Heliyon 10 (2024) e25999

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

EEG-based identification system using deep neural networks with frequency features

Yasaman Akbarnia, Mohammad Reza Daliri *

Biomedical Engineering Department, School of Electrical Engineering, Iran University of Science and Technology (IUST), Tehran, Iran

ARTICLE INFO

Keywords:
Identification
Electroencephalography (EEG)
Deep neural network (DNN)
Optimization
Regularization

ABSTRACT

Improving system security can be achieved through people identification. Among various methods, electroencephalography-based (EEG-based) identification is a dependable way to prevent identity theft and impersonation. Due to the distractions present in the identification environment, such as lack of focus, mental engagement, small body movements, blinking, and other noises, it is essential to analyze data that reflects these conditions. The present research aims to advance practical EEG-based identification by studying data with mental preoccupation and developing a suitable algorithm.

In this article, data from a study conducted on a group of 109 individuals has been analyzed. The data is categorized into two groups: focused data and waiting data. The article describes preprocessing the data and extracting three types of features, including Statistical, Frequency, and Wavelet. Then, a deep neural network (DNN) is used to classify the data. The DNN utilizes a multilayer, fully-connected neural network, with the number of layers and neurons varying based on the data type. Optimization and regularization methods are employed to improve the accuracy of the results. The DNN achieved an average accuracy of 99.19% for frequency features over all subjects in the focused data category, while the waiting data category showed an accuracy of 97.81%.

1. Introduction

In today's world, identification passwords are commonly used for various purposes. When authenticating, the local operating system or server compares the user's information with the database. If the information matches, the user is granted access to resources according to the rules. This way, organizations can safeguard their information by ensuring that only authorized users can access it.

Traditional methods, like user passwords and smart cards, are susceptible to attacks. Biometric systems have become an essential means to enhance security. These systems measure personal traits to verify individuals. The most significant biometric systems include fingerprint scanning, voice recognition, iris recognition, and face detection. The dependability of an identification system is crucial, and its confidence level should match its intended application.

Many security systems use biometric technology, such as fingerprint recognition. However, this technology has some security vulnerabilities [1]. Voice recognition is another low-cost option that can be forged through simulators [2]. Iris detection is a dependable option, but it is more complicated and also vulnerable to counterfeiting [2]. Face detection is quick and precise, but it can be defeated by anti-surveillance masks [2].

* Corresponding author.
E-mail address: daliri@iust.ac.ir (M.R. Daliri).

https://doi.org/10.1016/j.heliyon.2024.e25999
Received 23 June 2023; Received in revised form 23 December 2023; Accepted 6 February 2024
Available online 11 February 2024
2405-8440/

The EEG is a cost-effective and non-invasive way of recording brain signals. However, analyzing these signals for identification purposes can be challenging due to the influence of a person's mood, tension, and mental state. This makes it difficult to forge EEG signals in threatening situations [3]. EEG is particularly resistant to decoding and cannot be replaced with another person's brain signals. These factors make EEG an essential tool for developing biometric systems for high-security purposes [4].

When defining a biometric system, it is important to consider essential features that make it effective against forgery and imitation. The EEG's existence of these features is examined by the

Following [5].

- **Resilience**

EEG-based identification systems are suitable for authentication due to the invisibility and non-copyable nature of EEG data, providing resistance against forged identities.

- **Universality**

Every person's EEG signals are always present, making security applicable at all times.

- **Uniqueness**

Each person's EEG signals are unique, which can enhance accuracy in detection.

- **Accessibility**

There is a growing trend towards developing affordable and easy-to-use EEG headsets, which has led to increased convenience and reduced prices for EEG products.

In addition to structural differences, the human brain also has functional differences that arise from variations in memory, personality, and cognitive patterns. A study conducted in 1998 by Stassen et al. [6] on patients with mental illnesses revealed that the mental traits of family members were more alike during periods of rest. Research in 2007 demonstrated the heritability of alpha and beta power in the occipital region [7].

Research conducted by DeGenrau et al. [8] in 2008 revealed that the 8–16 Hz frequency is inherited. This study led to the concept of "brainprinting", which suggests using EEG as a form of authentication. This approach takes into consideration differences in brain structure and genetic inheritance.

In 2013, the Yeom team [9] achieved 86.2% accuracy for ten users using event-related potential (ERP). In this study, N170 and N250 were used as temporal features.

In 2015, Dai et al. [4] recorded the motor cortex signals of 16 individuals while at rest. They achieved an accuracy rate of 93% using the support vector machine (SVM) method. However,

The limitation of this study was the small number of participants in the experiments.

In 2016, Bashar et al. [10] were able to achieve an accuracy rate of 87.3% by filtering the signals in the frequency range of 0.5–59 Hz. This range includes delta, theta, alpha, beta, and a part of the gamma pattern. The classification for this study was performed using statistical features with the SVM.

In 2016, Corb et al. [11] examined the open-eye and closed-eye state data recorded by Physionet. They utilized the phase-lag index and graph theory in their study, which resulted in 94.3% and 90.5% accuracies for the gamma and alpha bands, respectively.

In 2016, Rodrigues et al. [12] conducted a study on 109 individuals using the imagery-movement section and obtained an 87% performance on 32 channels. They utilized the autoregressive model and the flower pollination algorithm in their research. The data used in the present article is the same as that used in the study of Rodrigues and his colleagues.

A study conducted by Mo et al., in 2017 [13] involved 16 participants and focused on feature extraction. The features were divided into four groups: spectral entropy, approximate entropy, sample entropy, and fuzzy entropy. Their results indicated that the fuzzy entropy feature presented the highest accuracy of 90.7% when used with SVM classification.

In 2019, Yu et al. [14] conducted EEG recordings using nine channels. The study aimed to extract the low-frequency component in SSVEP signals, which includes ERP transient responses. Visual stimulation was used on eight people, and the convolutional neural network (CNN) achieved 97% performance.

In 2020, a study led by Zhang [15] had 46 participants in a 7-min resting experiment with a single channel. The data was categorized into three groups of features. The Rayleigh quotient (RQ) method was used to determine the best features. The most accurate outcome was achieved using SVM, with an accuracy of 95.48%.

An experiment was conducted in 2020 using auditory evoked potential, achieving 94.5% accuracy through frequency features and LDA classification [16].

In 2020, Barayeu et al. [17] conducted a study on data presented in this article. They used a VGG-like NN-SVM decoder and an inception-like NN-SVM decoder on 64 channels, achieving accuracies of 90.68% and 93.40%, respectively.

The present study aims to investigate EEG-based systems for adults with normal conditions who can maintain self-control and remain calm for a certain period. The research is specifically focused on analyzing data while the participants are at rest. Conducting EEG-based identification outside of a laboratory setting can be difficult as it is hard to determine the users' thoughts, imagery, or activities before resting. Hence, it is essential to investigate EEG-based identification in waiting conditions. This article examines the effects of mental busyness on outcomes. Hence, it is necessary to create identification tasks that take into mental busyness. The primary objective of this article is to create an appropriate algorithm and analyze data that considers mental busyness.

The second section of the article covers various processing techniques, including preprocessing, feature extraction, dimensional reduction, and classification. The baseline data classification is demonstrated in the third section, proving that the algorithm remains unaffected by eye conditions. Next, the waiting data is analyzed, which is contaminated with disturbances caused by the mind and body movement. Finally, the last section delves into the content presented in the previous sections.

2. Materials and methods

2.1. EEG dataset

The data presented is taken from the EEG Motor Movement/Imagery Dataset on the Physionet website. A significant number of subjects are required to create a biometric database. Although the Physionet dataset is not designed for this purpose, it can be effective due to its extensive subject pool.

