



# On the Minimal Amount of EEG Data Required for Learning Distinctive Human Features for Task-Dependent Biometric Applications

Carlos Gómez-Tapia\*, Bojan Bozic and Luca Longo\*

Artificial Intelligence and Cognitive Load Research Lab, Applied Intelligence Research Centre, School of Computer Science, Technological University Dublin, Dublin, Ireland

Biometrics is the process of measuring and analyzing human characteristics to verify a given person's identity. Most real-world applications rely on unique human traits such as fingerprints or iris. However, among these unique human characteristics for biometrics, the use of Electroencephalogram (EEG) stands out given its high inter-subject variability. Recent advances in Deep Learning and a deeper understanding of EEG processing methods have led to the development of models that accurately discriminate unique individuals. However, it is still uncertain how much EEG data is required to train such models. This work aims at determining the minimal amount of training data required to develop a robust EEG-based biometric model (+95% and +99% testing accuracies) from a subject for a task-dependent task. This goal is achieved by performing and analyzing 11,780 combinations of training sizes, by employing various neural network-based learning techniques of increasing complexity, and feature extraction methods on the affective EEG-based DEAP dataset. Findings suggest that if Power Spectral Density or Wavelet Energy features are extracted from the artifact-free EEG signal, 1 and 3 s of data per subject is enough to achieve +95% and +99% accuracy, respectively. These findings contribute to the body of knowledge by paving a way for the application of EEG to real-world ecological biometric applications and by demonstrating methods to learn the minimal amount of data required for such applications.

**Keywords:** biometrics, EEG, feature extraction, machine learning, deep learning, graph neural networks

## OPEN ACCESS

### Edited by:

Giuseppe Placidi,  
University of L'Aquila, Italy

### Reviewed by:

Maciej Szymkowski,  
Bialystok University of Technology,  
Poland

Vinay Kumar,  
Thapar Institute of Engineering &  
Technology, India

### \*Correspondence:

Carlos Gómez-Tapia  
carlos.g.tapia@myTUDublin.ie  
Luca Longo  
luca.longo@tudublin.ie

**Received:** 28 December 2021

**Accepted:** 19 April 2022

**Published:** 10 May 2022

### Citation:

Gómez-Tapia C, Bozic B and Longo L (2022) On the Minimal Amount of EEG Data Required for Learning Distinctive Human Features for Task-Dependent Biometric Applications. *Front. Neuroinform.* 16:844667. doi: 10.3389/fninf.2022.844667

## 1. INTRODUCTION

Among the different techniques used for measuring brain activity, Electroencephalography (EEG) is a method used for measuring voltage fluctuation inside the electrical field generated by a subject's brain. The sampling process involves placing surface electrodes onto a subject's scalp, capturing brain activity at different regions. The resulting EEG signals are non-stationary, meaning frequency values along time are not constant but variable, making them unpredictable. Recent advances in Deep Learning (DL) (Scarselli et al., 2008; Goodfellow et al., 2014; Vaswani et al., 2017) have allowed for new and improved applications of EEG signals. These applications include but are not limited to: EEG-based biometric systems (DelPozo-Banos et al., 2015), Brain-Computer Interfaces (BCIs) (Vaid et al., 2015), emotion recognition systems (Jenke et al., 2014), alertness analysis (Subasi, 2005) and medical diagnosis and progression assessment for neurological diseases such as Alzheimer's

(Cassani et al., 2018) or epilepsy (Acharya et al., 2015). These systems are often composed of an automatic feature extraction layer that extracts representative features from the raw EEG signal, and a classifier for fitting a specific target feature. For example, Wilaiprasitporn et al. (2019) used a combination of Convolutional Neural Networks along with different type of temporal layers to learn distinctive features, and a fully connected layer for the purpose of classification. Similarly, Ullah et al. (2018) used a model composed of several 1-D convolutional layers for feature extraction and denoising to obtain representative features subsequently handled by a fully connected layer with a majority voting scheme.

In detail, this research focuses on biometrics, a process consisting of measuring and analyzing unique physical characteristics from humans to authenticate their unique identity. Among the methods employed for such a task, EEG has proven to be a robust biometrics method (Jayarathne et al., 2017). As pointed by Campisi and La Rocca (2014), EEG poses several advantages when compared to traditional exposed biometric modalities such as fingerprint or iris scanning. Namely: The advantage of being more robust against spoofing attacks (Revett, 2012), given the technological complexity of synthetically generating brain signals replicating the response for a specific individual. The advantage of being universal, meaning the sampling process is valid for anyone with no brain pathological conditions. Lastly, EEG poses the advantage of the aliveness detection problem not being present, meaning we can assume that the user is alive to produce brain signals sensor can read. This assumption does not necessarily hold in other biometric methods. There are two main applications for EEG in biometric systems: Person identification (PI), identifying an individual from a group of known subjects, and person authentication (PA), accepting or denying the identity of a particular individual. Both tasks' objective is to extract features with discriminating properties from EEG signals that a statistical model can later classify. Depending on the task subjects perform while their EEG recordings are taking place, it is possible to divide biometric systems into two groups, namely task-dependent (Wilaiprasitporn et al., 2019) and task-independent (Kong et al., 2018). Training a task-dependent system often means presenting the model data where all subjects perform a particular task. Whereas task-independent systems use data sampled from various tasks, making them more robust at handling unseen tasks at inference time. This work is centered around using EEG signals applied to task-dependent PI and is devoted to answering the following research question:

- **RQ:** What is the minimal amount of data required to train an affective EEG-based person identification model for biometric applications with and without an explicit feature extraction method?

The structure of this document is as follows: Section 2 provides a brief literature review on the different methods used for EEG-Based Person Identification. Section 3 explains the design and methodology, with reproducibility and replicability in mind. Section 4 discusses the results and findings and finally Section 5 summarizes these informing possible future directions.

## 2. RELATED WORK

There exists evidence suggesting that different subjects produce different EEG responses (Marcel and Millán, 2007), with studies dating as back as the 1970's suggesting the uniqueness in EEG responses depends on each subject genetic information (Vogel, 1970; Anokhin et al., 1992). This unique property from EEG data makes it suitable for biometrics applications. However, building such applications has proven challenging to build and deploy given the non-stationary nature of EEG signals and the high quantity of noise generated during recording by, for example, muscle movements, eye blinking, electrode displacement. Additionally, there has been evidence showing that the EEG responses from an individual may vary depending on their emotions. For example, EEG responses may vary if the participant is emotionally attached to the watched person (Koelstra et al., 2009). EEG analysis often requires the extraction of high-level features fed into statistical models that automatically learn to find hidden patterns within the features. There are several different feature extraction techniques present in the literature. The feature extraction methods that are applied most often to this domain are Power Spectral Density (PSD) (Riera et al., 2007), Auto-Regressive (AR) model coefficients (Mohammadi et al., 2006; Zivot and Wang, 2006), Discrete Wavelet transform (DWT) (Guo et al., 2009; Murugappan et al., 2009), and Hilbert-Huang Transform (HHT) (Li et al., 2009). Among these techniques, wavelet-based feature extraction stands up given its performance at characterizing non-stationary data.

The available literature on person identification from EEG features dates as far as the 90's [see (Poulos et al., 1999a,b,c)]. Scholars used sets of features obtained from an AR model and a set of different Machine Learning (ML) classifiers obtaining accuracy scores ranging from 80 to 100% at classifying unique individuals from a pool of four total subjects. Mohammadi et al. (2006) trained a competitive neural network model with features obtained from an AR. In this case, scholars performed experiments training models for single-channel and multi-channel settings with up to three channels. They collected 24 s of EEG readings for 10 participants and used 15 s to train a model and a total of 24 s for testing. Results demonstrated that it was possible to obtain accuracies ranging from 80% up to 100% by using this technique on a larger pool of subjects. Brigham and Kumar (2010) used coefficients obtained for each channel from an AR model along with a Support Vector Machine (SVM) for PI. The scholars obtained testing classification accuracies close to 99% at identifying subjects from a pool of 120 subjects, suggesting that it is possible to obtain separable features from each individual. Shedeed (2011) used a combination of features obtained from applying the Discrete Fourier Transform and the Discrete Wavelet Transform followed by a voting scheme to retain the most discriminating features, which are input to a multi-layer perceptron classifier. However, their dataset was composed solely of three subjects, casting doubt on the generalization ability of their model. Thomas and Vinod (2018) used PSD features obtained from a single channel from the gamma ( $\gamma$ , 30–50 Hz) band along with simple correlation-based matching operations on a dataset composed of 109 subjects with

an error rate of 0.0196. These results suggest that it is possible to apply a non-parametric operation over the raw EEG signal that maps input data onto a set of discriminative features for each subject.

