Evaluation of Features and Channels of Electroencephalographic Signals for Biometric Systems

Dustin Carrión-Ojeda[a,*], Paola Martínez-Arias[a], Rigoberto Fonseca-Delgado[b], Israel Pineda[a], Héctor Mejía-Vallejo[a]

[a]*School of Mathematical and Computational Sciences, Yachay Tech University, Urcuquí, Ecuador*
[b]*Electrical Engineering Department, Metropolitan Autonomous University Iztapalapa, Mexico City, Mexico*

[*]Corresponding author.
Email addresses: dustin.carrion@yachaytech.edu.ec (Dustin Carrión-Ojeda), paola.martinez@yachaytech.edu.ec (Paola Martínez-Arias), rfonseca@izt.uam.mx (Rigoberto Fonseca-Delgado), ipineda@yachaytech.edu.ec (Israel Pineda), hector.mejia@yachaytech.edu.ec (Héctor Mejía-Vallejo)

Preprint submitted to Expert Systems with Applications *November 26, 2021*

Abstract

Biometric systems are essential tools in current day society where most of our personal information lives in digital form. Although there is a significant variety of systems, electroencephalogram (EEG) signals are a popular technique to tackle biometric problems during recent years. Nevertheless, EEG processing needs to address some challenges to become a viable technique to build production-ready biometric systems. These challenges include the adequate selection of features and channels that maximize the quality of the results and optimize resources. This work explores the most relevant features and channels for the correct operation of a biometric system. This study worked with two datasets and analyses 19 features belonging to three groups, wavelet-based, spectral, and complexity. Moreover, five classifiers were trained: Multilayer Perceptron, AdaBoost, Random Forest, Support Vector Machine, and K-Nearest Neighbors. The experimental results of this work found that the best feature for developing a biometric system is the standard deviation extracted from the coefficients of a three-level discrete wavelet transform using a multilayer perceptron. Additionally, the analysis of the two datasets showed that the proposed method for channel selection is capable of reducing the necessary number of channels and maintaining the performance regarding using all available channels. This reduction was up to 21 channels and indicated that the best channels to develop biometric systems seem to be those located on the central area of the scalp.

Keywords: EEG Biometric, Features and Channels Evaluation, Discrete Wavelet Transform, Spectral Features, Complexity Features, DWT Features

1. Introduction

We currently live in a digital era where systems require reliable security methods that only grant access to authorized individuals to protect data and critical information privacy. Due to Covid-19, we have witnessed how many organizations, including companies, schools, universities, banks, and others, have migrated their services to the internet. Therefore, personal information protection and data security have become essential. Today, most common authentication systems rely on identification numbers (PIN), image-based techniques such as fingerprints, facial patterns, iris patterns, and hand geometry (Kang et al., 2016). However, these techniques could be easily tricked or simulated, or in the case of PINs, they are also prone to be forgotten or stolen.

Electroencephalographic signals have been proposed as an alternative technique for biometric systems since they are universal, which means that every alive human being produces EEG signals. Moreover, these signals can show information about cognitive processes, and subject-specific genetics (Kang et al., 2016). The disadvantages of this kind of biometrics are related to the noise associated with brain signals which makes them more complex and unstable than traditional biometric traits. Nonetheless, some features extracted from the EEGs can help process and classify them to differentiate among subjects.

Another concern when working with electroencephalograms for biometric systems development is how to record and use them with a functional, portable, and low-cost

implementation. For this reason, biometric systems need a selection of channels to extract the most essential and helpful information for the classification between subjects. It is beneficial to avoid a full set of 14 or 32 electrodes, which are the typical number of channels found in commercial EEG headsets. The channel selection process could result in constructing biometric systems based on EEG signals with fewer channels to decrease the computational and monetary cost while keeping high performance and portability.

The present study aims to find a new approach to deal with EEGs in biometric systems that consists of two main parts. The first part focuses on identifying the best feature or set of features to extract from the EEG signals that effectively distinguish one person from another. This evaluation compares the subject classification results using five machine learning algorithms for 19 different EEG extracted features. The second part of this study concentrates on selecting the most appropriate number of EEG channels to identify between subjects and analyze their impact on the performance of the system.

