is to use five to ten times the number of samples for each class with respect to the dimension of the feature vector. To make this problem clearer, let's examine it with an example. Suppose that we record a one-second trial with a sampling rate of 250 Hz from 32 EEG channels. As a result, we have stored 250 dimensions for each channel and the total dimension of the feature vector will be 32×250=8000, in which case the amount of training data will be at least 40000 samples. It is clear that this amount of samples cannot be recorded from the subject for a mental state. It should be borne in mind that only one second of signal is considered. This is where the importance of feature extraction becomes apparent. For this reason, special attention has been paid to feature extraction and dimensionality reduction in the architecture of the proposed method.

In the next step, the Reshape operation should be performed. The purpose of this step is to transform the CNN output data into a sequence for the LSTM. It should be noted that this function does not change the data sizes and only changes the data appearance. This function is defined as follows:

$$LSTM_{data} = Reshape\ (CNN_{data}) \tag{12}$$

The output of the CNN_{data} network is defined as follows, where X is the frequency dimension, Y is the time dimension, and Z is the number of filters.

$$CNN_{data} \sim [X \times Y \times Z] \tag{13}$$

$LSTM_{data}$ is also defined as follows:

$$LSTM_{data} \sim [M \times N] \tag{14}$$

Where M is the number of Time Steps and N is the number of features. Accordingly, the Reshape function is defined as follows:

$$Reshape \quad \rightarrow \quad M = Y, \qquad N = X.Z \tag{15}$$

Therefore, we have:

$$LSTM_{data} \sim [Y \times (X.Z)] \tag{16}$$

Which finally gives us:

$$[Y \times (X.Z)] = Reshape\ ([X \times Y \times Z]) \tag{17}$$

Then there are two LSTM stages whose purpose is to extract sequential features. In fact, temporal features are learned by these two layers of the LSTM network. The final classification is also performed by Dense and Softmax layers. One of the four classes will be activated at the output. It should be noted that in the proposed method, the STFT and CWT transforms are classified separately and then their results are combined and the final classification result will be determined. The data combination strategy in the proposed method is the weighted average of the classifier results for the two STFT and CWT inputs as follows, where coefficients A and B are weight coefficients that are equal to 0.5 by default.

$$OutClass = \frac{A.Softmax\big(X_{cwt}(a,b)\big) + B.Softmax\left(X_{stft}(a,b)\right)}{2} \tag{18}$$

where $OutClass$ is a four-class variable:

$$OutClass \equiv 0\text{: Left hand, } 1\text{: Right hand, } 2\text{: Both feet } \ 3\text{: Tongue} \tag{19}$$

In this way, the EEG input signal from 8 different channels will finally be converted into one of the 4 output classes.

Simulation of the Proposed Method

The BCI Competition IV dataset was used for simulation in this study [26]. This dataset has been used in most of the methods and has four output classes, which are summarized in **Table 1**.

Table 1 Four classes in the dataset

Class 1	Left hand
Class 2	Right hand
Class 3	Both feet
Class 4	Tongue

This dataset consists of 9 files, each containing EEG recordings from a different subject recorded at Graz University of Technology. The recordings consist of 22 EEG channels and 3 EOG channels sampled at 250 Hz while subjects performed four distinct motor imagery tasks: left hand, right hand, both feet, and tongue. Hence, this dataset has four classes. The specifications of the BCI Competition IV dataset are given in **Table 2**. A more detailed description of these specifications can be found in [26].

Table 2 Specifications of the BCI Competition IV dataset

Specifications	Value-Type
Raw Channels	22 EEG+ 3 EOG
Trials	48=4*12
Inter-electrode Distance	3.5 cm
Sampling Frequency	250 Hz
Band Pass Filter	0.5-100 Hz
Amplifier Sensitivity	100 μV
Notch Filter	50 Hz
Participant	9
Number of trials per Session	288
Electrode Type	Ag/AgCl

2-1 Hyperparameters and Constants

The choice of hyperparameters and constants is very effective on the quality of the proposed method. In machine learning, a hyperparameter is a parameter whose value is determined before the learning process begins. Unlike other model parameters, which are determined during the learning process. Constants also refer to variables that have a predetermined size in different methods. The values of hyperparameters and research constants are specified in **Table 3**.