The dataset contains over 1500 EEG recordings lasting one to 2 min each, collected from 109 volunteers. Brain signals were recorded using the BCI2000 system (http://www.bci2000.org) during various imagining-motion tasks. The data includes 64 EEG channels with a 10-10 system with a sampling rate of 160 Hz. Each volunteer participated in 14 runs, as follows [18,19].

- In the baseline state, two tasks last for 1 min each - one with open eyes and the other with closed eyes.

The following tasks last for 2 min. Each task is done with three different runs [18,19]. The execution pattern for these tasks is shown in Fig. 1.

- During the task, a target will appear on either the left or right side of the screen. The subject will then open and close his/her corresponding fist until the target disappears. After this, there will be a 4-s relaxation period before repeating the task for another 4 s.
- In another task, the target will appear on the left or right side of the screen, and the subject will imagine opening and closing his/her corresponding fist until the target disappears. Again, there will be a 4-s relaxation period before repeating the task for another 4 s.
- When a target appears at the top or bottom of the screen, the user should open and close both fists (for top targets) or both feet (for bottom targets) until the target disappears. Then, the subject should take a 4-s break before repeating the task for another 4 s.
- When a target appears at the top or bottom of the screen, the subject should imagine opening and closing both fists (for top targets) or both feet (for bottom targets) until the target disappears. Then, they should take a 4-s break before repeating the task for another 4 s.

a) Open- and closed-eye Data

The open-eye and closed-eye data to create the OC dataset are combined, which is defined for 109 individuals. The dataset includes a 2-min task with a sampling rate of 160 Hz. With 4-s trials, there are a total of 30 trials.

b) Waiting Data

The duration of tasks was mentioned as 2 min at a sampling rate of 160 Hz. However, for subjects 88, 92, 100, 104, and 106, the tasks were completed in less than 2 min. By removing the data of these subjects, data gathering remains for 104 people. The data is then used to create a 6-min dataset called BW by extracting the expectancies between imagery-movement and movement tasks. Considering 4-s trials, 90 trials exist. Fig. 2 provides different steps to the proposed algorithm.

Two main steps are required to complete the identification process. The first step involves identifying the username and adding the necessary information to the database. In the second step, a unique password is created for the user to use for identification. This article will focus on the second step. The objective is to classify each individual, assuming all 109 individuals have been correctly identified. To achieve this, EEG is recorded for 4 s and then analyzed to identify the individual.

2.2. Preprocessing

Proper preprocessing of EEG signals is essential to enhance the accuracy and reliability of the classification outcomes. EEG signals

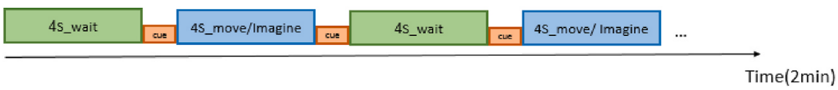

Fig. 1. The general pattern of performing a task.

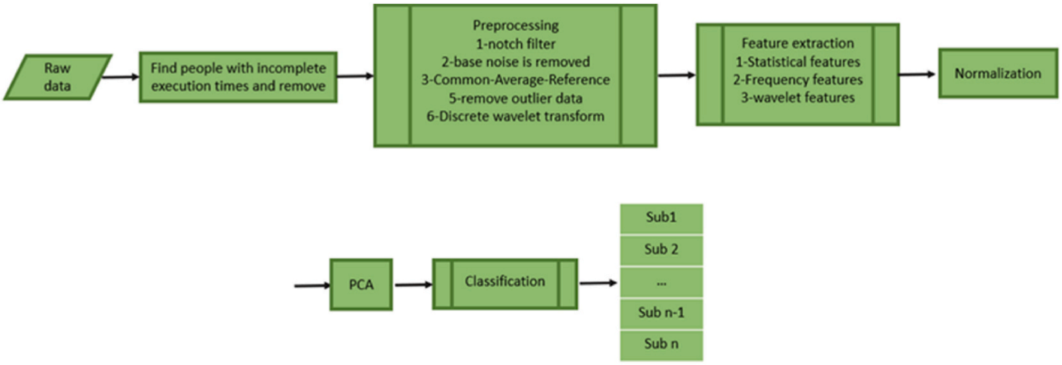

Fig. 2. Different steps of the proposed algorithm.

are frequently plagued with various types of noise, artifacts, and unwanted physiological activities, which may hinder the classification process and cause imprecise results [20].

The notch filter is used to remove the 50 Hz noise. Another type of noise is the base noise signal, caused by respiration, electrode impedance changes, and excessive body movements. Removing the base noise signal before processing to avoid adverse effects on results is important. The removal process involves analyzing 1-s windows and creating a 5-order polynomial model for each channel. The model is then subtracted from the main signal [20].

To improve accuracy in noisy EEG recordings, researchers use a technique called common average reference (CAR). This technique involves calculating the average of each sample across all channels and then subtracting that value from each channel. CAR helps to detect small effects and improves the signal-to-noise ratio [20,21].

The removed outlier filter eliminates anomalies in EEG signals from various sources like muscle activity, eye blinks, or electrode movements. This method produces a more consistent signal and enhances the signal-to-noise ratio without reducing data. The filter calculates 10% of upper and lower data amplitudes at each electrode and removes the top and bottom 10% as outliers, replacing them with the highest and lowest values of the middle 90% [22].

The Discrete Wavelet Transform (DWT) is a signal processing technique that helps to eliminate unwanted noise while retaining significant information in the signal. The signal is divided into two segments - approximation and detail. The approximation segment contains the low-frequency components of the signal, while the detail segment contains the high-frequency components or noise. The first step in this technique is to remove the detail segment, which contains the noise, while retaining the approximation segment. The approximation segment represents a cleaner version of the original EEG signal [23].

2.3. Feature extraction

Feature extraction is a process that transforms raw data into a more useable format for further analysis. Its main objective is to extract the most significant information from the data that can distinguish between different categories or classes. By selecting the most informative features, the classification algorithm can concentrate on the essential aspects of the data, which ultimately leads to improved accuracy in the classification process [24].

a) Statistical Features

This article discusses seven categories of statistical features for 64 channels, which add up to a total of 448 features in the time domain. Additionally, the trials are split into windows, with each window consisting of a quarter of a trial. The seven statistical feature sets are then recalculated for each window, resulting in 28 features. These calculations are performed for all 64 channels, resulting in a total of 1792 features. Altogether, 2240 statistical features are extracted. In equations (1) to (7), Xi represents each sample in the column of the feature matrix and N denotes the length of the columns [24].

- Mean

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{1}$$

- Standard deviation (STD)

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} |X_i - \mu|^2} \tag{2}$$

- Median

$$M = \begin{cases} X\left[\dfrac{n+1}{2}\right] & \text{if } n \text{ is odd} \\ \dfrac{X\left[\frac{n}{2}\right] + X\left[\frac{n}{2}+1\right]}{2} & \text{if } n \text{ is even} \end{cases} \tag{3}$$

- Variance

$$V = \frac{1}{N-1} \sum_{i=1}^{N} |X_i - \mu|^2 \tag{4}$$

- The skewness or asymmetry of the data around the mean, where σ is the standard deviation.

$$Skew = \frac{\sum_{i=1}^{N} (X_i - \mu)^3}{(N-1)\sigma^3} \tag{5}$$

- Kurtosis

$$K = \frac{\sum_{i=1}^{N} (X_i - \mu)^4}{(N-1)\sigma^4} \tag{6}$$

- Power

$$power = \frac{\sum (input \times input)}{Input\ length} \tag{7}$$

b) Frequency Features

The Fourier transform is a mathematical technique that changes signals from the time domain to the frequency domain [24]. The classifier can use the frequency value and amplitude height of the peaks in the frequency spectrum as inputs. To create features, we select the top five frequencies with the highest amplitudes for each channel, leading to a total of 10 features (five frequencies with their corresponding amplitudes). Therefore, with 64 channels, there are 640 features available.