The main contribution of this work is a methodology to select the best feature or features and the optimal number of EEG channels. In this work, the best feature was the standard deviation extracted from the coefficients of a three-level discrete wavelet transform, and the optimal number of channels was different for each dataset, reaching a maximum reduction of 21 channels. The article is organized as follows. Section 2 provides a summary of the related work on the feature and channel selection. Section 3 describes the datasets used in this work and the proposed methodology. Section 4 presents the experimental results and their corresponding analysis. Finally, Section 5 contains the conclusions and future work.

2. Related Works

The ability of the features to capture the most distinctive characteristics of the subjects is critical to achieving high performance in EEG-based applications (Gui et al., 2019). Due to the wide range of applications based on EEGs, researchers have focused on finding a suitable method for feature selection during recent years. The most common methods determine the best feature and include a channel selection approach (Alotaiby et al., 2015). However, most feature or channel selection methods target specific activity classification where the goal is to differentiate between activities rather than people. For this reason, the results of those methods cannot be directly applied in the development of biometrics since the neurological stimulus used for recording the signals does not matter from a security point of view but rather the precision to distinguish one person from another.

Schröder et al. (2003) proposed an automatic feature selection combining a genetic algorithm with a Support Vector Machine (SVM). They considered each channel as a feature, so their method aimed to find the best channels for developing Brain-Computer Interfaces (BCIs). The authors concluded that the optimal set of features for developing BCI is strongly dependent on the subjects and the experimental paradigm. In the development of biometric systems, a requirement is to establish a method independent of the involved people.

Even though the channels can be used as features, most EEG-based applications compute some features capable of representing the channels' most relevant information while reducing the dataset dimension. For this reason, Liu et al. (2014) first evaluated a set

of features including the autoregression (AR) model, the power spectrum of the time-domain (TPS), the power spectrum of the frequency-domain (FPS), and phase-locking value (PLV). These features were extracted from all the available channels and then evaluated with SVM. After the feature evaluation, they employed a support vector machine recursive feature elimination (SVM-RFE) to select the most discriminative channels, but only for the AR model because it showed the best performance in the previous step. Despite the selection of channels, the authors proposed using 32 EEG channels. Considering that using 32 channels still makes it hard to develop portable biometric systems, more features need to be evaluated to further reduce this number.

Another approach for selecting the best features and channels is finding a measurement representing intra-subject and inter-subject variability. In this kind of measurement, the best features and channels have minimum intra-subject variability and maximum inter-subject variability. Following this approach, Kang, Lee, and Kim (Kang et al., 2016; Lee et al., 2016) provided two analyses of features and channels for EEG-biometrics. In the first analysis, they evaluated Power Spectral Density (PSD) features and Lyapunov exponents with a criteria index (CI) that consisted of three types of variances. Using this index, they found that the maximum Lyapunov exponent had the maximum CI value among all types of EEG features. Furthermore, two (T4, T6) out of sixteen channels had the highest CI values overall brain areas (Kang et al., 2016).

A second study of the authors (Lee et al., 2016) evaluates the performance of seven features (alpha/theta, alpha/beta, theta/beta power ratio, sample entropy, permutation entropy, entropy, and median values of distribution) extracted for each of the sixteen EEG channels. As a result, 112 features composed their initial set of features. Then they used mutual information to select the best ones. After their analysis, the set of features was reduced to nine, and the user identification results were higher than using PCA reduction (Lee et al., 2016). Regardless of the proposed measurements' efficiency, both works' main limitations were the dataset's size because it only contained the information of seven subjects with recordings of sixteen channels.

A different approach for features and channels selection for a subject identification is presented by Kang et al. (2018). The authors first performed a channel selection guided by the assumption that electrooculographic (EOG) interference in the resting-state with eyes open and the bursting alpha activity in the resting-state with eyes closed could lead to low authentication. Using the spectral power, they investigated the effect of these two factors and finally reduced the number of channels from 56 to 34. They analyzed, using the selected channels, ten single-channel features (seven spectral and three nonlinear) and ten multichannel features by conducting network analysis based on phase synchronization. The authors were able to select the best feature for the dataset in the study.