Recent advances in DL have allowed the development of models with more expressing capabilities. Unlike approaches using hand-crafted features, DL models are often composed of two parts: a parametric feature extraction block and a classifier. The feature extraction block aims to learn representative features from the raw input signal automatically. The classifier uses the extracted features as input and predicts a target value such as a specific subject identity. Automatic feature extraction layers use spatial layers to reduce the dimensionality of the input data while trying to maintain informative features. This process usually consists of employing a combination of convolutional and pooling layers (Das et al., 2017, 2018). The input dimensionality of the data is usually not very large, given that there is a limited number of electrodes. The small input size leads to convolutional models with a relatively small number of layers, often ranging from 1 to 5 (Maiorana, 2020). Özdenizci et al. (2019) used a dataset composed of 128 subjects with readings obtained from different sessions and proposed an adversarial network from an invariant representation-learning perspective. Their goal was to create features that would be invariant between different recording sessions. Their approach consisted of an encoder convolutional layer to encode the raw signals and two classification layers. One of these layers is the identifier that predicts the correct person ID. The other one is the adversary layer that tries to predict the recording session ID. The assumption was that the encoder obtained features that the identifier could analyse but the adversary network could not, which led to session-invariant features. Additionally to the spatial layers, some approaches consider temporal layers to include temporal information in the features. Recurrent Neural Networks often include a combination of spatial and temporal layers (Chen et al., 2020). Das et al. (2019) proposed a model based on a combination of spatial layers (Convolutional) and temporal layers by employing Long Short Term Memory (LSTM). Results pointed to accuracies close to 100% on a pool of 109 subjects using the raw data from recordings of 64 channels. Wilaiprasitporn et al. (2019) proposed an alternative method using Gated Recurrent Units instead of the LSTM, achieving 99.9–100% test accuracy on the DEAP dataset (Koelstra et al., 2011) using all available channels. They also achieved 99.17% test accuracy, reducing the number of channels from 32 to 5, using their Frontal and Parietal (F3, F4, Fz, F7, and F8) configuration. Chen et al. (2020) demonstrated that it was possible to train models that automatically learn feature extraction techniques from the raw signal using spatial and temporal layers. These rich features are then injected into a classifier, where the final target feature is the subject ID. They achieved 96% test accuracy on a pool of 157 subjects sampled from four different experiments, demonstrating their model's ability to generalize.

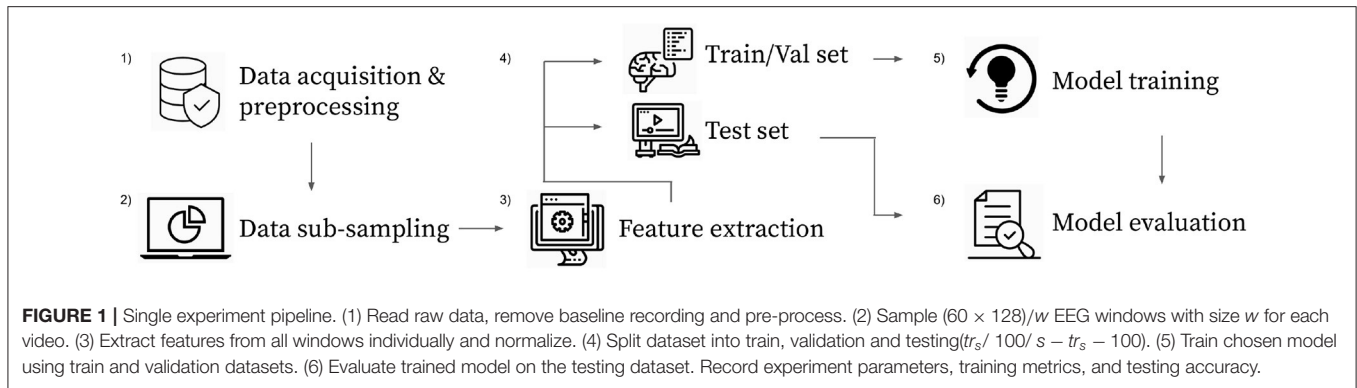
The recent surge in interest in Graph Neural Network (GNN) architectures has also affected EEG-based applications. Applying GNNs to domains where elements and relationships exist has proven to be very effective. These applications

include but are not limited to traffic flow prediction, point cloud classification, and text classification (Cui et al., 2019; Yao et al., 2019; Shi and Rajkumar, 2020). These systems also pose the advantage of implicitly stating the relationships between the different elements, making them potentially more useful for explainability purposes (Pope et al., 2019; Vilone and Longo, 2020, 2021a,b). The use of graphs for EEG-based applications remains underexplored. Zhong et al. (2020) used a Graph Convolutional Network model for EEG-based emotion recognition obtaining results comparable to other non-graph-based methods. Wang et al. (2020) used a graph to map brain functional connectivity for Participant Classification without making use of message-passing Graph Neural Networks. Their work consists of creating brain connectivity networks using multi-channel settings, node centrality, and global network metrics to compute features for the different nodes composing the graph. They prove that generated networks have significant inter-individual distinctiveness, making them suitable for biometric applications.

As these studies suggest, it is possible to obtain meaningful distinguishable patterns of behavior, as gathered by EEG signals, of a specific user, for biometric applications, with current approaches demonstrating error rates close to 0% (Wilaiprasitporn et al., 2019; Seha and Hatzinakos, 2021). However, these approaches rely on datasets composed of several minutes of recordings for each participant, sampling significant EEG windows or using a high number of training samples. Neglecting that sampling data for each participant is an expensive process that could burden real-world scenarios. Carrión-Ojeda et al. (2019) performed a similar study to find out the smallest EEG window size for a biometric system to perform subject identification. The researchers used a fixed amount of training samples per participant, leaving open the question of the minimum total time in seconds required for a PI system to be trained. Similarly, Seha and Hatzinakos (2021) developed a PI model with session-invariant features. Researchers emphasize using smaller windows than previous approaches while obtaining better results. However, they also used most of their available data for training the model, still not focusing on the total training dataset size but solely on the size of the epochs. These reasons motivated the focus on this work to provide hindsight on the minimal amount of affective-based EEG data required from each subject to train task-dependent biometric models. This minimal amount is a combination of the number of samples used for training and the window size for each sample. An exhaustive experiment have been designed, and described in the next section.

### 3. DESIGN AND METHODOLOGY

This research work is devoted to understanding the minimal EEG data required to identify unique discriminative person-specific features. In order to tackle this goal, the experiment includes a combination of different training data sizes, which are defined based on the size of the sampled EEG windows and the number of training samples available per participant. In detail,



these experiments include the training of models by employing several training data sizes along with different machine learning techniques, including a *Logistic Regressor* (LR), *Multi-Layer Perceptron* (MLP), *1-D Convolutional Neural Network* (CNN), and a *Graph Convolutional Network* (*Graphconv*) and a set of feature extraction methods, namely *Power Spectral Density* (PSD) and *Wavelet Energy* (WE). **Figure 1** depicts the experimental pipeline for a given configuration. Each experimental configuration is used 10 times with different, randomly sampled data to obtain results distributions. Each pipeline aims to train and evaluate a classification model that can perform Person Identification given a sub-sampled window of EEG recordings. Training a discriminative model with data for a pool of  $p$  participants that automatically learns to differentiate input features depending on the participant from whom the features have been sub-sampled. More formally, the goal is to approximate a parametric function  $f_\sigma$  that maps an input feature matrix  $\mathbf{X}$  to a vector of predicted probabilities  $\mathbf{p}$  with one entry per known participant.  $\mathbf{p}_i$  represents the probability of the input features belonging to the subject with ID  $i$ . All together we can represent it as  $f_\sigma(\mathbf{X} \in \mathbb{R}^{n \times d}) \rightarrow \mathbf{p}_s \in [0, 1]^p$  where  $n$  indicates the number of EEG channels used,  $d$  depends on the feature extraction method and denotes the number of obtained features per channel and  $p$  is the total number of participants. Models are trained and validated on a randomly sub-sampled portion of the complete dataset, where there are  $tr_s$  training samples and 100 validation samples for each user. The evaluation process for a trained model consists of making predictions on the remaining unseen data and measuring classification accuracy. In detail, classification is correct when the highest value for a predicted probability vector  $\mathbf{p}_s$  matches the ID of the subject the recordings were sampled from.

We hypothesize that by running a set of experimental configurations and evaluating the results, it is possible to provide an answer to the research question stated in Section 1. The overall experiment, consisting of the different configuration settings, is performed to test the following hypothesis:

- **H1:** IF different affective EEG-based person identification models are trained using a variable dataset size (number of training samples and window sizes) and pre-process data extracting wavelet energy or power spectral density THEN it exists a minimum amount of data that can lead these

models to achieve 95% and 99% of accuracy, and this amount is significantly lower than the amount associated to the same models trained without the pre-processing methods.