There are six powers in the second group of frequency features, namely Delta (0.4–4), Theta (4–8), Alpha (8-13), Beta (13–30), Gamma 1 (13–50), and Gamma 2 (50–60). These bands create 384 features for every 64 channels. Additionally, the trials are split into windows, with each window consisting of a quarter of a trial. Then, the six power sets are recalculated for each window, resulting in 24 features. These calculations are performed for all 64 channels, resulting in a total of 1536 features. Therefore, a total of 2560 frequency features can be extracted.

c) Wavelet Features

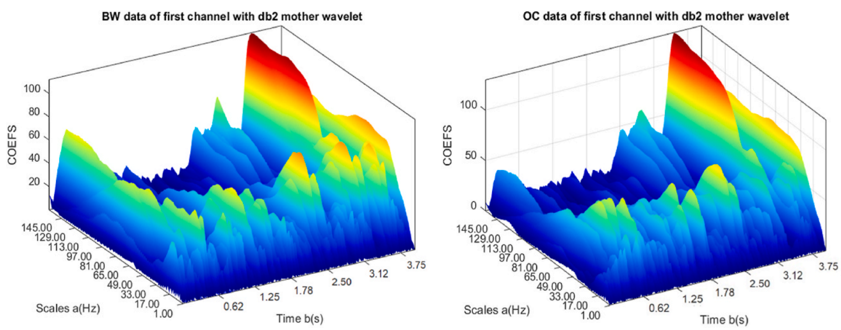

Fig. 3. Coefficients CWT in the first channel, first trial, first person. left) BW data, right) OC data.

For signals like EEG that are constantly changing, wavelet transforms are appropriate for analysis. Unlike sine waves, wavelets are localized and can provide both time and frequency information. Wavelet algorithms require selecting a mother wavelet with a short, transient waveform. In this article, wavelet db2 is chosen as the mother wavelet [24].

- Continuous Wavelet Transforms

The use of continuous wavelet transform (CWT) for feature extraction is not a widely used method. In Fig. 3, the coefficients obtained through the use of the db2 mother wavelet can be seen. Fig. 3 has three axes: Coefficients, Scales a, and Time b. From each axis, we extract the five highest amplitudes, resulting in a total of 15 features. These features are obtained by converting the "Scales a" axe to the frequency domain, from "Time b" and from the coefficients. When these three feature groups are combined, they produce 960 features for 64 channels [23,24].

- Discrete Wavelet Transforms

This paper uses the discrete wavelet transform (DWT) technique, which involves sampling wavelet functions at the sampling rate of 160 Hz. The optimal number of divisions in DWT needs to be determined. The length of the signal is denoted as N, while the number of divisions is represented by m. As per the concept of DWT, the signal undergoes high-pass and low-pass filtering. Formula (8) can be used to calculate the length of the signal N′, after downsampling by a factor of 2 [23,24]:

$$N' = \text{round}\left(\frac{N-1}{2}\right) + m \tag{8}$$

In this method, the main signal is divided into approximations and details. This process is repeated until the remaining signal matches the mother wavelet. In each step, the length of the approximation is compared to the previous step. When the length of the approximation in stage t matches the length in stage t-1, the number of divisions in the previous step is considered the final limit of divisions [23,24].

In this dataset, the division rate is 10. According to Formula (9), the optimal number of divisions with useful information can be defined:

$$\textit{Best decompose} = \frac{\textit{decompose}}{2} \tag{9}$$

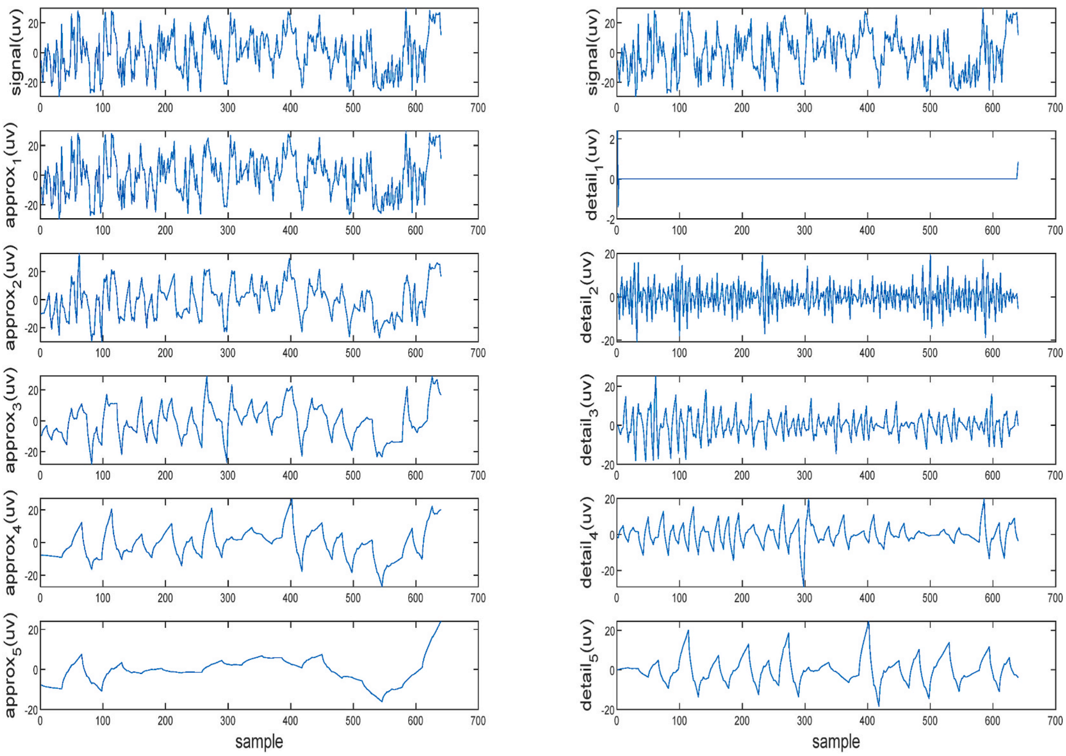

Fig. 4. Signal with the best number of divisions, from the first channel, first trial, and first person in the OC data.

Based on the data, it appears that using five divisions is the optimal choice, as illustrated in Figs. 4 and 5. The top row of these figures depicts the primary signal for the first channel, trial, and person. The second row displays the approximation and detail for the first division. As mentioned in the preprocessing section, the first detail is removed as it is considered to be noise.

2.4. Normalization and dimension reduction

Normalization before the classification process prevents certain features from dominating the learning algorithm. It helps ensure that all features contribute equally to the learning algorithm. Z-Score normalization can be defined according to Formula (10), where norm data is the scaled value, Xi is the original value, μi is the mean of the feature, and σi is the standard deviation of the feature [25]:

$$Norm\ data = \frac{x_i - \mu_i}{\sigma_i} \tag{10}$$

Dimension reduction involves transforming a large number of variables into a smaller set that retains most of the information. One common technique for dimension reduction is principal component analysis (PCA). This technique shows the directions with the highest variance in data analysis. The number of components required to analyze the data depends on the number of features and the variance present. In this paper, the dimensions of all features are reduced by considering 90% of their variance. Formula (11) can indicate the required quantity [26]:

$$explain\ variance = \frac{the\ sum\ of\ the\ eigenvalues\ of\ the\ selected\ features}{sum\ of\ total\ eigenvalues} \tag{11}$$

2.5. Classification

a) Validation Method

It's necessary to reserve some data and their labels for testing sets to prevent overfitting. One way to achieve this is by using the repeated random sub-sampling validation method, which involves dividing the data into training and testing subsets. The model is trained on the training data, while the accuracy of predictions is evaluated using the test data. This method is not affected by the number of divisions. In this article, the test data is divided into testing and validation [27,28]. For BW data, the ratio of training to test is 85 to 15, while for OC data, the ratio is 75 to 25. The upcoming section looks into the use of DNN in classification.