Table 3 Hyperparameter values and research constants

Raw Channels	22
Selected Channels	8 (C3, C4, Cz, CP4, CPz, FC3, FC4, CP3)
τ (Shift Value)	5 ms
S (Noise Coefficient)	100
θ (Rotation Angle)	5 degrees
Window Type (STFT)	Hamming
FFT Size (STFT)	128

Window Length (STFT)	128
Frequency Range (STFT)	6−30Hz
Mother Wavelet (CWT)	Morlet
Scales (CWT)	1-64
Fs (STFT-CWT)	250 Hz
Image Dimension	64 × 128
Input Tensor	64 × 128 × 8
CNN Filter Size	3×3
Feature Map	16 × 32 × 32
Reshape	32 × 512
Learning Rate	0.001

2-2 Performance Evaluation Metrics

For Accuracy Evaluation And the speed of the proposed method, four metrics are used. The first three metrics indicate the accuracy of the proposed method in detecting the correct class of the EEG signal.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (20)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \quad (21)$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (22)$$

Where T indicates the correctly classified samples and F the wrong samples. In addition to the three criteria of Accuracy, Sensitivity, and Specificity, time in seconds is also important in quantifying the proposed method. Hence, the average test time (Average Test Time), which is specified by the variable ATT, is the fourth evaluation criterion. The average test duration for an input is in seconds and is obtained from the following equation, where K is the number of test inputs, and $TT(D_j)$ is the test duration of the D_j sample.

$$ATT = \frac{\sum_{j=1}^{K} TT(D_j)}{K} \quad (23)$$

2-2-1 STFT and CWT Transformation Simulation Results

The BCI Competition IV – Dataset 2a dataset has 22 channels. As mentioned, not all 22 channels were used in this study, because in this case the data would have been too large and difficult to work with. A comparison of the selected channels with other channels is shown in **Figure 5**.

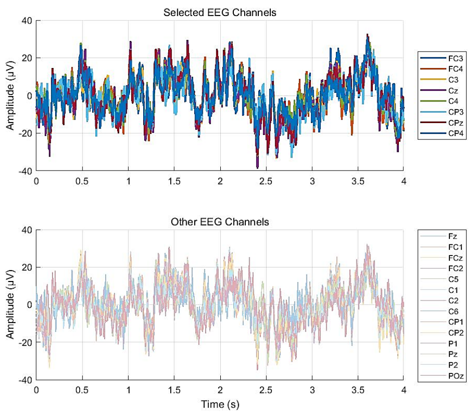


Figure 5 Comparison of selected channels with other channels

Natural Resources for
Human Health
eISSN: 2583-1194

The results of the STFT transformation for four different classes are shown in **Figure 6**. For these transformations, the first trial of each class and also the C3 channel are used. This channel theoretically has more information than the other classes.

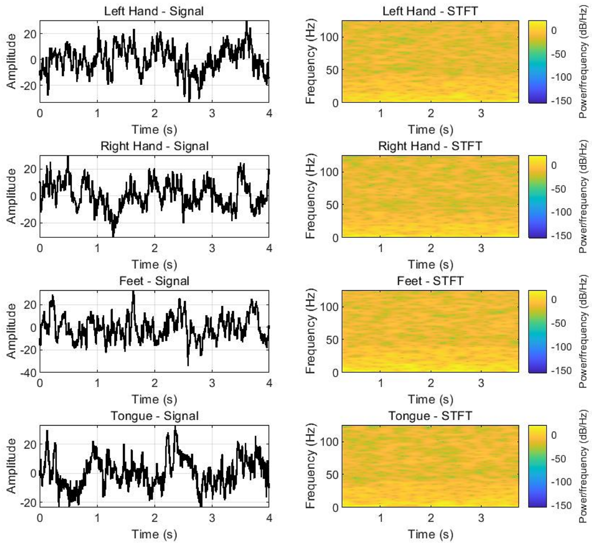


Figure 6 STFT transformation results for four different classes

The results of the CWT transformation for four different classes are shown in **Figure 7**. Here, similarly to the previous case, the first trial of each class and also the C3 channel are used for the transformations.