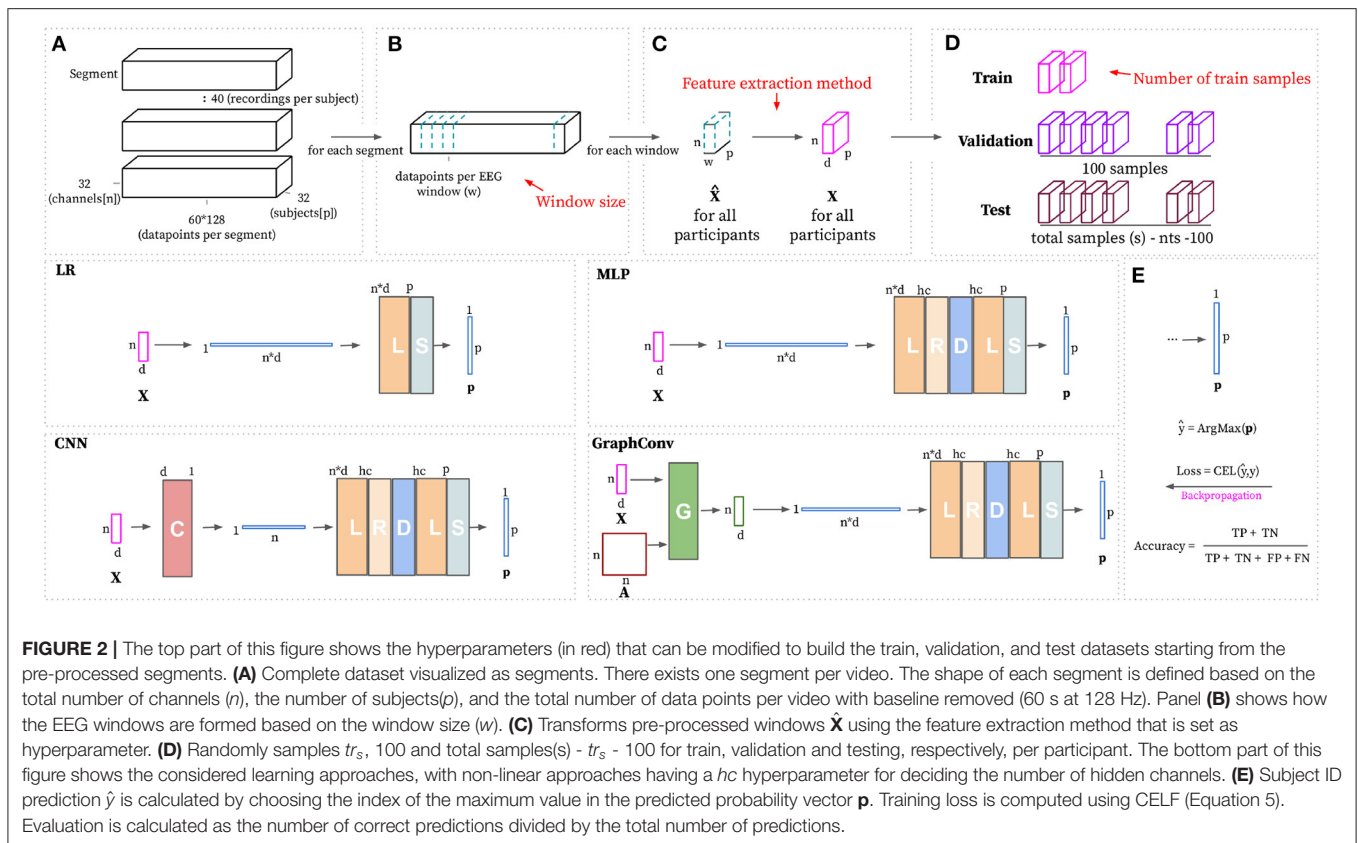
In the context of training person identification models, the minimum required amount of data is defined based on the total number of train EEG windows used ( $tr_s$ ) and the size for each window in seconds ( $w/128$ ). The minimal amount of data and the best learning approach and feature extraction combination can be determined by comparing the result distributions and picking the experimental configuration able to obtain +95% and +99% accuracy on unseen data with the least amount of total training data. Moreover, performing a Mann–Whitney  $U$ -test over the results distributions where training data time is the same but the window size differs would answer whether it is better to have fewer oversized windows or many smaller ones.

### 3.1. Data Acquisition and Pre-processing

The experiment uses the dataset for emotion analysis using EEG, physiological, and video signals (DEAP) (Koelstra et al., 2011). It consists of a set of EEG signals and physiological and face information from 32 different participants, recorded as they watched 40 one-minute music videos meant to trigger different kinds of emotions, making this dataset task-dependent in general and affective-based in particular. The DEAP dataset poses the advantage of being sampled under naturalistic, ecological conditions. Instead of evoking a certain kind of response in a short time window, the DEAP dataset provides a continuous stream of EEG data that better resembles a real-world environment.

Thirty-two electrodes ( $n = 32$ ) were positioned using the international 10-20 system. The current research study only considers the EEG signals. The DEAP dataset has a total of 1,280 segments (32 participants  $\times$  40 videos), where each segment consists of 32 channel readings, each with 3 s of baseline signal followed by the main recordings 60 s. The pre-processing stage consists of applying a set of pre-processing techniques to the original EEG recordings, including:

- Downsample to 128 Hz
- High-pass filter to 4.0 Hz
- Low-pass filter to 45.0 Hz.



A band-pass filter is applied in order to extract the main frequencies of EEG human waves (Abhang et al., 2016) (Theta [4–8 Hz], Alpha [8–12 Hz], Beta [12–35 Hz], Gamma), dismissing the Delta (0.5–4 Hz) band due to it being associated with sleep activity (Amzica and Steriade, 1998). Non-parametric down-sampling is used to compress the data size. The first 3 s representing baseline recordings are removed from the rest of the signal and a sliding, non-overlapping window of varying size  $w$  is applied in order to split each segment into  $(60 * 128)/w$  windows (see **Figure 2B**). After the pre-processing steps, the total number of samples  $s$  is equal to 32 (participants)  $\times$  40 (videos)  $\times$   $(60 * 128)/w$  (windows per segment). Each pre-processed window  $\hat{X} \in \mathbb{R}^{n \times w}$  consists of readings for  $n$  channels with  $w$  readings for each of them. The DEAP dataset performed its sampling with the following electrodes: [Fp1, AF3, F7, F3, FC1, FC5, T7, C3, CP1, CP5, P7, P3, Pz, PO3, O1, Oz, O2, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, Fp2, Fz, Czz].

### 3.2. Feature Extraction

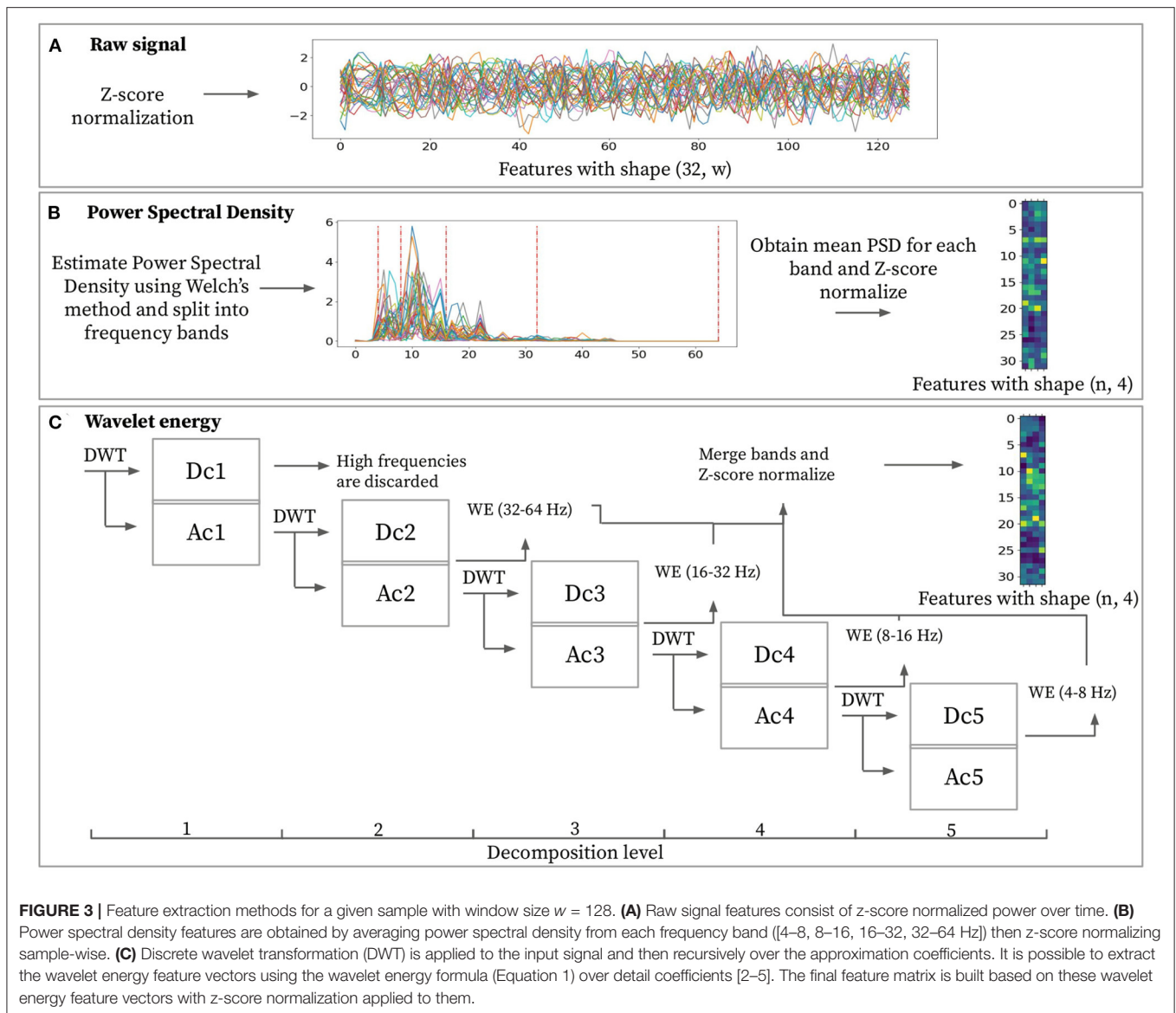
All windows sub-sampled from the original segments are input to the feature extraction block. This block extracts a set of discriminating normalized features  $\mathbf{X}$  from a pre-processed window  $\hat{X}$  (see **Figure 2C**). This research work considers two feature extraction methods: Power Spectral Density (PSD) and Wavelet Energy (WE). The rationale for selecting these two techniques is that wavelet-based methods are often applied to characterize non-stationary signals, hence making it suitable for