Although some previous works proposed different methods for selecting the best features and channels, to our knowledge, no work verifies its results using multiple datasets recorded with different EEG headsets. Additionally, an analysis that includes different classification algorithms to find the best feature-channel combination is still lacking. Moreover, features based on wavelet transform have been proven to be efficient for user identification (Shedeed, 2011; Bashar, 2018; Yang et al., 2019). For this reason, the inclusion of these kinds of features to the set of initial features could improve the understanding of EEG-based biometric systems.

3. Methodology

The proposed method is divided into four main steps: data acquisition, feature extraction, feature selection, and channel selection. The following subsections discuss the details of each step. Moreover, it is important to mention that this study extends the EEG-Based Applications using Python (EBAPy) framework (Carrión-Ojeda et al., 2021b). Due to the flexibility of this framework, this study added specific functionalities to it (Carrión-Ojeda et al., 2021c).

3.1. Data Acquisition

This work employed two datasets to train, test, and experiment with our methodology. This section describes each one of these.

The first dataset is the open-access dataset "DEAP" (Koelstra et al., 2012). It has recordings of 32 healthy participants (50% female), aged between 19 and 37 (mean age is 27). Each recording has 32 EEG channels, 12 peripheral channels, three unused channels, and one status channel. All subjects recorded 40 trials, and each trial has a duration of 63 seconds (3 seconds of baseline and 60 seconds of trial). In each trial, the participants were presented with a YouTube video aimed at evoking particular sentiments. After each video, the subjects scored their level of valence, arousal, dominance, and liking.

Even though the DEAP dataset aims to analyze human affective states, it can be correctly used for biometrics development. It is only needed to know which subject each EEG corresponds to, regardless of the task performed at recording time. Furthermore, using a dataset with different affective states is advantageous because each subject's EEGs can vary significantly between the trials providing a more realistic environment.

DEAP dataset has two versions, one with the raw signals and another one with already preprocessed signals. This work uses the preprocessed version of the DEAP dataset. In this version, the EEG signals were downsampled from 512 Hz to 128 Hz, passed through a band-pass frequency filter from 4-45 Hz, and through a common average reference filter to improve the signal-to-noise ratio. Additionally, the electrooculographic (EOG) artifacts were removed using a blind source separating technique.

The second dataset was the private "BIOMEX-DB"[1]. The main reason to use another dataset was to increase the reliability of this study. In contrast to DEAP, this dataset aims to develop multimodal biometric systems by providing EEG, audio, and video recordings. However, our research only analyzes the EEGs.

The information of BIOMEX-DB corresponds to 51 healthy participants (49% female), aged between 16 and 61 (mean age 29). Each participant registered 135 trials of 2.5 seconds, where each trial was the pronunciation of a number between 1 to 10. Each EEG was recorded with a frequency of 2048 Hz and 14 channels. However, the sampling frequency was downsampled to 128 Hz.

As with DEAP, the original labels of BIOMEX-DB were replaced by the subjects of each EEG. Based on the results of previous works (Carrión-Ojeda et al., 2019, 2021a), this research uses 1.75 seconds of recording taken randomly from each trial.

[1] A detailed description available at `http://inaoe.repositorioinstitucional.mx/jspui/handle/1009/1604`

3.2. Feature Extraction

This study evaluates nineteen features belonging to three categories: spectral, complexity, and wavelet-based. Each feature is detailed below.

3.2.1. Spectral Features

Dauwels et al. (2010) use the spectral features to quantify the changes in the signal given by its power. When working with spectral features, it is necessary to work with the Fourier transform of the signal.

Power Spectral Density (PSD): The PSD shows the strength of the energy variation of a signal as a function of the frequency. It is defined as the Discrete-time Fourier Transform of the covariance sequence (Stoica et al., 2005). Nevertheless, in this work, the PSD was obtained using the "Periodogram", which is a Matlab function. The periodogram is a nonparametric estimate of the PSD, and it is defined as the Fourier transform of the biased estimate of the autocorrelation sequence (Auger & Flandrin, 1995). To calculate this, we use the frequency sample of the signals, which is 128Hz.