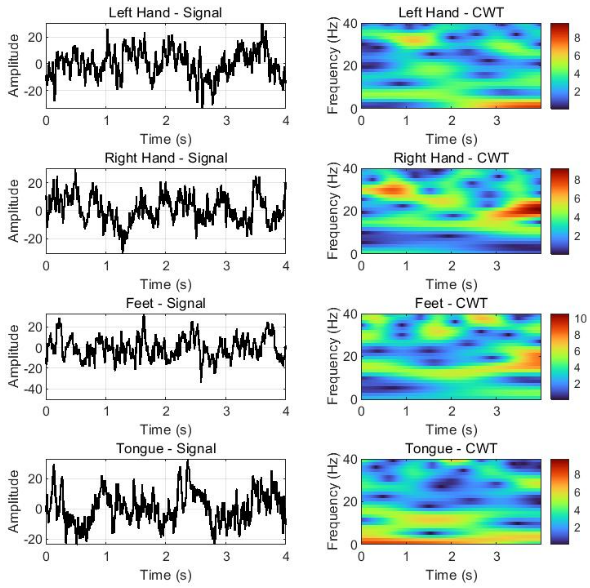


Figure 7 STFT transformation results for four different classes

2-2-2 Results

The main challenge of CNN+LSTM+STFT+CWT architecture is its high computational load and consequently its slowness. In applications based on "taking action based on brain decision", the electrical activity of the individual's brain (brain stripe) is measured and analyzed, and the individual's decision is converted into commands controlling the desired device through processing methods. Obviously, in such applications, the speed of classification is an important criterion. Therefore, to solve the challenge of the slowness of CNN+LSTM+STFT+CWT architecture, hyperparameter optimization (HPO) can be used. HPO is one of the important methods for improving the results in classification methods. The graphs in **Figure 8** show the changes in Accuracy and ATT as a function of the changes in CNN Filter Size and FFT Size hyperparameters.

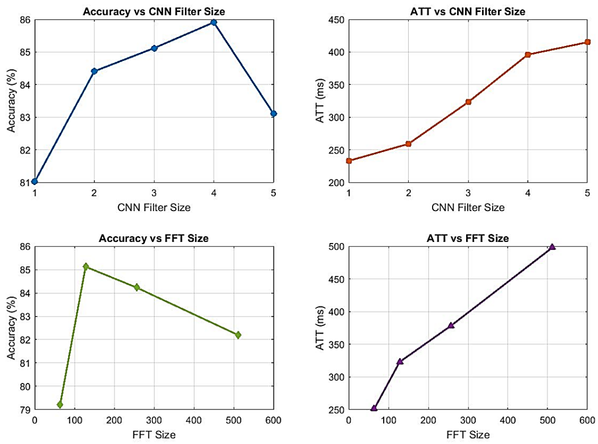


Figure 8 Changes in Accuracy and ATT as a function of the changes in CNN Filter Size and FFT Size hyperparameters

As is clear, smaller filters do not capture enough spatial information, and larger filters cause over-smoothing and reduce the details of EEG patterns. Obviously, the larger the filter, the greater the number of multiplications and additions in each convolution layer and the slower the speed. On the other hand, a small FFT has low frequency resolution, and the EEG band information is not well separated. Conversely, a very large FFT also reduces the temporal resolution. It is clear from the changes in the graph that as the FFT becomes larger, the STFT calculations increase and the feature extraction time before entering the CNN becomes longer.

These graphs show that the choice of hyperparameters has a great impact on the final accuracy and speed. The simulation results of the four evaluation criteria for the CNN+LSTM+STFT+CWT+HPO architecture are shown in **Table 4**. Here, the hyperparameters are selected based on trial and error. As is clear, in this case, the speed has increased slightly, but the price has been a decrease in accuracy.