EEG data (Guo et al., 2009; Chai et al., 2017). On the other hand, using power spectral density to characterize EEG data is widely used in literature, given their high expressive power at characterizing all kinds of signals (Carrier et al., 2001). PSD features are based on the Fourier Transform (FT), whereas WE features are computed using the Wavelet Transform (WT). These transforms are invertible frequency decompositions of the original signal. The FT is often more suitable for stationary signals. In contrast, the WT better captures the characteristics of non-stationary data (Sifuzzaman et al., 2009). The main difference between FT and WT is that FT is localized within the frequency domain, and WT is both within frequency and time domains. The WT includes information about the point in time the frequencies occurs, whereas the FT does not. Whether time information makes a difference for biometric purposes is subject to study.

To investigate their respective characterization power, raw signals are used as baseline, as depicted in **Figure 3A**. Raw signal features represent normalized signal power over time. Applying normalization over the input signal results in feature matrix  $\mathbf{X} \in \mathbb{R}^{n \times d} = (\hat{X} - \mu_{\hat{X}})/\sigma_{\hat{X}}$  with  $d = w$ .

#### 3.2.1. Power Spectral Density (PSD) vs. Wavelet Energy (WE)

The PSD represents signal power over frequencies. Spectral density estimation is performed over the input data  $\hat{X}$  in order to transform the time domain signal into the frequency domain.



This domain transformation is performed using Welch's method and aims at extracting the four desired frequency bands ([4–8, 8–16, 16–32, 32–64 Hz]). Using PSD, it is possible to extract any desired frequency band. This specific band choice has the objective of matching the ones obtained by performing the wavelet decomposition for obtaining the WE feature. The final feature for every frequency band is obtained by averaging the PSD from each frequency band. Computing this feature for each channel and each frequency band results in feature matrix  $\mathbf{X} \in \mathbb{R}^{n \times d}$  with  $d = 4$ . Each row in the feature matrix represents a channel, and each column represents a frequency band as depicted in **Figure 3B**.

Wavelet energy feature is computed by splitting a pre-processed window  $\hat{\mathbf{X}}$  into different frequency ranges using a series of DWTs (see **Figure 3C**). The wavelet transform requires the selection of the mother wavelet and the decomposition

level ( $N$ ). The mother wavelet acts as a filter that is applied recursively to the input signal. In order to extract highly representative features, the mother wavelet will ideally have similar characteristics as the input signal. The “db4” mother wavelet has proven to be the most effective for EEG data (Subasi, 2007) and hence was used for all experiments. Given pre-processed signals are filtered from 4 to 45 Hz, setting the maximum decomposition level equal to 5 avoids adding noise to the features since choosing a higher decomposition level would include frequencies lower than 4 Hz. **Table 1** displays the frequency range obtained from each decomposition level. Each decomposition level consists of two filters and two down samples. These filters produce an approximation coefficient containing low-frequency information and a detail coefficient containing high-frequency information. By performing the wavelet energy computations on the coefficients ranging from D2 to D5 and

**TABLE 1** | Wavelet signal frequencies for different decomposition levels.

Decomposed signal	Frequencies (Hz)	Decomposition level
D1 (noise)	64–128	1
D2	32–64	2
D3	16–32	3
D4	8–16	4
D5	4–8	5

disregarding the D1 coefficient, we expect to reduce the total noise added onto the features, as pointed by Mohammadi et al. (2017). These coefficients for each EEG window by applying the DWT recursively over the signal's approximation coefficients at different decomposition levels. The shape of the coefficients will vary for each level due to the down sample applied at each step. With all the coefficients computed, it is possible to extract the wavelet energy feature using (1)

$$e_j = \sum_{i=1}^{k_j} (\mathbf{d}_j(i))^2 \quad (1)$$

where  $e_j$  is the wavelet energy feature for a certain channel at decomposition level  $j$ .  $k_j$  denotes the number of wavelet coefficients for decomposition level  $j$  and  $\mathbf{d}_j(i)$  is the value of the detail coefficient at point  $i$  for decomposition level  $j$ . Calculating the wavelet energy feature for all channels results in feature matrix  $\mathbf{X} \in \mathbb{R}^{n \times d}$  with  $d = 4$  where, similarly to the power spectral density features, each row represents a channel obtained from an electrode, and each column represents a particular frequency band.

### 3.3. Learning Approaches

Experiment settings include four different learning approaches, all of them sharing an identical goal: mapping an input sample  $\mathbf{X}$  to a vector of predicted probabilities  $\mathbf{p}$  indicating the probability of the input sample belonging to each unique individual. There is a linear model and three non-linear models inside the four different approaches. The *Logistic Regressor* (LR) is linear, meaning the transformation between input and output features does not include a non-linear activation function. It hence is a direct mapping between input and output. The MLP acts as a baseline for non-linear approaches. The CNN model consists of a 1-D Convolutional Layer followed by an MLP (CNN) to analyse whether merging the different frequency bands into one could reduce trainable parameters without sacrificing predictive performance. Finally, the *GraphConv* model consists of a Graph Convolutional layer followed by an MLP. This latter approach aims at modeling spatial dependencies in 3-D space among the different channels. All of the non-linear approaches have an additional hyperparameter  $h$  ranging from 64 up to 2048 that defines the number of hidden channels for different layers as summarized in **Table 2**. **Figure 2** displays a simplified visualization of every considered learning approach.

The *logistic regressor* consists of a single linear layer mapping the input sample  $\mathbf{X}$  to a vector with a length equal to the number

of participants. Input feature matrix gets flattened into a vector (with length  $n \times d$ ), input to the linear layer. Finally, a softmax activation function gets applied over the output of the linear layer in order to convert its output into a vector of probabilities  $\mathbf{p}$  with their combined values adding up to 1. The LR is a linear model meaning the output will be a linear input transformation. High testing accuracies would indicate that the classes (unique subjects in this case) are linearly separable for the input features.

The MLP has a single hidden layer with a varying number of hidden channels  $h$ . The input feature matrix is flattened and fed into the hidden layer, followed by a ReLU non-linear activation function. Dropout is applied after the hidden layer, essentially deactivating at random some of the neuron weight updates to reduce overfitting and improve the final model's generalization capabilities. A softmax activation function is used after the output layer to obtain the desired probability vector  $\mathbf{p}$ . As opposed to the LR, the MLP is non-linear, meaning it can apply learnable non-linear transformations to the input, which could be a benefit in some cases and is subject to study.

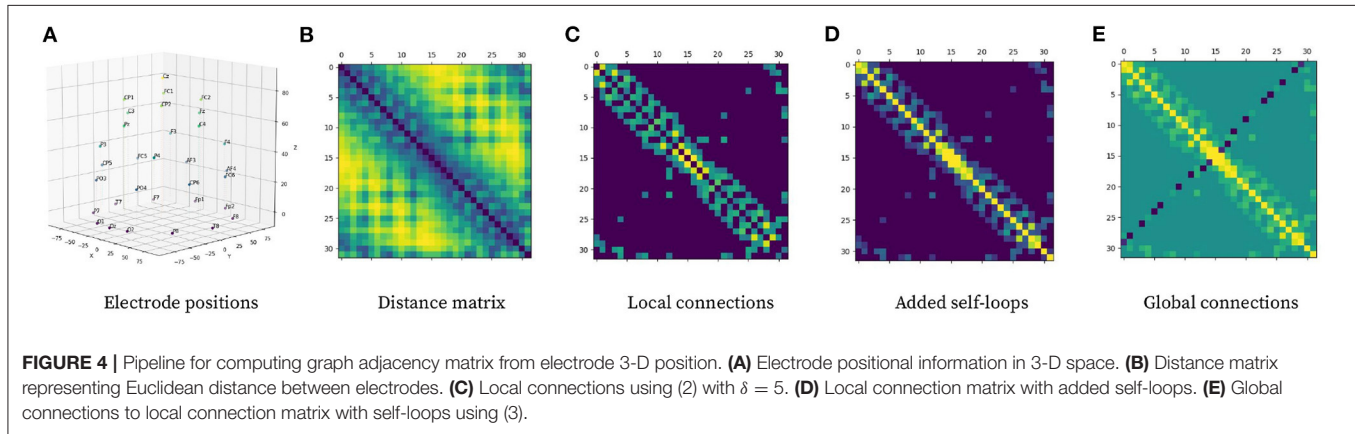
A 1-D Convolutional Neural Network merges all features into a single value by combining the different features before feeding this into the final classifier, an MLP with varying hidden channels  $h$ . The convolutional layer reduces the trainable parameter count by minimizing the shape of the input features from  $d$  to 1. The objective of the CNN is to map the input feature matrix  $\mathbf{X} \in \mathbb{R}^{n \times d}$  into a smaller, more informative feature vector  $\mathbf{x} \in \mathbb{R}^n$ . A ReLU function gets applied to the CNN's output and after the first layer of the MLP classifier. Similar to the MLP, dropout gets applied after the hidden layer of the MLP. A Softmax function acts as a final activation after the output layer in order to obtain the probability vector  $\mathbf{p}$ .