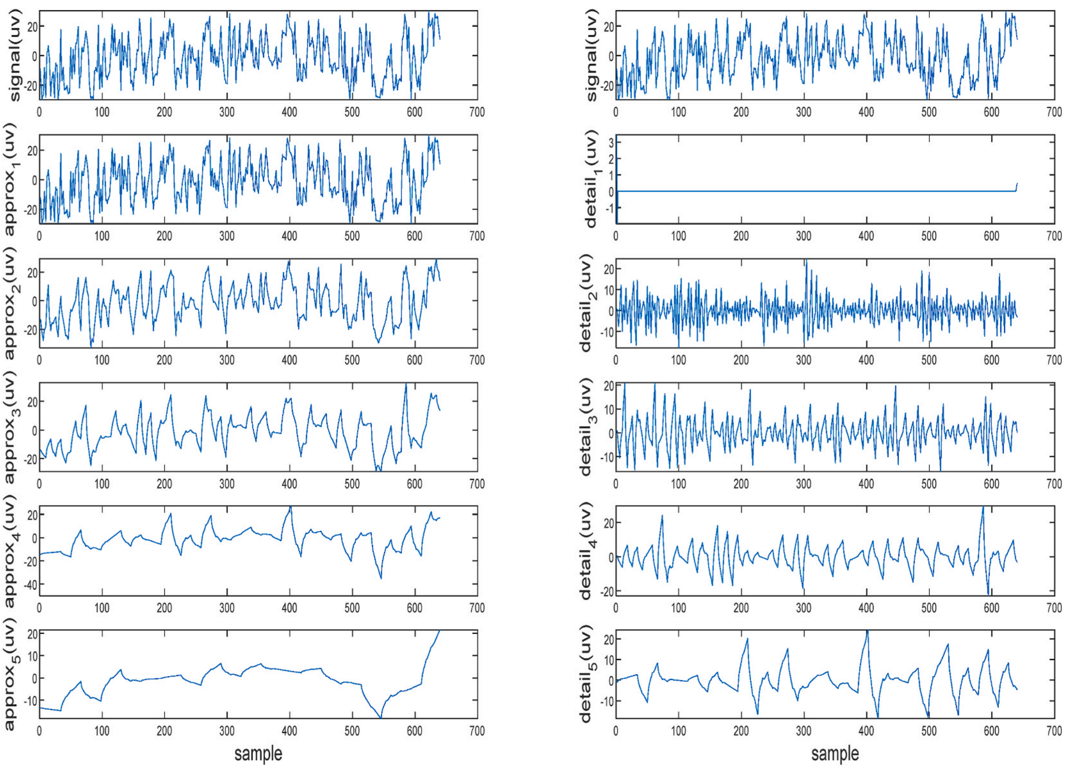

Fig. 5. Signal with the best number of divisions, from the first channel, first trial, and first person in the BW data.

b) Deep Neural Network (DNN)

The human brain receives information from the surrounding environment, processes it, and then produces a response. Neural networks are a type of machine learning that mimics this process. They take in multiple inputs, process the information through neurons in hidden layers, and produce an output using an output layer [29].

The perceptron model is a basic type of artificial neural network in machine learning. It has four main parameters: input nodes, weights and bias, net sum, and an activation function. At the start, the model multiplies the input values by their respective weights and then adds them together to create the weighted sum. This sum is then passed through the activation function 'f' to produce the desired output. When multiple perceptrons are combined, they create a layer. Several layers can be admixtured to form a multilayer perceptron (MLP). An MLP can have multiple hidden layers between the input and output layers [29].

In this article, the activation function in the initial layers is considered RELU, and the output layer is Softmax. The learning rate decreases according to formula (12) per epoch, in which t represents the period in the neural network [29]:

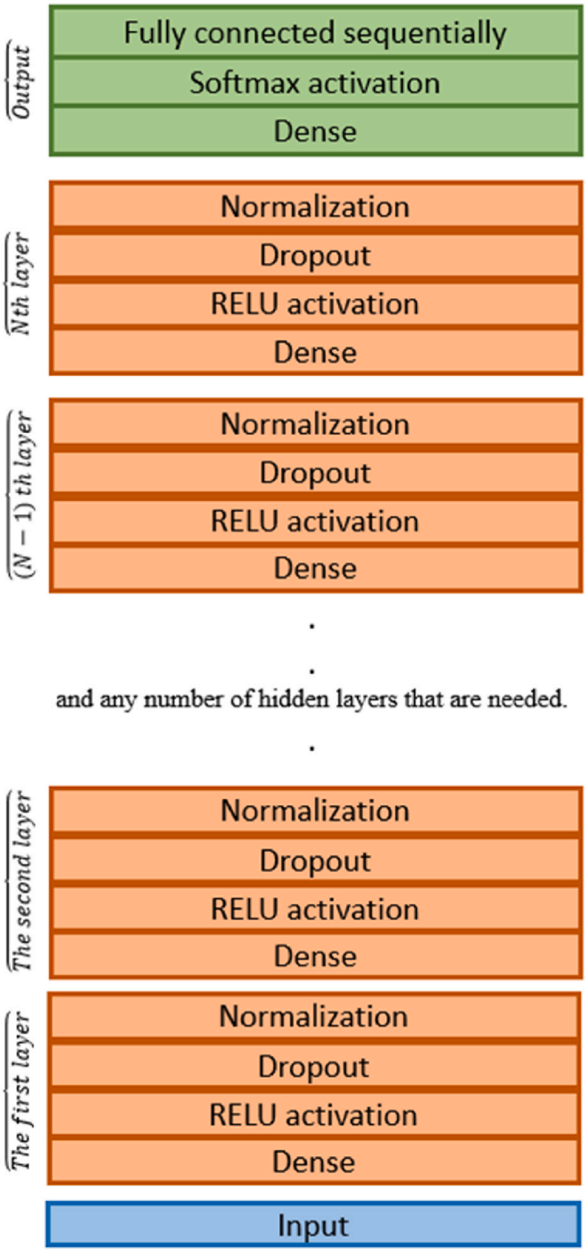

Fig. 6. The general model of deep neural network structure.

$$learning\ rate_t = learning\ rate_{t-1} * 0.099 \tag{12}$$

Neural networks often encounter the issue of overfitting, which causes high accuracy in predicting training data but poor accuracy in predicting test data. This occurs when the model concentrates too much on features unique to the training data. Regularization methods can effectively eliminate overfitting.

Dropout is a crucial regularization method in deep learning. It involves randomly setting some weights to zero, eliminating certain perceptrons and their connections during each epoch. As a result, each iteration produces a unique set of outputs, leading to more accurate information about the training data and improving accuracy when dealing with new data [30]. In this study initially, 50% of the neurons are removed in each layer, and this percentage is ultimately determined based on the neural network's performance.

In a neural network, the distribution of inputs changes every time the previous layer's parameters are updated, which slows down the training process. Therefore, high-quality initialization and a low learning speed are necessary. Batch normalization is a technique used in neural networks to improve convergence rates and performance. It normalizes the input of each layer on every batch, resulting in a Gaussian distribution in neurons [31].