Table 4 Simulation results of the four evaluation criteria for the CNN+LSTM+STFT+CWT+HPO architecture

Metric	Value
Acc (%)	84.23
Sen(%)	83.23
Spe (%)	84.11
ATT (s)	256

2-3 Discussion and Comparison

Table 5 presents a comparison of different architectures in terms of methodology. Also, **Table 6** presents the simulation results of the four evaluation criteria for different architectures in full. As can be seen, the theory of the first and second architecture was introduced in [16] and the third architecture in [11].

Table 5 Comparison of different architectures in terms of methodology

Architecture	Transforms	Augmentation	Enhancement	Hybrid Net	HPO
CNN+FSST [16]	1	×	×	×	×

CNN+WSST [16]	1	×	×	×	×
CNN+HHT [11]	1	×	×	×	×
CNN+LSTM+STFT	1	✓	✓	✓	×
CNN+LSTM+CWT	1	✓	✓	✓	×
CNN+LSTM+STFT+CWT	2	✓	✓	✓	×
CNN+LSTM+STFT+CWT+HPO	2	✓	✓	✓	✓

Our method uses a hybrid network consisting of two CNN and LSTM models, while the methods [16] and [11] used a single CNN model. The difference between the Transform methods between our method and other methods is also highlighted. On the other hand, our proposed method uses the data augmentation algorithm and the ESPCN image quality enhancement mechanism, which [16] and [11] lack.

Table 6 Simulation results of four evaluation criteria for different architectures

Architecture	Acc	Spe	Sen	ATT
CNN+FSST [16]	68.45	65.10	69.00	128
CNN+WSST [16]	70.11	71.12	72.36	149
CNN+HHT [11]	76.22	76.56	75.11	326
CNN+LSTM+STFT	60.92	63.44	71.02	**118**
CNN+LSTM+CWT	75.17	72.15	71.2	221
CNN+LSTM+STFT+CWT	**85.12**	**84.00**	**84.19**	323
CNN+LSTM+STFT+CWT+HPO	84.23	83.23	84.11	256

The simulation results of the three accuracy evaluation criteria for different architectures are shown in the form of a bar graph in **Figure 9**. According to this figure, it is clearly clear that the CNN+LSTM+STFT+CWT architecture is the most accurate architecture. Also, the two CNN+LSTM+STFT+CWT and CNN+HHT architectures are the slowest architectures.

Natural Resources for
Human Health
eISSN: 2583-1194

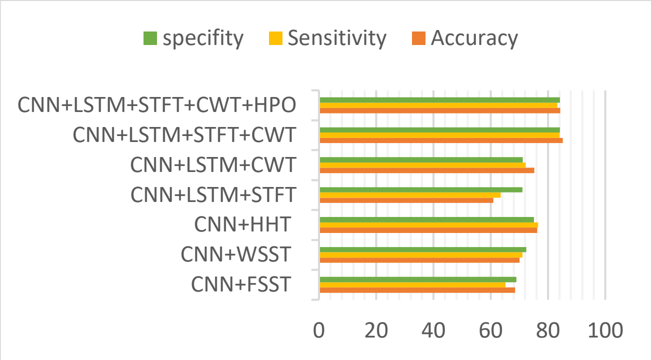


Figure 9 Simulation results of the three accuracy evaluation criteria for different architectures

In **Figure 10**, the architectures are also compared in terms of ATT (ms). It is clear from the graph that the most important limitation and implementation challenge of our method is its slowness compared to similar architectures. Obviously, our method uses a hybrid architecture and binary structure in the transformation phase. In other words, our method uses both STFT and CWT as EEG signal converters. This makes the proposed method about twice as slow. In brain-computer interaction (BCI) applications, this slowness is a big challenge.