Graph Neural Network refers to a learning approach that directly operates on graphs and can perform convolution-like operations efficiently over the input data by employing a message-passing algorithm. It is possible to model complex relationships between individual elements by modeling complex structures as graphs. A *graph* is defined as  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$  where  $\mathcal{V}$  represents the set of nodes that form the graph and  $\mathcal{E}$  refers to the set of edges connecting said nodes.  $\{\mathcal{V}\}$  can be represented as a feature matrix  $\mathbf{X} \in \mathbb{R}^{n \times d}$  with  $n$  being the total number of nodes and  $d$  being the node feature dimensionality. An undirected edge set  $\{\mathcal{E}\}$  can be viewed as a symmetric weighted adjacency matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ .  $\mathbf{A}_{ij}$  represents the importance of the relationship between node  $i$  and node  $j$ . It is possible to model a sample  $\mathbf{X}$  into a graph by representing each electrode as a node, where each initial node feature  $x_i^{t=0} \in \mathbb{R}^d$  is directly obtained from each sample feature matrix  $\mathbf{X}$ . The adjacency matrix  $\mathbf{A}$  is constructed based on electrode 3-D positional information (see **Figure 4**), and it remains constant for all samples given they follow the same 10-20 electrode placement scheme.

The first step to building the adjacency matrix is to compute local connections among electrodes. The purpose of local connectivity is to allow nodes to spread information along with their neighborhood, which should theoretically help model spatial relationships inside the graph. This process consists of constructing a distance matrix based on electrode positions and then applying the local connection formula (2); with  $\delta = 5$

**TABLE 2** | Configuration for each experimental phase.

Setting name	Values main exp.	Values raw exp.	Description
Feature extraction method	raw, psd, wav	raw	Feature to extract from each raw sample to build the dataset
EEG window size ( $w$ )	0.25, 0.5, 1, 1.5, 2	0.5	Number of seconds per sub-sampled EEG window
Number of train samples ( $tr_s$ )	1, 2, 4, 8	16, 32, 64, 128, 256, 512	Number of train samples per participant
Model	MLP, CNN, GraphConv, LR	MLP, CNN, GraphConv, LR	Model architecture
Hidden channels	64, 128, 256, 512, 1,024, 2,048	512, 1,024	Number of hidden channels within non-linear models
Dropout rate	0.25	0.5	Probability for neurons before a Dropout layer not updating their weights after a backward pass



denoting a calibration constant used to keep around 20% of the possible connections to maximize the efficiency of the network topology according to Zhong et al. (2020).  $d_{ij}$  represents the Euclidean distance between electrode  $i$  and electrode  $j$ .

$$A_{ij} = \begin{cases} \min(1, \frac{\delta}{d_{ij}}) & \text{if } \min(1, \frac{\delta}{d_{ij}}) > 0.1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The last step is to add global connections, which aim at modeling EEG asymmetry, which are connections between distant channels in opposing parts of the brain.  $\{GC\}$  is the set of nodes that require a global connection between them ( $\{(FP1, FP2), (AF3, AF4), (F5, F6), (FC5, FC6), (C5, C6), (CP5, CP6), (P5, P6), (PO5, PO6), (O1, O2)\}$ ) given their superior performance at modeling EEG asymmetry as stated by Zhong et al. (2020).

$$A_{ij} = \begin{cases} A_{ij} - 1 & \text{if } (i, j) \in \{GC\} \\ A_{ij} & \text{otherwise} \end{cases} \quad (3)$$

The graph layer used in this research work is similar to that presented in Morris et al. (2019), where they proposed a higher-order graph convolutional network. This architecture proposes a node update function in which nodes are updated based on their features and the features from neighboring nodes. The graph adjacency matrix defines node neighborhoods by specifying relevant connections among the different nodes that compose the graph. Node features are updated at every layer using (4),

where  $x_i^t$  denotes node  $i$ 's feature vector at time  $t$ ,  $\Theta_1$  and  $\Theta_2$  are weight matrices that are updated using loss backpropagation and  $\mathcal{N}(i)$  indicates node  $i$ 's set of neighboring nodes. The number of updates performed on every node is equal to the total number of graph convolutional layers that is included onto the architecture, either with shared or un-shared weights. Having  $k$  convolutional layer implies considering  $k$ -hop neighborhoods around each node. This experiment only considers a single graph convolutional layer due to increased performance in preliminary experiments.

$$x_i^{t+1} = \Theta_1 x_i^t + \Theta_2 \sum_{j \in \mathcal{N}(i)} A_{ij} \cdot x_j \quad (4)$$

The *GraphConv* architecture's tail is the same as the MLP but uses the features obtained from its graph convolutional layer as input features instead of inputting the window directly. One message-passing step gets applied to the input feature matrix  $\mathbf{X}$  and this resulting feature matrix is then input to the MLP, which makes the final prediction for subject ID.

### 3.4. Training, Validation, and Testing

The training process consists of two phases: the feature extraction phase and the no-feature extraction or the raw phase. The objective of the feature extraction phase is to find out what is the minimum amount of data required to fit a Person Identification model. The raw phase is similar but aims to answer how much data is needed to train a PI model if a learned feature extraction method replaces the feature extraction step. **Table 2** displays



the different experimental variants. Each experiment was run 10 times with a Montecarlo sampling strategy, sub-sampling several training and validation samples from randomly selected videos at random timesteps for every participant (i.e., sample second 1 of video 23 for all participants). Combining the different experimental settings leads to 11,400 experiments for the feature extraction phase and 380 for the raw phase.

The first step is building the dataset for each desired configuration, according to the employed feature extraction method and the EEG window size. The output of such a feature extraction step is a set of feature matrices  $\mathbf{X}_s \in \mathbb{R}^{n \times d}$ . The target feature is equal to the subject ID. Each training sample also has an associated adjacency matrix (A) that Graph-based models use and non-graph models disregard. The number of train samples per participant ( $tr_s$ ) is a hyperparameter for each experiment. Setting  $tr_s = 1$  means having just one training example per participant, making this problem one-shot learning (see **Figure 2D**). The total number of samples per participant depends on each experimental setting, and it ranges from 1,200 ( $w = 2$ ) to 9,600 ( $w = 0.25$ ). The number of validation samples was kept constant at 100 for all experiments. The remaining data, unseen during training, is used for model evaluation. The number of validation samples was selected so that it is possible to keep it constant for all experiments. This was for ensuring that there are enough validation samples to avoid overfitting, and enough unseen testing samples to evaluate the models when the window size is largest ( $w = 2$ ). Every learning approach is presented with the same training and validation data to make the model comparison fairer. Additionally, the size for each window is an experimental setting. Whether the model performs better with smaller EEG windows or bigger ones is also subject to this study. We hypothesize that having more, smaller samples would lead the model to achieve higher prediction performance with the same total amount of time per subject. Including more samples introduces the model to examples of different mental states.

All approaches were trained using the Cross Entropy Loss function (CEL, Equation 5)

$$CEL = -\frac{1}{s} \sum_{i=1}^s y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i) \quad (5)$$

Given that PI is a supervised classification problem and every architecture output are the logits of the last layer passed through a SoftMax function, along with the Adam optimizer with a fixed learning rate of 0.0005 as it empirically demonstrated superior performance in preliminary experiments. The batch size was kept constant at 32 since this is the minimum amount the training dataset could have if  $tr_s = 1$ . The dropout rate did not seem to make much difference in the results. It was also kept constant at 0.25 in the main experimental phase and raised to 0.5 in the raw phase to improve the model's generalization capabilities. The values for the dropout probabilities were also decided based on the preliminary experiments. An early stopping patience limit was set equal to 30 to avoid wasting resources during training. The whole training took place using 4 Tesla P100 GPUs running Pytorch and Pytorch Geometric. All experiments run in 2 weeks, each taking around 8–10 min to complete on average.