One way to improve optimization is through batch gradient descent, which uses average gradients during training. The mini-batch method is more efficient, dividing samples into smaller categories and reducing variance in parameter updates. This approach requires less storage space than full-batch and results in more consistent outcomes [32].

As a last step in optimization, error optimization is employed. To achieve this, the adaptive gradient algorithm, also referred to as ADAM, is utilized to optimize the neural network swiftly and effectively. It's important to note that the foundation of this algorithm is gradient descent [33].

The number of layers and neurons in a neural network is determined by the training process and the type of data. Fewer layers are needed for OC data, while BW data needs more layers. In the next section, it will be demonstrated that both OC and BW data will require 3 and 5 layers, respectively.

Two widely used regularization techniques for models are L1 and L2. They both involve adding a regularization term to the general cost function to reduce the values of weight matrices and obtain a simpler model [34]. The L1 and L2 methods effectively improve the dispersion of results in the BW data. However, since this scatter is not observed in the OC data, these methods are not used for this data.

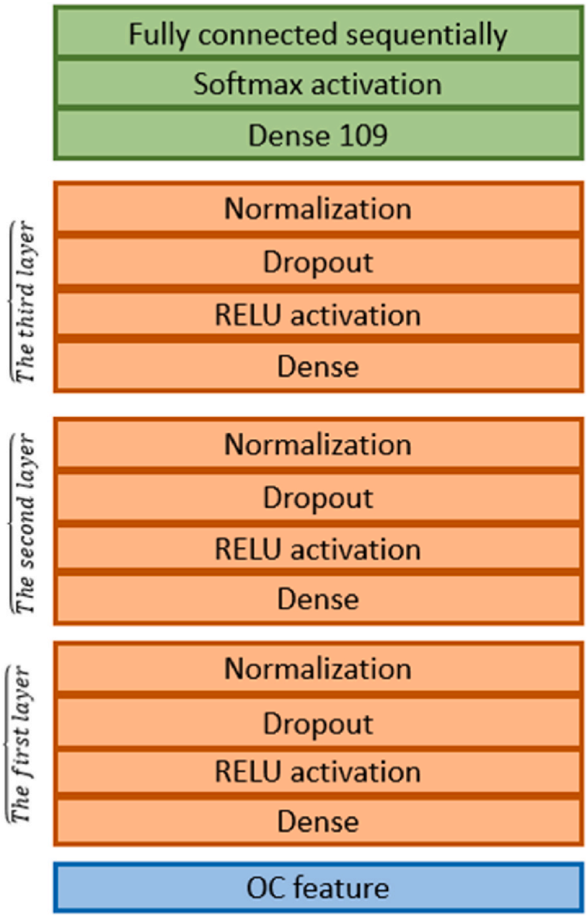

Fig. 7. The deep neural network structure of OC data.

$$Cost\ function = Loss + Regularization\ term \tag{13}$$

When trying to compress a model, L1 is more effective as it sets the weights to zero, while L2 only brings them closer to zero. However, if compression is not the goal, L2 is typically preferred [34]. The λ parameter controls the impact of L2 and is initially set to 0.001. It can be adjusted as needed until the desired results are achieved.

The general model of the neural network is illustrated in Fig. 6. The number of input neurons corresponds to the number of features obtained from PCA. Fig. 6 depicts the arrangement of layers with orange blocks, starting from the first layer to the Nth layer. The number of layers and neurons in the neural network varies depending on the type of data used. Fig. 6 also shows the activation function for each layer. The output is shown with a green block that employs the Softmax activation function. Subsequently, the parameters determined for each data will undergo examination based on the mentioned methods.

In Fig. 7, a 3-layer neural network that utilizes OC data analysis has been shown. Table 1 shows the parameters of the neural network structure, such as the number of layers and neurons, dropout percentage, and learning rate. The batch gradient considers 256 samples. The number of epochs for training data is 500 while utilizing PCA-derived features.

In Fig. 8, a 5-layer neural network that utilizes BW data analysis has been shown. Table 2 shows the parameters of the neural network structure, such as the number of layers and neurons, dropout percentage, and learning rate. The batch gradient considers 256 samples. The number of epochs for training data is 500 while utilizing PCA-derived features.

In this section, a comparison is made between different techniques used in deep neural networks. Each technique is removed from the neural network, and the neural network's parameters are adjusted based on the new conditions. The resulting outcomes are then examined. It is important to note that each technique is removed individually to observe its effect on the neural network, and no two methods are eliminated simultaneously. This comparison is done on the statistical features of BW data.

The initial stage involves utilizing a 5-layer neural network to scrutinize BW statistical features without some of the preprocessing. The neural network structure is outlined in the "Preprocessing effect" column of Table 3. The general design of the neural network structure is displayed in Fig. 6. The input signal for the "Preprocessing effect" neural network is derived from the 0.4–60 Hz range, and a notch filter is applied while no other filters are used.

A 5-layer neural network is used to analyze BW statistical features without normalization. The neural network structure is outlined in the "Batch normalization effect" column of Table 3.

The dropout method is used in the neural network to prevent overfitting. The absence of this method leads to an increase in overfitting. Additionally, the neural network's complexity needs to be reduced due to the over-training of neurons. Therefore, a 4-layer neural network with 300 neurons in each layer is the appropriate choice for this type. The neural network structure is outlined in the "Dropout effect" column of Table 3.

3. Results

The present study aims to enhance the practical application of EEG-based identification by analyzing data with mental preoccupation and creating an appropriate algorithm. It is essential to consider varying levels of focus while processing data rather than solely relying on entirely focused data. This section delves into the outcomes of DNN classification using both OC and BW data. Firstly, the data is preprocessed, and three sets of features are extracted for each dataset. The steps before the classification process are executed using Matlab 9.8, and the classification is accomplished using Python 3.9. The final result of the neural network is an average of 10 repetitions. A computer with a CPU corei 7, 2.60 GHz processor, Windows 11, 16 GB of RAM, and SSD storage are utilized for this study.

A. OC Data

The results obtained from the neural network used for OC data classification (Fig. 7) are shown in Table 4. The details about the neural network's architecture and parameters are presented in Table 1. Additional information about the architecture is presented in the Materials and Methods Section.

It is important to check the accuracy of results, but other performance metrics are also helpful. Table 5 displays these metrics.

B. BW Data

The results obtained from the neural network used for BW data classification (Fig. 8) are shown in Table 6. The architecture and parameters of the neural network are presented in Table 2. Additional information about the architecture is presented in the Materials

Table 1
The parameters of the deep neural network structure for OC data.