This challenge is due to the type of applications expected from the results of this research. For example, various neurological diseases such as spinal cord injury, cerebral palsy, stroke, etc. lead to partial or complete loss of voluntary muscle activity and cause movement or speech disorders [27, 28]. In such cases, a fast and accurate BCI can reduce the negative effects of disability and improve the quality of life of people with disabilities. Brainstem-based BCI is a new non-invasive communication channel for communicating between the brain and the environment, without using other natural communication channels such as hands, feet, and tongue [28]. In this method, the electrical activity of the person's brain (brainstem) is measured and analyzed, and the person's decision is converted into commands controlling the desired device through processing methods. In other words, the ultimate goal of a brain-computer interface is to help people with severe disabilities to interact effectively with their environment through external devices such as computers, speech-to-speech converters, assistive devices, and neural prostheses. Such an interface can increase the level of independence of these people and improve their quality of life. Now, if this BCI is slow, it may actually lose its efficiency. One solution to increase the speed and real-time of the system was to use the HPO mechanism. However, research in this area is still a serious need.

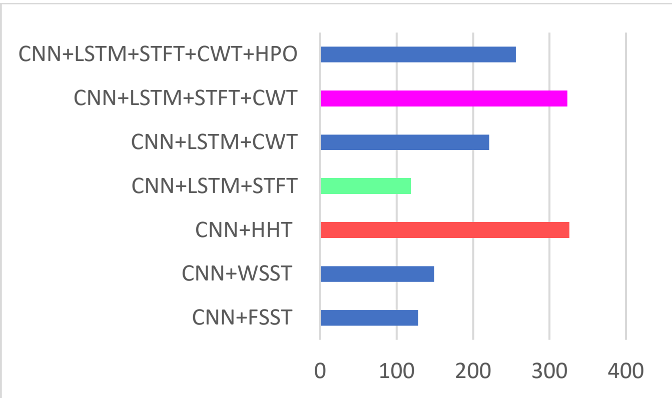


Figure 10 Comparison of architectures in terms of ATT (ms)

The graph in **Figure 11** compares the accuracy and speed of the 5 transforms used in the study (two proposed transforms STFT and CWT and three transforms in previous studies). For better comparison, these values have been normalized between 0 and 1. According to this graph, the fastest method is STFT and the slowest is HHT. Also, the most accurate method is HHT and the least accurate method is STFT. Hence, choosing the appropriate deep network architecture along with choosing an efficient method can greatly help improve the classification process.

Natural Resources for Human Health
eISSN: 2583-1194

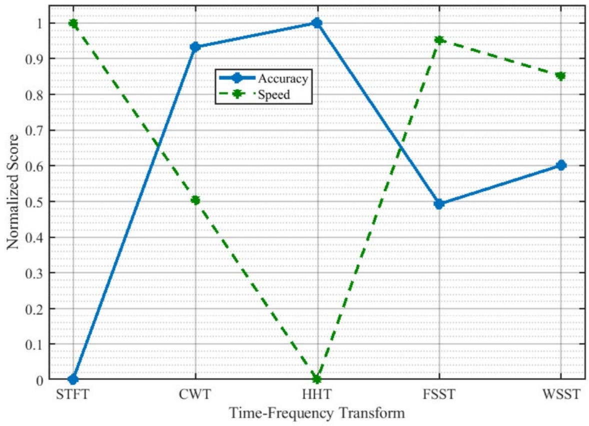


Figure 11 Comparison between accuracy and speed of the 5 transformations used in the study

Despite the challenges in speed, in summary, the simulation results on the BCI Competition IV dataset with an accuracy of 88.23% show that the three innovations of the proposed method led to an improvement in the classification process of EEG time-frequency representations.

CONCLUSION

In this paper, a hybrid deep learning method based on CNN+LSTM network was proposed to improve the classification of time-frequency representations of EEG signals with data augmentation and improved data quality.Simulation results on the BCI Competition IV dataset showed that combining spatial and temporal features, along with ESPCN methods and data augmentation, improved the classification performance and achieved an accuracy of 88.23%. EEG signal processing is a dynamic and developing field. Although significant progress has been made in this field, there are still limitations. One of the most important limitations is the complexity and high noise of the EEG signal. Separating brain signals from environmental noise and noise caused by physical activities requires advanced signal processing techniques. Thus, noise reduction techniques and adaptive filters can lead to increased accuracy. Also, in most methods, such as our method, full optimization of all hyperparameters has not been considered. It is recommended that more attention be paid to this issue in future research. The main limitation of our method was also the real-time nature of the classifier. This limitation may be compensated by fully optimizing all hyperparameters and using a newer feature extraction process in the future. It is hoped that the results of this research will take a small step in the development of EEG-based software and help patients with spinal cord injury, cerebral palsy, and stroke better communicate with the world around them.