Total Average Power (TAP): In Physics, power is defined as work per time, meaning the amount of energy transferred per unit of time. The following formula is used to calculate the total average power. Considering a signal x:

$$TAP = \frac{\sum |x|^2}{\text{length}(x)}. \tag{1}$$

Median Frequency (MFreq): The MFreq is defined as the frequency in which the total spectral power is halved (McBride et al., 2014). It is expressed by the following equation:

$$\sum_{j=1}^{MDF} P_j = \frac{1}{2} \sum_{j=1}^{M} P_j, \tag{2}$$

where P_j is the EEG power spectrum at the frequency bin j, MDF is the frequency value in which the power spectrum is divided in two parts with equal integrated power (Thongpanja et al., 2013), and M is the length of the frequency bin (Phinyomark et al., 2012).

Relative Power (RelPow): This method consists in the computation of the Spectral Power in each frequency band of the EEG, and then, it is necessary to compute the percentage of the total power that each band has. For computing the RelPow, this work used four frequency bands: δ (0-4 Hz), θ (4-8 Hz), α (8-12 Hz), and β (12-30 Hz).

3.2.2. Complexity Features

This category of features quantifies the entropy of a system. In the context of information theory, entropy is defined as the measure of the uncertainty associated with a random variable (Dauwels et al., 2010). The greater the entropy, the more complex the system is.

Tsallis Entropy (TsEn): The TsEn is the generalization of the Boltzmann-Gibbs entropy (Coronel et al., 2017) and it is determined by quantizing the amplitude of the EEG. As mentioned in (Tsallis, 1988), given a discrete set of probabilities $\{p_i\}$ with the

condition that $\sum p_i = 1$, the TsEn is defined as:

$$S_q(p_i) = \frac{k}{q-1}\left(1 - \sum_{i=1}^{W} p_i^q\right), \tag{3}$$

where $q \in \mathbb{R}$ is any real number, $W \in \mathbb{N}$ is the total number of possible configurations. and k is a conventional positive constant. For this work we used $k = 1$ and $q = 3$, as recommended in several studies for determining changes in a signal (Zhang et al., 2009).

Approximate Entropy (ApEn): It measures the logarithmic likelihood that runs of patterns that are close (within a window denominated by r) for m contiguous observations on subsequent incremental comparisons. In this work we used the values of $r = 0.01$ and $m = 1$ to determine the amplitude variation in the signal (Sarlabous et al., 2010). To compute the ApEn the following steps should be followed (Abásolo et al., 2008):

1. Form a time series data $\{u(n)\} = u(1), u(2), ..., u(N)$ with N data values equally spaced in time.
2. Fix $m \in \mathbb{Z}$ and $r \in \mathbb{R}^+$ where m is the length of the run data, and r the filtering level, or tolerance window.
3. Form a sequence of vectors $x(1), x(2), ..., x(N-m+1) \in \mathbb{R}^m$, a m-dimensional space defined by $x(i) = [u(i), u(i+1), ..., u(i+m-1)]$ where $i \in \{1, ..., N-m+1\}$.
4. Define $d[x(i), x(j)]$ which is the distance between $x(i)$ and $x(j)$, as the maximum absolute difference between their respective scalar components

$$d[x(i), x(j)] = \max_{k=1,2,...,m} [u(i+k-1) - u(j+k-1)]. \tag{4}$$

5. For a given $x(i)$ count the number of j ($j \in \{1, ..., N-m+1\}$) so that $d[x(i), x(j)] \leq r$ denoted as $N^m(i)$. Then:

$$C_r^m = \frac{N^m(i)}{N-m+1}, \tag{5}$$

   for $i \in \{1, ..., N-m+1\}$.
6. Find the natural logarithm of each $C_r^m(i)$ and average it over i