		Statistical	Frequency	Wavelet
1	Features	232	240	283
2	Layer	3	3	3
3	Neurons	200	300	350
4	Dropout	40	40	40
5	Learning rate	0.02	0.005	0.01

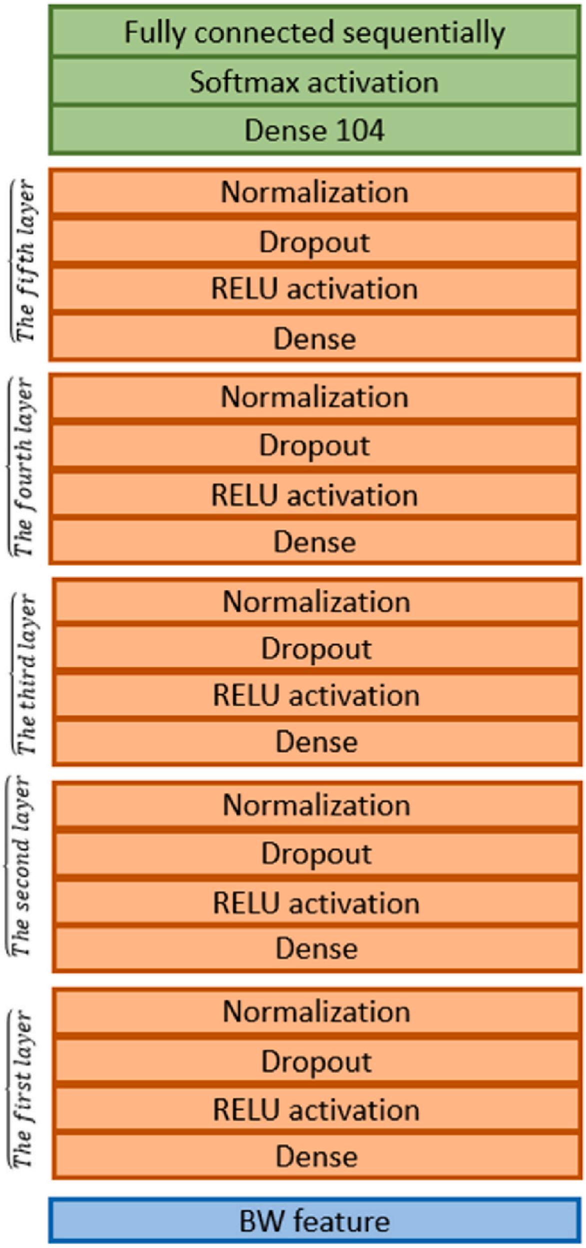

Fig. 8. The deep neural network structure of BW data.

Table 2
The parameters of the deep neural network structure for BW data.

		Statistical	Frequency	Wavelet
1	Features	302	351	365
2	Layer	5	5	5
3	Neurons	450	450	500
4	Dropout	35	35	30
5	Learning rate	0.01	0.01	0.01
6	L2	0.002	0.0015	0.004

Table 3
Neural network parameters, in examining the effect of removing each technique from the proposed neural network model for statistical BW data.

		Preprocessing effect	Batch normalization effect	Dropout effect	L2 effect
1	Features	173	302	302	302
2	Layer	5	5	4	5
3	Neurons	500	400	300	450
4	Dropout	20	30	–	35
5	Learning rate	0.01	0.01	0.01	0.01
6	L2	0.0005	0.001	0.002	–

The L2 method is crucial for achieving convergence, and removing it leads to increased variability in results. A 5-layer neural network to scrutinize BW statistical features without L2 is used. Table 3 shows the neural network structure in the "L2 effect" column.

Table 4
DNN classification results for OC data.

		Statistical	Frequency	Wavelet
1	Performance	96.79 %	99.19 %	98.26 %
2	STD	0.98	0.48	0.73
3	Time(s)	46.14	44.57	43.27

Table 5
Performance metrics of OC data.

	Accuracy (std)	Precision (std)	Recall (std)	F1_score (std)
Statistical	96.79% (0.98)	98.77% (0.70)	99.68% (0.87)	96.90% (0.73)
Frequency	99.19% (0.48)	99.20% (0.63)	99.85% (0.56)	99% (0.66)
Wavelet	98.26% (0.73)	98.30% (0.82)	99.84% (0.84)	98.20% (0.63)

Table 6
DNN classification results for BW data.

		Statistical	Frequency	Wavelet
1	Performance	93.18 %	97.81 %	97.35 %
2	STD	0.72	0.70	0.78
3	Time (s)	124.23	120.86	144.01

and Methods Section. Other performance metrics are presented in Table 7.

Table 8 shows the effect of applied techniques on the data. Refer to Table 3 for neural network configurations. Details are in the Materials and Methods Section. The succeeding section will delve into the consensus regarding the tables mentioned in this part.

C. Consensus of Results

Tables 4 and 6 show that after the frequency features, the wavelet features had the best results. Fig. 9 provides a comparison of the results from Tables 3 and 6 Researchers analyze the frequency characteristics of EEG data in different frequency bands to identify patterns and specific traits, such as attention, memory, decision-making, extroversion, introversion, emotional stability, and impulsivity [35–38]. Therefore, using the unique patterns and features captured through frequency analysis, it is possible to gain a deeper understanding of individual identification through EEG signals.

When working with DNN models, various challenges may arise. These challenges include maintaining high-quality data recordings, the number of subjects, and performing additional calibration steps, all of which can lead to errors.

There are ways to improve DNN accuracy, like Dropout and batch normalization. Simplifying the model with L2 can also help. EEG signals often have noise and artifacts, but preprocessing methods can improve this.

Fig. 10 compares the results from Table 8. Removing filters decreases accuracy by 7.93% in comparison to using all methods. Also, removing the normalization method reduces accuracy by 3.23% and dropout by 1.79%. Removing L2 decreases accuracy by 0.21% but increases scattering by 1.75%. Using the L2 method in the identification process reduces scatter hazards.

4. Discussion

Initially, the OC data was classified. The high results obtained from the data suggest that the presented process is not affected by the eye's position, which is significant for identification algorithms.

Movement, lack of concentration, and the movement decision before identification can impact brain signals. As a result, the brain

Table 7
Performance metrics of BW data.

	Accuracy (std)	Precision (std)	Recall (std)	F1_score (std)
Statistical	93.18% (0.72)	93.40% (0.84)	97.71% (0.99)	93.10% (0.87)
Frequency	97.81% (0.70)	98.00% (0.67)	97.47% (0.74)	97.70% (0.82)
Wavelet	97.35% (0.78)	97.50% (0.71)	98.61% (0.97)	97.20% (0.79)

Table 8
Comparison of results, in examining the effect of removing each technique from the proposed neural network model for statistical BW data.

		Preprocessing effect	Batch normalization Effect	Dropout effect	L2 effect
1	Performance	85.25%	89.95%	91.39%	92.97%
2	STD	1.10	0.89	0.87	2.54
3	Time (s)	115.32	89.32	89.32	117.26

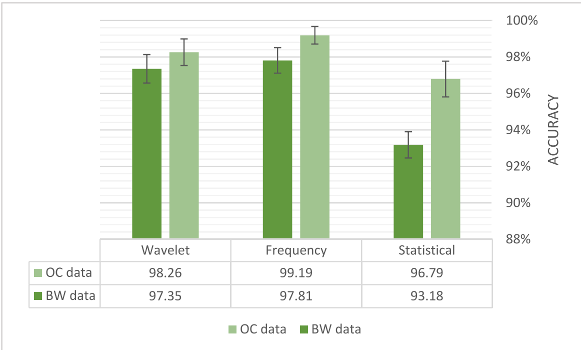

Fig. 9. The results of different feature groups are compared. Error bars show the standard deviation.

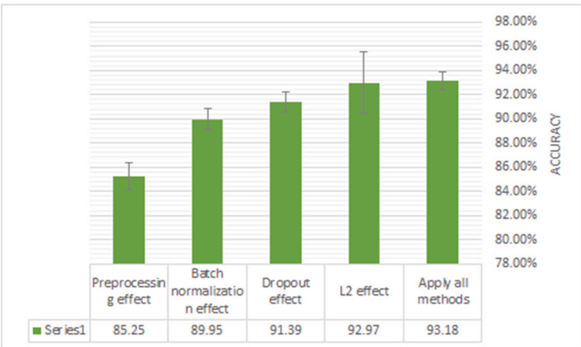

Fig. 10. The effects of removing each technique from the proposed neural network model for statistical BW data are compared. Error bars indicate standard deviation. It is important to note that each technique is removed individually to observe its effect on the neural network, and no two methods are eliminated simultaneously.

data collected during complete rest will be different from the data recorded when the individual is at rest between tasks. The reason for utilizing BW data is to account for varying levels of focus rather than relying solely on data from highly focused individuals.