REFERENCES

[1] M. Orban, M. Elsamanty, K. Guo, S. Zhang, and H. Yang, "A review of brain activity and EEG-based brain–computer interfaces for rehabilitation application," *Bioengineering,* vol. 9, no. 12, p. 768, 2022.

[2] G. Manoharan and D. R. Faria, "Enhanced mental state classification using EEG-based brain–computer interface through deep learning," in *Intelligent Systems Conference*, 2024: Springer, pp. 570-586.

[3] D. Lin *et al.*, "High accuracy EEG signal classification for brain computer interfaces using advanced neural architectures," *Frontiers in Neuroscience,* vol. 20, p. 1752176, 2026.

[4] J. W. Ahn, G. Y. Yu, S.-W. Kim, Y.-S. Seok, K.-M. Byun, and S. H. Choi, "A Cheonjiin Layout Mental Speller: Developing a Simple and Cost-Effective EEG-Based Brain–Computer Interface System," *Sensors,* vol. 26, no. 7, p. 2265, 2026.

[5] J. A. Mendivil Sauceda, B. Y. Marquez, and J. J. Esqueda Elizondo, "Emotion classification from electroencephalographic signals using machine learning," *Brain Sciences,* vol. 14, no. 12, p. 1211, 2024.

[6] S. Aslam, M. Hussain, X. Li, H. Li, and H. Wu, "HAMD-Net: A biologically informed hybrid attention model for EEG classification of multiple neurological disorders," *Biomedical Signal Processing and Control,* vol. 120, p. 110049, 2026.

[7] J. Altamirano, "Motor Imagery EEG Classification using Common Spatial Patterns and Machine Learning: A Cross-Subject Study," *preprint, Mar,* 2026.

[8] K. K. Ang *et al.*, "Brain-computer interface-based robotic end effector system for wrist and hand rehabilitation: results of a three-armed randomized controlled trial for chronic stroke," *Frontiers in neuroengineering,* vol. 7, p. 30, 2014.

[9] H.-T. Lee *et al.*, "A review of hybrid EEG-based multimodal human–computer interfaces using deep learning: applications, advances, and challenges," *Biomedical Engineering Letters,* vol. 15, no. 4, pp. 587-618, 2025.

[10] M. A. Pfeffer, S. S. H. Ling, and J. K. W. Wong, "Exploring the frontier: Transformer-based models in EEG signal analysis for brain-computer interfaces," *Computers in Biology and Medicine,* vol. 178, p. 108705, 2024.

[11] A. Eltvik, "Deep learning for the classification of EEG time-frequency representations," *[Norwegian]: Norwegian University of Science and Technology,* 2018.

[12] P. Samal and M. F. Hashmi, "Role of machine learning and deep learning techniques in EEG-based BCI emotion recognition system: a review," *Artificial Intelligence Review,* vol. 57, no. 3, p. 50, 2024.

[13] J. Samuel, T. K. Murugan, L. Govindaraj, M. Balaji, V. SenthilKumar, and S. Sundararajan, "Adversarial robust EEG-based brain–computer interfaces using a hierarchical convolutional neural network," *Scientific Reports,* vol. 16, no. 1, p. 4353, 2026.

[14] Y. Narayan, S. Ananthi, S. Khujamqulov, H. K. Thind, D. Juneja, and R. Kaur, "Deep Learning-Based EEG Signal Classification for Enhancing Brain-Computer Interface Accuracy," in *2026 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, 2026, vol. 4: IEEE, pp. 1-6.

[15] M. Bensimon, Y. Hadad, Y. Ben-Shimol, and S. Greenberg, "Time-Frequency Analysis for Feature Extraction Using Spiking Neural Network," *Authorea Prepr,* 2023.