## 4. RESULTS AND DISCUSSION

This research work's main objective is to provide insight into the minimal amount of data required for training an accurate Person Identification model. Each section provides results for an experimental phase. The first phase aims at testing the hypotheses that consider an explicit feature extraction block. This first phase also aims to test the hypothesis of having more minor windows over fewer bigger ones. The second phase tests the hypothesis with no explicit feature extraction block. Results are presented in these two first sections and interpreted in the subsequent discussion section.

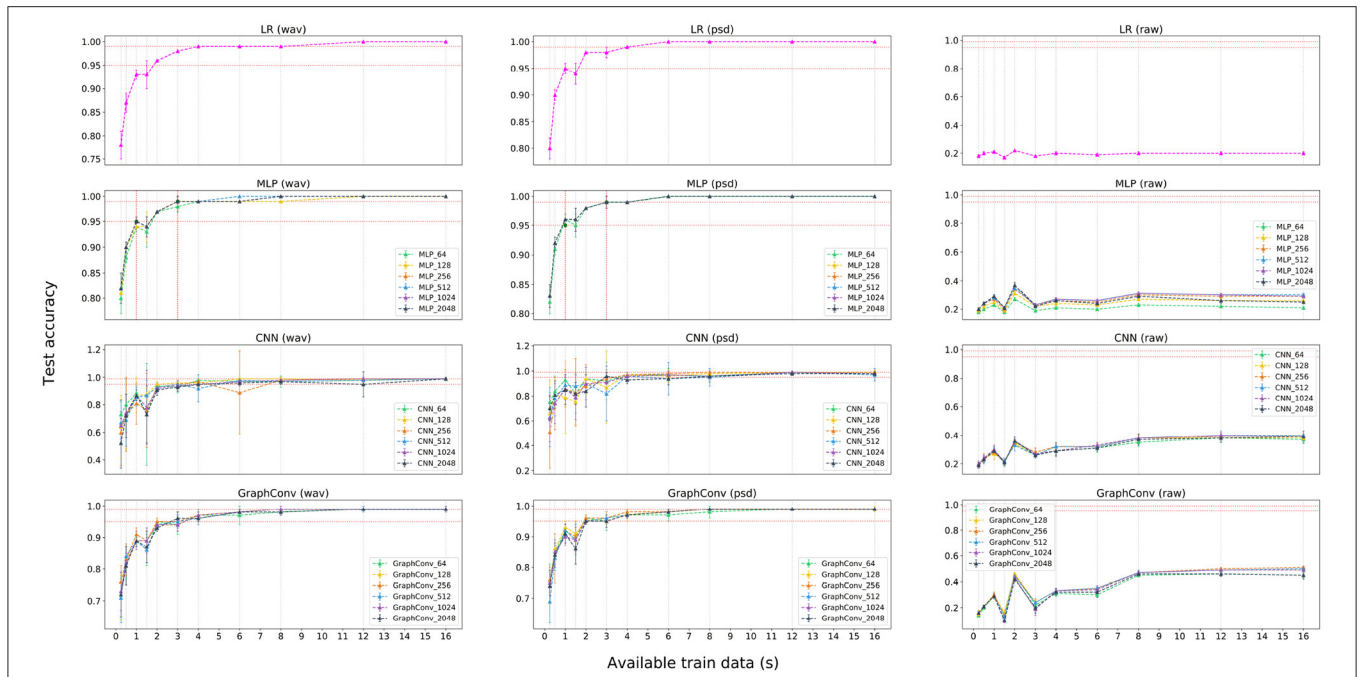
### 4.1. Feature Extraction Phase Results

The first step is to analyse whether it is better to have more minor EEG windows with more available examples or bigger windows with fewer examples. Answering this question is possible by performing a Mann–Whitney  $U$ -test over the result distributions to compare testing accuracy for different window sizes and the number of train samples. Results show that there was a significant difference in test accuracy between experiment configurations that used ( $2 \times 0.25$  s) as opposed to ( $1 \times 0.5$  s) [ $t = 2.618, p < 0.01$ ]. Similarly, experiments using ( $4 \times 0.25$  s) also performed significantly better than the ones using ( $1 \times 1.0$  s) [ $t = 3.584, p < 0.01$ ]. For the 2-s combination, smaller windows again performed better, using ( $8 \times 0.25$  s) obtained better results than ( $1 \times 2.0$  s) [ $t = 4.629, p < 0.01$ ]. Same was the case for ( $8 \times 0.5$  s) vs. ( $2 \times 2.0$ ) [ $t = 2.705, p < 0.01$ ]. **Supplementary Figure 1** displays the complete results for the independent  $U$ -Test.

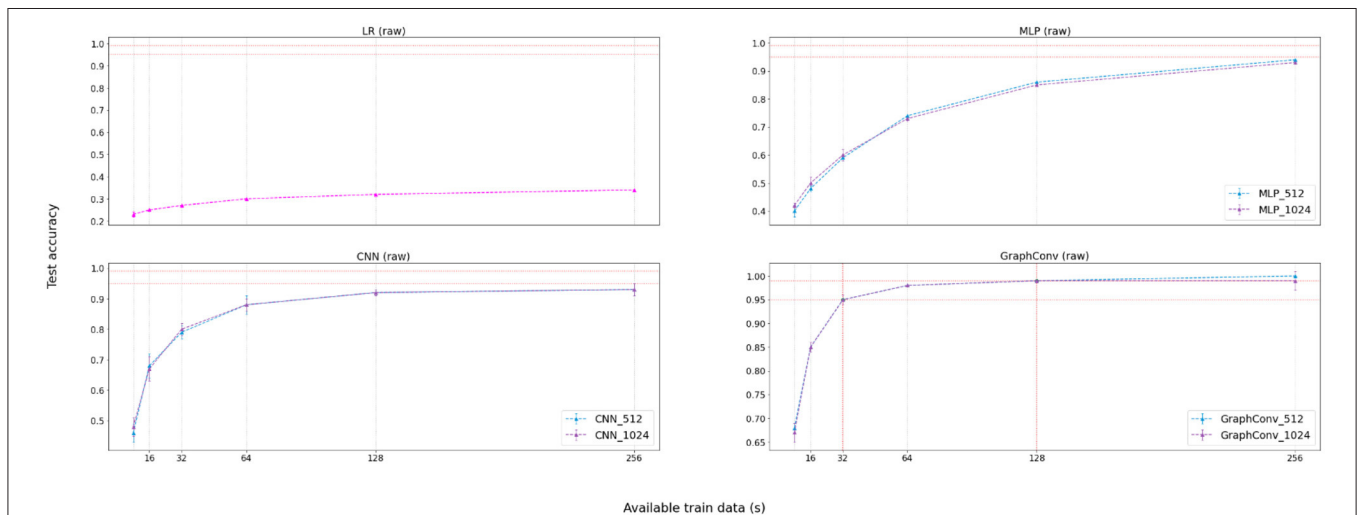
**Supplementary Tables 1–3** display the results for the main experimental phase, listing the mean and standard deviation of the test accuracy distributions for every model. These results include a row for every learning approach and a column indicating the number of seconds of data available for each architecture during training. **Figure 5** provides a visualization of these three tables for easier understanding. From these results, it is possible to observe that the LR model achieved very high performance using pre-processed features, meaning these features have a very high discriminating power that makes them linearly separable. We observe little difference in performance between the same models with different hidden channels. The MLP was able to get outstanding performance, superior to the LR but requiring more trainable parameters. The CNN achieved lower mean performance than the other models while having a significantly higher significant standard deviation. Finally, *GraphConv* had a slightly worse performance than the LR and the MLP whilst superior to the CNN. All models achieved +99% test accuracy using 8 s of data for training.

### 4.2. No Feature Extraction (Raw) Phase

**Table 2** displays the results for the raw phase. **Figure 6** provides a visualization of these results for easier understanding. Results suggest that the Graph Convolutional model handles raw data better than the other models. We attribute *GraphConv*'s success to its ability to extract compressed, representative features in its upper layers from the raw data alone. Similarly, the CNN model also creates features from raw data automatically. However,



**FIGURE 5 |** Results visualization for the feature extraction phase. Plots represent test accuracy against available training data for each participant (in seconds). Each row represents one learning approach with a legend accompanying non-linear approaches and displaying the number of hidden channels. Each column represents the employed feature extraction method, considering the raw signal as a baseline for comparison.



**FIGURE 6 |** Results visualization for the no-feature extraction (raw) phase. Graphs represent accuracy against available training data for each participant (in seconds). There exists a legend accompanying non-linear approaches and displaying the number of hidden channels.

results demonstrate that features learned by the CNN contain less expressive information than the *GraphConv*, at least when constraining the training dataset size. Results also imply that the automatic feature extraction method is better than feeding the raw data into a classifier directly, given that both LR and MLP performed poorly when the data was not processed.

In this case, the minimal amount of data required for the best performing model to achieve 95% is 32 s per participant, with 64 EEG windows with a size of 0.5 s. As for 99% test accuracy, the model requires at least 128 s with 256 windows of size 0.5 s. In this case, it was only possible to achieve accuracy higher than 95% employing the *GraphConv* learning approach, indicating

that this model's learned feature extraction step is superior to other learning approaches.