The use of 64 channels makes EEG-based identification difficult, but reducing the number of channels will be done in future studies. In this article, the main focus is on mental busyness and its effect on results. Hence, the paper aims to develop a suitable algorithm and analyze the BW data.

While this method has some benefits, it also comes with disadvantages. In this research, efforts were made to reduce their disadvantages.

This method requires several preprocessing steps that may remove certain high-frequency components that are not necessarily noise. Consequently, this can lead to the loss of important signal details or features that could be useful for classifying purposes. The selection of features in certain categories may create a bias that affects classification performance. Future research should examine feature interaction. Interpreting DNN classification can be challenging due to its black-box nature, which makes it hard to understand how decisions are made. This lack of interpretability can limit the insights we can gain from analyzing such models. Moreover, it is crucial to be mindful of overfitting, particularly when dealing with high-dimensional data.

The method proposed in this paper has several advantages. The approach combines preprocessing, feature extraction, dimensionality reduction, normalization, and DNN classification for brain signals, providing a comprehensive solution. Accurately extracting relevant features from brain signals is crucial, and the preprocessing phase plays a vital role in achieving this goal. Eliminating unwanted noises and artifacts can greatly enhance the signal-to-noise ratio and enable more accurate and meaningful feature extraction. This is particularly important in EEG-based identity recognition, where capturing subtle patterns and features in brain signals is necessary to differentiate between individuals. Utilizing DNN to analyze brain signals offers numerous advantages. These computational models have shown exceptional performance in various fields and can improve the accuracy of classification by learning complex patterns. The techniques described in the Materials and Methods section have been implemented to prevent overfitting and aid in resulting convergence.

The process of identification depends on time. While training DNN networks may take a while, the testing phase is usually short. Tables 4 and 6 provide the duration of the test phase, depending on the system's features. By utilizing robust systems and decreasing the number of channels, time can be further reduced.

As mentioned earlier, EEG-based identification can be categorized into two types: rest-based and task-based. This section will analyze the outcomes discussed in previous articles. Table 9 displays the results from the OC data in a resting state. The findings in this article show better performance compared to the other studies. In Table 10, BW data is considered as imagery-movement tasks and movement task data. As can be seen in this table, the obtained result in this case is also superior to other articles.

5. Conclusions and future work

In this paper, an appropriate algorithm is discussed that considers mental busyness while analyzing data. The data of 109 participants underwent a five-step preprocessing in the initial stage, followed by computation of statistical properties, frequency, and wavelet.

The performance of the data was evaluated using DNN classifications. This involved the application of normalization, dropout, L2, and batch gradient methods to improve neural network performance. The effectiveness of these methods in improving the neural network's performance and achieving convergence of results was demonstrated.

Based on the OC data, the algorithm displayed low sensitivity to eye conditions and showed a 99.19% accuracy for frequency features using DNN. After identifying the optimal algorithm, it was used to assess the performance of BW data. The neural network used for this data was notably more complex than that of OC data. The highest success rate achieved was 97.81% for frequency features on BW data. Furthermore, it was clear that the DNN approach performed better than shallow methods such as SVM. The neural network method is superior to other classification methods because it can efficiently learn and extract complicated patterns from the data. Neural networks are non-linear models by nature, which enables them to capture intricate relationships and non-linear dependencies present in the data. This flexibility allows them to model complex patterns more accurately than linear classifiers. Additionally, neural networks can adapt and learn from new data, making them suitable for scenarios where patterns change over time.

In the case of EEG identification, an important factor is the ease of recording information. The performance of a DNN for EEG analysis can be affected both positively and negatively by increasing the number of channels. To summarize, increasing the number of channels in an EEG can enhance feature extraction, improve spatial resolution, and potentially improve DNN performance. However, it can also introduce challenges related to data quality, higher dimensionality, and data collection limitations. Therefore, future research aims to pinpoint specific areas of the brain with personality characteristics and adjust channels accordingly. Generally, identification is a process where even a small increase in results can be significant, so it is important to implement methods to improve the performance of the identification process.

Ethical approval

Not applicable.

Table 9
Biometric systems using resting-state EEG patterns.

Reference/ Year	Channels	Subjects	Features	Methodology	Performance
[4], 2015	14	16	power spectral density (PSD)	SVM	93%
[10], 2016	5	9	multiscale shape description (MSD), multiscale wavelet packet statistics (WPS) and multiscale wavelet packet energy statistics (WPES)	SVM	87.3%
[11], 2016	64	109	phase locking value (PLV)	core sub-network	94.3%
[15], 2020	1	46	statistical, frequency, and wavelet	SVM	95.48%
This paper	64	109	statistical, frequency, and wavelet	DNN	99.19%

Table 10
Biometric systems using mental tasks.

Reference/ Year	Channels	Subjects	Mental Activity	Features	Methodology	Performance
[12], 2016	32	109	imagery-movement	Autoregressive model (AR)	Flower pollination	93%
[13], 2017	32	16	self-face and non-self-face images	spectral entropy, approximate entropy, sample entropy, and fuzzy entropy	SVM	87.3%
[14], 2019	9	8	SSVEP	Low-frequency SSVEP components	convolutional neural network (CNN)	94.3%
[16], 2020	11	40	auditory evoked potential	Auditory Evoked Potentials (AEPs) in different frequency bands	LDA	95.48%
[17], 2020	64	109	imagery-movement	Power Spectral Density (PSD) of the first 5 Intrinsic Mode Functions (IMFs)	Inception-like NN-SVM decoder	93.40%
This paper	64	109	Waiting between imagery-movement	statistical, frequency, and wavelet	DNN	97.81%

Funding

Not applicable.

Availability of data and materials

Data analyzed in the current research were a re-investigation of existing data, which are openly available on the Physionet website: https://physionet.org/.