[16] S. Madhavan, R. K. Tripathy, and R. B. Pachori, "Time-frequency domain deep convolutional neural network for the classification of focal and non-focal EEG signals," *IEEE Sensors Journal,* vol. 20, no. 6, pp. 3078-3086, 2019.

[17] J. J. Bird, D. R. Faria, L. J. Manso, P. P. Ayrosa, and A. Ekart, "A study on CNN image classification of EEG signals represented in 2D and 3D," *Journal of neural engineering,* vol. 18, no. 2, p. 026005, 2021.

[18] W. Qing-Hua, W. Li-Na, and X. Song, "Classification of EEG signals based on time-frequency analysis and spiking neural network," in *2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2020: IEEE, pp. 1-5.

[19] D. Truong, M. Milham, S. Makeig, and A. Delorme, "Deep convolutional neural network applied to electroencephalography: Raw data vs spectral features," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021: IEEE, pp. 1039-1042.

[20] G. Lekkas, E. Vrochidou, and G. A. Papakostas, "Time–frequency transformations for enhanced biomedical signal classification with convolutional neural networks," *BioMedInformatics,* vol. 5, no. 1, p. 7, 2025.

[21] M. Bensimon, Y. Hadad, Y. Ben-Shimol, and S. Greenberg, "Time–frequency analysis using spiking neural network," *Neuromorphic Computing and Engineering,* vol. 4, no. 4, p. 044001, 2024.

[22] V. Srimadumathi and M. Ramasubba Reddy, "Classification of Motor Imagery EEG signals using high resolution time-frequency representations and convolutional neural network," *Biomedical Physics & Engineering Express,* vol. 10, no. 3, p. 035025, 2024.

[23] J. Zheng *et al.*, "Time-frequency analysis of scalp EEG with Hilbert-Huang transform and deep learning," *IEEE Journal of biomedical and health informatics,* vol. 26, no. 4, pp. 1549-1559, 2021.

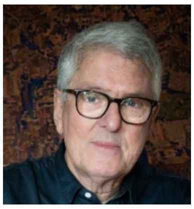
[24] H. B. Ahmed, M. S. Mohammed, and A. H. Saleh, "Classification of EEG Signal Using Deep Neural Network Techniques," *Mathematical Modelling of Engineering Problems,* vol. 12, no. 3, 2025.

[25] A. E. Top, "Classification of Eeg signals using transfer learning on convolutional neural networks via spectrogram," Ankara Yıldırım Beyazıt Üniversitesi Fen Bilimleri Enstitüsü, 2018.

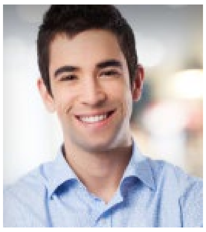
[26] G. Schalk, "Review of the BCI competition IV," 2012.

[27] A. K. Selvam and A. Loganathan, "An intelligent EEG-based ensemble framework for communication assistance in Locked-In Syndrome patients," *Scientific Reports,* 2026.

[28] Y. Peng, Y. Jiang, Z. Zuo, S. Gu, Z. Wang, and Z. Tian, "A rehabilitation design concept based on brain–computer interface and McKibben artificial muscle," *Healthcare and Rehabilitation,* vol. 2, no. 1, p. 100066, 2026.



Scientific papers and research records are one of the most important elements of any researcher's resume. Including these records correctly and using Scientific papers and research records are one of the most important elements of any researcher's resume. Including these records correctly and using a systematic and attractive method can differentiate you from the rest of the competition and offer you more job and research opportunities. Scientific papers demonstrate your ability to research and analyze data, innovate and produce new knowledge. But including these papers on your resume requires an effective method that can showcase your abilities well.



Scientific papers and research records are one of the most important elements of any researcher's resume. Including these records correctly and using Scientific papers and research records are one of the most important elements of any researcher's resume. Including these records correctly and using a systematic and attractive method can differentiate you from the rest of the competition and offer you more job and research opportunities. Scientific papers demonstrate your ability to research and analyze data, innovate and produce new knowledge. But including these papers on your resume requires an effective method that can showcase your abilities well.