### 4.3. Discussion

Findings suggest that having more training examples per participant leads to better accuracy on test data than having fewer, bigger windows. We assume these results are because training the model with data from different trials introduces different mental states for every subject, enhancing the model's ability to recognize discriminative user features and improving predictive performance.

The minimal amount of data required to train a Person Identification model is 1 and 3 s to obtain 95% and 99% test accuracy, respectively. The model that was able to achieve 95% test accuracy and 99% with this minimal amount of data was the MLP. The results for the LR were similar to the MLP while requiring fewer parameters to be trained. Both feature extraction methods performed similarly, with PSD obtaining slightly better performance. Using raw data alone and avoiding the feature extraction step, it was possible to train a Graph convolutional model with 32 and 128 s of total data to obtain the target accuracies. These results confirm that deep learning approaches require more data than machine learning to allow robust training. We attribute the *GraphConv* model's success to its ability to learn complex feature representation automatically from raw data alone. *GraphConv* introduces spatial relationships along the different electrodes dismissed in other learning approaches. By performing a single graph convolution, the model could obtain rich features that were input to the MLP. Compared to the MLP alone, *GraphConv* shows superior performance. This superior performance indicates that the message-passing operation along the graph adds discriminative information onto the features, beneficial to the overall model predictive performance.

There exist limitations to the experiments, specifically related to the DEAP dataset. Firstly, as stated in Maiorana (2020), having single-session recordings might lead the model to learn session-specific exogenous conditions instead of personal biometric traits. However, the DEAP dataset could be considered multi-session. It consists of several independent recordings with a break at the half-mark, where electrodes are recalibrated. These facts make us believe the non-stationary nature of EEG data is preserved. Furthermore, exogenous conditions are minimized due to the experiment taking place for a long time duration. Secondly, the dataset is affective-based and sampled under naturalistic conditions. These characteristics are inherent to this dataset and should be addressed if comparing our results against any other work.

## 5. CONCLUSIONS AND FUTURE WORK

Biometric systems using EEG data have seen a rise in popularity with recent advances in ML and DL, along with a deeper understanding of feature extraction methods. These techniques have allowed researchers to obtain low error rates in a wide variety of settings including using recordings sampled from different sessions and having a larger pool of unique subjects. However, there exists little work exploring how much data

is needed to train accurate person identification models. This work focuses on answering this question. For this purpose a set of experiments was run comparing different train data sizes, learning approaches and feature extraction methods. The predictive performance of the trained models is then measured based on their accuracy for the unseen, test dataset. This work demonstrates that, in the context of affective EEG-based person identification, having just 1 s of data per participant is enough for training a model to achieve +95% test accuracy. Similarly, 3 s of data suffices to train a model with +99% test accuracy, if a suitable feature extraction method is chosen. The use of raw data is also explored comparing three simple models against a Graph Neural Network which is able to achieve +95% test accuracy using 32 s and +99% test accuracy using 128 s. EEG-based biometrics poses several advantages over traditional biometric systems, the main one it being secure, however, it also poses several disadvantages. Sampling EEG data is not as straightforward as scanning an iris or a fingerprint, several electrodes have to be attached to each subject and their signals must be recorded for a certain amount of time. Data collection is an expensive process. If real-world EEG-based biometric applications were to be implemented, data for every subject would have to be gathered and this data collection process should be as efficient as possible. Current research work focuses mainly on the method but fails to address this fact. These results pave the way for EEG biometrics used in real world applications, the objective being providing insight on how the data quantity changes the performance of the predictive models. Our findings suggest that in order to train an affective EEG-based biometric model, sampling data for a few seconds for every subject would be enough, as opposed to other methods found in literature which use several minutes from each subject to train their respective models. Each subject would not need to sit in a room with electrodes attached to its head for more than a few seconds. As a future direction, the number of electrodes attached to every subject to measure brain signals could also be studied, making the sampling process for this kind of application more straightforward and hence more efficient. Furthermore, task-independent methods considering a wider variety of tasks and session invariant features considering the non-stationary nature of EEG signals on multi-session datasets could also be subject to exploration.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.eecs.qmul.ac.uk/mmv/datasets/deap/>.

## AUTHOR CONTRIBUTIONS

The main research work was developed by CG-T. LL and BB provided great insight and feedback throughout the project as well as revising the different article drafts that were developed, providing suggestions for potential improvements. All authors contributed to the article and approved the submitted version.

## FUNDING

This publication has emanated from research supported in part by a grant from SFI Centre for Research Training in Machine Learning at Technological University Dublin under Grant number 18/CRT/6183. For the purpose of OpenAccess, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. We thank Giuliano Anselmi from

IBM for granting us access to computing resources and helping us configure the IBM powerstations our models were trained on.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fninf.2022.844667/full#supplementary-material>

## REFERENCES

- Abhang, P. A., Gawali, B. W., and Mehrotra, S. C. (2016). *Introduction to EEG-and Speech-Based Emotion Recognition*. 1st ed. Academic Press. doi: 10.1016/B978-0-12-804490-2.00007-5
- Acharya, U. R., Fujita, H., Sudarshan, V. K., Bhat, S., and Koh, J. E. (2015). Application of entropies for automated diagnosis of epilepsy using EEG signals: a review. *Knowledge Based Syst.* 88, 85–96. doi: 10.1016/j.knsys.2015.08.004
- Amzica, F., and Steriade, M. (1998). Electrophysiological correlates of sleep delta waves. *Electroencephalogr. Clin. Neurophysiol.* 107, 69–83. doi: 10.1016/S0013-4694(98)00051-0
- Anokhin, A., Steinlein, O., Fischer, C., Mao, Y., Vogt, P., Schalt, E., et al. (1992). A genetic study of the human low-voltage electroencephalogram. *Hum. Genet.* 90, 99–112. doi: 10.1007/BF00210751
- Brigham, K., and Kumar, B. V. (2010). “Subject identification from electroencephalogram (EEG) signals during imagined speech,” in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS) (IEEE)*, 1–8. doi: 10.1109/BTAS.2010.5634515
- Campisi, P., and La Rocca, D. (2014). Brain waves for automatic biometric-based user recognition. *IEEE Trans. Inform. Forensics Secur.* 9, 782–800. doi: 10.1109/TIFS.2014.2308640
- Carrier, J., Land, S., Buysse, D. J., Kupfer, D. J., and Monk, T. H. (2001). The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20–60 years old). *Psychophysiology* 38, 232–242. doi: 10.1111/1469-8986.3820232
- Carrión-Ojeda, D., Mejía-Vallejo, H., Fonseca-Delgado, R., Gómez-Gil, P., and Ramírez-Cortés, M. (2019). “A method for studying how much time of EEG recording is needed to have a good user identification,” in *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)* (Guayaquil: IEEE), 1–6. doi: 10.1109/LA-CCI47412.2019.9037054
- Cassani, R., Estarellas, M., San-Martin, R., Fraga, F. J., and Falk, T. H. (2018). Systematic review on resting-state EEG for Alzheimer’s disease diagnosis and progression assessment. *Dis. Mark.* 2018, 5174815. doi: 10.1155/2018/5174815
- Chai, X., Wang, Q., Zhao, Y., Li, Y., Liu, D., Liu, X., et al. (2017). A fast, efficient domain adaptation technique for cross-domain electroencephalography (EEG)-based emotion recognition. *Sensors* 17, 1014. doi: 10.3390/s17051014
- Chen, J., Mao, Z., Yao, W., and Huang, Y. (2020). EEG-based biometric identification with convolutional neural network. *Multimedia Tools Appl.* 79, 10655–10675. doi: 10.1007/s11042-019-7258-4
- Cui, Z., Henrickson, K., Ke, R., and Wang, Y. (2019). Traffic graph convolutional recurrent neural network: a deep learning framework for network-scale traffic learning and forecasting. *IEEE Trans. Intell. Transport. Syst.* 21, 4883–4894. doi: 10.1109/TITS.2019.2950416
- Das, B. B., Kumar, P., Kar, D., Ram, S. K., Babu, K. S., and Mohapatra, R. K. (2019). A spatio-temporal model for EEG-based person identification. *Multimedia Tools Appl.* 78, 28157–28177. doi: 10.1007/s11042-019-07905-6
- Das, R., Maiorana, E., and Campisi, P. (2017). “Visually evoked potential for EEG biometrics using convolutional neural network,” in *2017 25th European Signal Processing Conference (EUSIPCO)* (Rome: IEEE), 951–955. doi: 10.23919/EUSIPCO.2017.8081348
- Das, R., Maiorana, E., and Campisi, P. (2018). “Motor imagery for EEG biometrics using convolutional neural network,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Rome: IEEE), 2062–2066. doi: 10.1109/ICASSP.2018.8461909
- DelPozo-Banos, M., Travieso, C. M., Weidemann, C. T., and Alonso, J. B. (2015). EEG biometric identification: a thorough exploration of the time-frequency domain. *J. Neural Eng.* 12, 056019. doi: 10.1088/1741-2560/12/5/056019
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). “Generative adversarial nets,” in *Advances in Neural Information Processing Systems 27* (Curran Associates, Inc.). Available online at: <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b1226f1f8f06494c97b1afccf3-Paper.pdf>
- Guo, L., Rivero, D., Seoane, J. A., and Pazos, A. (2009). “Classification of EEG signals using relative wavelet energy and artificial neural networks,” in *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation* (New York, NY: Association for Computing Machinery), 177–184. doi: 10.1145/1543834.1543860
- Jayarathne, I., Cohen, M., and Amarakeerthi, S. (2017). “Survey of EEG-based biometric authentication,” in *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)* (Fukushima: IEEE), 324–329. doi: 10.1109/ICAWS.2017.8256471
- Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 5, 327–339. doi: 10.1109/TAFFC.2014.2339834
- Koelstra, S., Mühl, C., and Patras, I. (2009). “EEG analysis for implicit tagging of video data,” in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* (Queen Mary: IEEE), 1–6. doi: 10.1109/ACII.2009.5349482
- Koelstra, S., Mühl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/T-AFFC.2011.15
- Kong, X., Kong, W., Fan, Q., Zhao, Q., and Cichocki, A. (2018). “Task-independent EEG identification via low-rank matrix decomposition,” in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Tokyo: IEEE), 412–419. doi: 10.1109/BIBM.2018.8621531
- Li, Y., Yingle, F., Gu, L., and Qinye, T. (2009). “Sleep stage classification based on EEG Hilbert-Huang transform,” in *2009 4th IEEE Conference on Industrial Electronics and Applications* (Hangzhou: IEEE), 3676–3681.
- Maiorana, E. (2020). Deep learning for EEG-based biometric recognition. *Neurocomputing* 410, 374–386. doi: 10.1016/j.neucom.2020.06.009
- Marcel, S., and Millán, J. R. (2007). Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 743–752. doi: 10.1109/TPAMI.2007.1012
- Mohammadi, G., Shoushtari, P., Molaei Ardekani, B., and Shamsollahi, M. B. (2006). “Person identification by using AR model for EEG signals,” in *Proceeding of World Academy of Science, Engineering and Technology* (Prague), 281–285.
- Mohammadi, Z., Frounchi, J., and Amiri, M. (2017). Wavelet-based emotion recognition system using EEG signal. *Neural Comput. Appl.* 28, 1985–1990. doi: 10.1007/s00521-015-2149-8
- Morris, C., Ritzert, M., Fey, M., Hamilton, W. L., Lenssen, J. E., Rattan, G., et al. (2019). Weisfeiler and Leman go neural: higher-order graph neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 4602–4609. doi: 10.1609/aaai.v33i01.33014602
- Murugappan, M., Nagarajan, R., and Yaacob, S. (2009). “Comparison of different wavelet features from EEG signals for classifying human emotions,” in *2009*