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_r^m(i). \tag{6}$$

7. Increase the dimension of m to $m+1$ and then find $C_r^{m+1}(i)$ and $\Phi^{m+1}(r)$
8. Define Approximate Entropy as:

$$ApEn = \Phi^m(r) - \Phi^{m+1}(r). \tag{7}$$

Sample Entropy (SampEn): The SampEn is a refinement of the ApEn designed to have a smaller bias since it does not include self-similar patterns (Richman et al., 2004). Assuming a time-series sequence of length $N, u(i) = \{u_1, ..., u_N\}$, $N-m+1$ vectors $x_m(i)$ with $i \in \{1, ..., N-m+1\}$ where $x_m(i) = \{u(i+k) : 0 \leq k \leq m-1\}$ are formed. Then,

the distance function $d[x_m(i), x_m(k)]$ is defined to be the maximum difference between the components of the vectors:

$$d[x_m(i), x_m(k)] = \max\{|u(i+j) - u(k+j)| : 0 \leq j \leq m\}. \tag{8}$$

Finally, the SampEn is defined as:

$$SampEn = -\log \frac{A}{B}, \tag{9}$$

where A is the number of template vector pairs having $d[x_{m+1}(i), x_{m+1}(k)] < r$, and B is the number of template vector pairs having $d[x_m(i), x_m(k)] < r$ (Richman et al., 2004). The result is always greater than zero, and a small value is an indicator of less noise and more self-similarity. For this feature, we also used the values of $r = 0.01$ and $m = 1$.

Lempel-Ziv Complexity (LZC): This method counts the number of different patterns in a signal of length n. The fewer such patterns, the better a signal may be compressed. The compression rate is a measure of the regularity of a signal (Zhang et al., 2016).

The original signal must be coarse-grained and transformed into a symbol sequence for simplifying the computation. To generate a two-state sequence, signal ($R = \{r(1), r(2), ..., r(n)\}$) the following equation is applied:

$$r(i) = \begin{cases} 0, & \text{if } x(i) < T_h \\ 1, & \text{if } x(i) \geq T_h \end{cases}$$

where n is the length of the signal $x(n)$ and T_h is the threshold, which normally is the mean value of the sequence.

Using the binary sequence R, the vector $c(n)$, which is a counter of the different patterns, for a binary symbol sequence, is calculated following the below process:

1. Let S and Q denote two strings. SQ is the concatenation of S and Q. $SQ\pi$ is the concatenated string with the last character deleted, and $v(SQ\pi)$ is the vocabulary of all the different substrings of $SQ\pi$. For example, consider $c(n) = 1$, $S = S_1$ and $Q = S_2$, so $SQ\pi = S_1$.
2. If $Q \in v(SQ\pi)$ then Q is a substring of $SQ\pi$.
   $S = S_1 S_2 S_3 \dots S_r$
   $Q = S_{r+1}$
   $SQ\pi = S_1 S_2 ... S_r$
3. $Q = S_{r+1} S_{r+2} \dots S_{r+i}$ is not a substring of $SQ\pi = S_1 S_2 S_3 ... S_{r+i-1}$, so $c(n)$ increases by one.
4. $c(n)$ is updated because a new pattern was found, then S and Q are also updated: $S = S_1 S_2 \dots S_{r+i}$ and $Q = S_{r+i+1}$

As a result, $c(n)$ is the number of different substrings contained in R, meaning that $c(n)$ represent the different patterns in a sequence. Also, $c(n)$ may vary with length, therefore, it should be normalized. The upper bound of $c(n)$ is:

$$c(n) < \frac{n}{(1-\epsilon_n)\log_\alpha(n)}, \tag{10}$$

where n is the length of the sequence, α the number of different symbols, and ϵ_n is a small quantity. Additionally, $\epsilon \to 0$ when $n \to \infty$.