CRediT authorship contribution statement

Yasaman Akbarnia: Writing – review & editing, Writing – original draft, Methodology, Formal analysis. **Mohammad Reza Daliri:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Yang, N. Xiong, A.V. Vasilakos, Two-stage enhancement scheme for low-quality fingerprint images by learning from the images, IEEE Transactions on human-machine systems 43 (2) (2012) 235–248, https://doi.org/10.1109/TSMCC.2011.2174049.
[2] P. Mishra, Development of Biometric Verification Algorithm Using Electroencephalogram EEG, Thapar Institute of Engineering and Technology, 2016 a reservoir of Indian theses @ INFLIBNET, http://hdl.handle.net/10603/301566.
[3] R. Palaniappan, D.P. Mandic, Biometrics from brain electrical activity: a machine learning approach, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 738–742, https://doi.org/10.1109/TPAMI.2007.1013.
[4] Y. Dai, X. Wang, X. Li, Y. Tan, Sparse EEG compressive sensing for web-enabled person identification, Measurement 74 (2015) 11–20, https://doi.org/10.1016/j.measurement.2015.07.008.
[5] X. Zhang, L. Yao, S.S. Kanhere, Y. Liu, T. Gu, K. Chen, Mindid: person identification from brain waves through an attention-based recurrent neural network, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2 (3) (2018) 1–23, https://doi.org/10.1145/3264959.
[6] J. Püschel, H.H. Stassen, G. Bomben, C. Scharfetter, D. Hell, Speaking behavior and speech sound characteristics in acute schizophrenia, J. Psychiatr. Res. 32 (2) (1998) 89–97, https://doi.org/10.1016/S0022-3956(98)00046-6.
[7] B.P. Zietsch, J.L. Hansen, N.K. Hansell, G.M. Geffen, N.G. Martin, M.J. Wright, Common and specific genetic influences on EEG power bands delta, theta, alpha, and beta, Biol. Psychol. 75 (2) (2007) 154–164, https://doi.org/10.1016/j.biopsycho.2007.01.004.
[8] L. De Gennaro, F. Fratello, C. Marzano, F. Moroni, G. Curcio, D. Tempesta, P.M. Rossini, Cortical plasticity induced by transcranial magnetic stimulation during wakefulness affects electroencephalogram activity during sleep, PLoS One 3 (6) (2008) e2483, https://doi.org/10.1371/journal.pone.0002483.
[9] S.K. Yeom, H.I. Suk, S.W. Lee, Person authentication from neural activity of face-specific visual self-representation, Pattern Recogn. 46 (4) (2013) 1159–1169, https://doi.org/10.1016/j.patcog.2012.10.023.
[10] M.K. Bashar, I. Chiaki, H. Yoshida, Human identification from brain EEG signals using advanced machine learning method EEG-based biometrics, in: 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), IEEE, 2016, December, pp. 475–479, https://doi.org/10.1109/IECBES.2016.7843496.
[11] A. Crobe, M. Demuru, L. Didaci, G.L. Marcialis, M. Fraschini, Minimum spanning tree and k-core decomposition as measures of subject-specific EEG traits, Biomedical Physics & Engineering Express 2 (1) (2016) 017001, https://doi.org/10.1088/2057-1976/2/1/017001.
[12] D. Rodrigues, G.F. Silva, J.P. Papa, A.N. Marana, X.S. Yang, EEG-based person identification through binary flower pollination algorithm, Expert Syst. Appl. 62 (2016) 81–90, https://doi.org/10.1016/j.eswa.2016.06.006.
[13] Z. Mu, J. Hu, J. Min, J. Yin, Comparison of different entropies as features for person authentication based on EEG signals, IET Biom. 6 (6) (2017) 409–417, https://doi.org/10.1049/iet-bmt.2016.0144.
[14] T. Yu, C.S. Wei, K.J. Chiang, M. Nakanishi, T.P. Jung, EEG-based user authentication using a convolutional neural network, in: 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE, 2019, March, pp. 1011–1014, https://doi.org/10.1109/NER.2019.8716965.
[15] R. Zhang, B. Yan, L. Tong, J. Shu, X. Song, Y. Zeng, Identity authentication using Portable electroencephalography signals in resting states, IEEE Access 7 (2019) 160671–160682, https://doi.org/10.1109/ACCESS.2019.2950366.
[16] S.N.A. Seha, D. Hatzinakos, EEG-based human recognition using steady-state AEPs and subject-unique spatial filters, IEEE Trans. Inf. Forensics Secur. 15 (2020) 3901–3910, https://doi.org/10.1109/TIFS.2020.3001729.

[17] U. Barayeu, N. Horlava, A. Libert, M. Van Hulle, Robust single-trial EEG-based authentication was achieved with a 2-stage classifier, Biosensors 10 (9) (2020) 124, https://doi.org/10.3390/bios10090124.
[18] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, BCI2000: a general-purpose brain-computer interface (BCI) system, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 51 (6) (2004) 1034–1043, https://doi.org/10.1109/TBME.2004.827072.
[19] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P.C. Ivanov, R. Mark, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation [Online] 101 (23) (2000), https://doi.org/10.1161/01.CIR.101.23.e215 e215–e220.
[20] A.G. Reddy, S. Narava, Artifact removal from EEG signals, Int. J. Comput. Appl. 77 (13) (2013), https://doi.org/10.5120/13543-1175.
[21] K.A. Ludwig, R.M. Miriani, N.B. Langhals, M.D. Joseph, D.J. Anderson, D.R. Kipke, Using a common average reference to improve cortical neuron recordings from microelectrode arrays, J. Neurophysiol. 101 (3) (2009) 1679–1689, https://doi.org/10.1152/jn.90989.2008.
[22] D. Wang, D. Miao, G. Blohm, Multi-class motor imagery EEG decoding for brain-computer interfaces, Front. Neurosci. 6 (2012) 151, https://doi.org/10.3389/fnins.2012.00151.
[23] D. Zhang, Wavelet transform, in: Fundamentals of Image Data Mining, Springer, Cham, 2019, pp. 35–44, https://doi.org/10.1007/978-3-030-17989-2_3.
[24] Z. Zhang, Spectral and time-frequency analysis, in: EEG Signal Processing and Feature Extraction, Springer, Singapore, 2019, pp. 89–116, https://doi.org/10.1007/978-981-13-9113-2_6.
[25] S. Sanei, J.A. Chambers, EEG Signal Processing, John Wiley & Sons, 2013. ISBN: 1118691237, 9781118691236.
[26] S. Karamizadeh, S.M. Abdullah, A.A. Manaf, M. Zamani, A. Hooman, An overview of principal component analysis, J. Signal Inf. Process. 4 (3B) (2013) 173, https://doi.org/10.4236/jsip.2013.43B031.
[27] M.W. Browne, Cross-validation methods, J. Math. Psychol. 44 (1) (2000) 108–132, https://doi.org/10.1006/jmps.1999.1279.
[28] M. Ojala, G.C. Garriga, Permutation tests for studying classifier performance, J. Mach. Learn. Res. 11 (6) (2010), https://doi.org/10.1109/ICDM.2009.108.
[29] M.A. Nielsen, in: Neural Networks and Deep Learning, vol. 25, Determination press, San Francisco, CA, USA, 2015, https://doi.org/10.4236/gep.2018.65003.
[30] J. Kukačka, V. Golkov, D. Cremers, Regularization for Deep Learning: A Taxonomy, 2017, https://doi.org/10.48550/arXiv.1710.10686 arXiv preprint arXiv:1710.10686.
[31] P. Luo, X. Wang, W. Shao, Z. Peng, Towards Understanding Regularization in Batch Normalization, 2018 00846, https://doi.org/10.48550/arXiv.1809.00846 arXiv preprint arXiv:1809.
[32] S. Ruder, An Overview of Gradient Descent Optimization Algorithms, 2016, https://doi.org/10.48550/arXiv.1609.04747 arXiv preprint arXiv:1609.04747.
[33] Z. Zhang, Improved Adam optimizer for deep neural networks, in: 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), IEEE, 2018, June, pp. 1–2, https://doi.org/10.1109/IWQoS.2018.8624183.
[34] C. Cortes, M. Mohri, A. Rostamizadeh, L2 Regularization for Learning Kernels, 2012, https://doi.org/10.48550/arXiv.1205.2653 arXiv preprint arXiv:1205.2653.
[35] J.I. Schmidtke, W. Heller, Personality, affect and EEG: predicting patterns of regional brain activity related to extraversion and neuroticism, Pers. Indiv. Differ. 36 (3) (2004) 717–732, https://doi.org/10.1016/S0191-8869(03)00129-6.
[36] J. Jacobs, G. Hwang, T. Curran, M.J. Kahana, EEG oscillations and recognition memory: theta correlates of memory retrieval and decision making, Neuroimage 32 (2) (2006) 978–987, https://doi.org/10.1016/j.neuroimage.2006.02.018.
[37] F.Q. Annisa, E. Supriyanto, S. Taheri, Personality dimensions classification with EEG analysis using support vector machine, in: 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), IEEE, 2020, December, pp. 79–82, https://doi.org/10.1109/ISRITI51436.2020.9315507.
[38] R. Zhang, Z. Wang, Y. Liu, The research of EEG feature extraction and classification for subjects with different organizational commitments, in: MATEC Web of Conferences, vol. 355, EDP Sciences, 2022 03042, https://doi.org/10.1051/matecconf/202235503042.