- IEEE Symposium on Industrial Electronics & Applications* (Perlis; Jajawi;Kangar: IEEE), 836–841. doi: 10.1109/ISIEA.2009.5356339
- Özdenizci, O., Wang, Y., Koike-Akino, T., and Erdoğan, D. (2019). Adversarial deep learning in EEG biometrics. *IEEE Signal Process. Lett.* 26, 710–714. doi: 10.1109/LSP.2019.2906826
- Pope, P. E., Kolouri, S., Rostami, M., Martin, C. E., and Hoffmann, H. (2019). “Explainability methods for graph convolutional neural networks,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Los Alamitos, CA: IEEE Computer Society), 10764–10773. doi: 10.1109/CVPR.2019.011103
- Poulos, M., Rangoussi, M., and Alexandris, N. (1999a). “Neura network based person identification using EEG features,” in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings* (Paphos: IEEE), 1117–1120. doi: 10.1109/ICASSP.1999.759940
- Poulos, M., Rangoussi, M., Chrissikopoulos, V., and Evangelou, A. (1999b). “Parametric person identification from the EEG using computational geometry,” in *ICECS’99. Proceedings of ICECS’99. 6th IEEE International Conference on Electronics, Circuits and Systems* (Paphos: IEEE), 1005–1008.
- Poulos, M., Rangoussi, M., Chrissikopoulos, V., and Evangelou, A. (1999c). “Person identification based on parametric processing of the EEG,” in *ICECS’99. Proceedings of ICECS’99. 6th IEEE International Conference on Electronics, Circuits and Systems*. (Paphos: IEEE), 283–286.
- Revett, K. (2012). Cognitive biometrics: a novel approach to person authentication. *Int. J. Cogn. Biometr.* 1, 1–9. doi: 10.1504/IJCB.2012.046516
- Riera, A., Soria-Frisch, A., Caparrini, M., Grau, C., and Ruffini, G. (2007). Unobtrusive biometric system based on electroencephalogram analysis. *EURASIP J. Adv. Signal Process.* 2007, 1–8. doi: 10.1155/2007/143728
- Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2008). The graph neural network model. *IEEE Trans. Neural Netw.* 20, 61–80. doi: 10.1109/TNN.2008.2005605
- Seha, S. N. A., and Hatzinakos, D. (2021). “Longitudinal assessment of EEG biometrics under auditory stimulation: a deep learning approach,” in *2021 29th European Signal Processing Conference (EUSIPCO)* (Toronto, ON: IEEE), 1386–1390. doi: 10.23919/EUSIPCO54536.2021.9616098
- Shedeed, H. A. (2011). “A new method for person identification in a biometric security system based on brain EEG signal processing,” in *2011 World Congress on Information and Communication Technologies* (Cairo: IEEE), 1205–1210. doi: 10.1109/WICT.2011.6141420
- Shi, W., and Rajkumar, R. (2020). “Point-GNN: graph neural network for 3D object detection in a point cloud,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Pittsburgh, PA), 1711–1719. doi: 10.1109/CVPR42600.2020.00178
- Sifuzzaman, M., Islam, M., and Ali, M. (2009). Application of wavelet transform and its advantages compared to Fourier transform. *J. Phys. Sci.* 13.
- Subasi, A. (2005). Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Expert Syst. Appl.* 28, 701–711. doi: 10.1016/j.eswa.2004.12.027
- Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Syst. Appl.* 32, 1084–1093. doi: 10.1016/j.eswa.2006.02.005
- Thomas, K. P., and Vinod, A. P. (2018). Eeg-based biometric authentication using gamma band power during rest state. *Circ. Syst. Signal Process.* 37, 277–289. doi: 10.1007/s00034-017-0551-4
- Ullah, I., Hussain, M., Aboalsamb, H., et al. (2018). An automated system for epilepsy detection using EEG brain signals based on deep learning approach. *Expert Syst. Appl.* 107, 61–71. doi: 10.1016/j.eswa.2018.04.021
- Vaid, S., Singh, P., and Kaur, C. (2015). “EEG signal analysis for BCI interface: a review,” in *2015 Fifth International Conference on Advanced Computing & Communication Technologies* (Chandigarh: IEEE), 143–147. doi: 10.1109/ACCT.2015.72
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). “Attention is all you need,” in *Advances in Neural Information Processing Systems* (Long Beach, CA: Curran Associates, Inc.), 5998–6008. Available online at: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- Vilone, G., and Longo, L. (2020). Explainable artificial intelligence: a systematic review. *arXiv [Preprint]*. arXiv:2006.00093. doi: 10.48550/arXiv.2006.00093
- Vilone, G., and Longo, L. (2021a). Classification of explainable artificial intelligence methods through their output formats. *Mach. Learn. Knowledge Extract.* 3, 615–661. doi: 10.3390/make3030032
- Vilone, G., and Longo, L. (2021b). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Inform. Fusion.* 76, 89–106. doi: 10.1016/j.inffus.2021.05.009
- Vogel, F. (1970). The genetic basis of the normal human electroencephalogram (EEG). *Humangenetik* 10, 91–114. doi: 10.1007/BF00295509
- Wang, M., Hu, J., and Abbass, H. A. (2020). Brainprint: EEG biometric identification based on analyzing brain connectivity graphs. *Pattern Recogn.* 105, 107381. doi: 10.1016/j.patcog.2020.107381
- Wilaiprasitporn, T., Ditthaporn, A., Matchaparn, K., Tongbuasirilai, T., Banluesombatkul, N., and Chuangsuwanich, E. (2019). Affective EEG-based person identification using the deep learning approach. *IEEE Trans. Cogn. Dev. Syst.* 12, 486–496. doi: 10.1109/TCDS.2019.2924648
- Yao, L., Mao, C., and Luo, Y. (2019). “Graph convolutional networks for text classification,” in *Proceedings of the AAAI Conference on Artificial Intelligence* (Chicago, IL), 7370–7377. doi: 10.1609/aaai.v33i01.33017370
- Zhong, P., Wang, D., and Miao, C. (2020). EEG-based emotion recognition using regularized graph neural networks. *IEEE Trans. Affect. Comput.* 1–1. doi: 10.1109/TAFFC.2020.2994159
- Zivot, E., and Wang, J. (2006). “Vector autoregressive models for multivariate time series,” in *Modeling Financial Time Series with S-Plus* (New York, NY: Springer New York), 385–429. doi: 10.1007/978-0-387-32348-0\_11

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gómez-Tapia, Bozic and Longo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.