$$\lim_{n\to\infty} c(n) = b(n) = \frac{n}{\log_\alpha(n)}. \tag{11}$$

Finally, $c(n)$ can be normalized as:

$$C(n) = \frac{c(n)}{b(n)}. \tag{12}$$

Higuchi Fractal Dimension (HFD): According to Giannakakis et al. (2014), the Fractal Dimension is a nonlinear measure on the time domain that is used to characterize the complexity of a time series. It is helpful to quantify the complexity and self-similarity from a signal (Gómez et al., 2009). The degree of complexity of the sequence increases as the fractal dimension increases. The following procedure was used to obtain this measure. Given a one dimensional time series $X = x(1), x(2), ..., x(N)$, form k new time series X_k^m defined by:

$$X_k^m = \left\{ x(m), x(m+k), x(m+2k), ..., x\left(m + int\left(\frac{N-m}{k}\right) \times k\right)\right\}, \tag{13}$$

where k and m are both integers, and $int(*)$ makes reference to the integer part of '$*$'. k represents the discrete time interval between points, and $m = 1, 2, ..., k$ is a representation of the initial time value (Gómez et al., 2009). In this sense, taking as an example $k_{max} = 3$ and $N = 100$:

$$X_1^3 : X(1), X(4), X(7), \ldots, X(94), X(97), X(100)$$
$$X_2^3 : X(2), X(5), \ldots, X(95), X(98)$$
$$X_3^3 : X(3), X(6), \ldots, X(96), X(99)$$

For each of the time series constructed by the previous equation, the length is computed in the following way:

$$L(m,k) = \frac{1}{k}\left[\left(\sum_{i=1}^{int(\frac{N-m}{k})} |X[m+ik] - X[m+(i-1)\times K]|\right) \times \frac{N-1}{int(\frac{N-m}{k})k}\right], \tag{14}$$

with N as the length of the original time series X, and $\frac{N-1}{int(\frac{N-m}{k})k}$ is a normalization factor.

$$L(k) = \frac{1}{k} \times \sum_{m=1}^{k} L(m,k). \tag{15}$$

Those steps are repeated k_{max} times for each $k = 1, 2, \ldots, k_{max}$ The next step in the procedure, is to plot $L(k)$ against $1/k$ on a double logarithmic scale, with $k = 1, 2, \ldots, k_{max}$. The result should be the data falling on a straight line that has a slope equal to the Fractal Dimension of the time-series X. Therefore, Higuchi Fractal Dimension (HFD) is defined as the slope of the line that fits the pairs $\{\ln(L(k)), \ln(1/k)\}$ using a least-square method to determine it. The value of k_{max} is chosen at the point in which the Fractal Dimension is considered a saturation point. A value of $k_{max} = 60$ was chosed for this study.

3.2.3. Wavelet-based Features

This study analyses the ten features proposed by Hamad et al. (2016). All features are calculated after applying a discrete wavelet transform (DWT). Previous experimental results by Carrion-Ojeda et al. (Carrión-Ojeda et al., 2019, 2021a) showed that a three-level DWT with Daubechies-4 as mother wavelet provides efficient features. For this reason, this study uses those parameters for the DWT. All the detail coefficients (D) and the last approximation coefficient (A) were analyzed to compute the features. Below are the equations for all the extracted features and in all of them $i \in \begin{cases} \{1,2,3\}, & \text{if } C = D \\ \{3\}, & \text{if } C = A \end{cases}$, and N is the length of the coefficients.

Maximum per Wavelet Coefficient (max):

$$\max(C_i) \tag{16}$$

Minimum per Wavelet Coefficient (min):

$$\min(C_i) \tag{17}$$

Mean per Wavelet Coefficient:

$$\mu_{C_i} = \frac{1}{N}\sum_{j=1}^{N} C_{i_j} \tag{18}$$

Standard Deviation per Wavelet Coefficient:

$$\sigma_{C_i} = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(C_{i_j} - \mu_{C_i})^2} \tag{19}$$

Variance per Wavelet Coefficient:

$$V_{C_i} = \sigma_{C_i}^2 \tag{20}$$

Median per Wavelet Coefficient: Median of each coefficient.

Skewness per Wavelet Coefficient:

$$Skewness_{C_i} = \frac{3 \times (\mu_{C_i} - \text{median}_{C_i})}{\sigma_{C_i}} \tag{21}$$

Energy per Wavelet Coefficient:

$$E_{C_i} = \sum_{j=1}^{N} |C_{i_j}|^2 \tag{22}$$

Relative Wavelet Energy: For computing the RWE, first the total energy is need:

$$E_T = \left(\sum_{i=1}^{3} E_{D_i}\right) + E_{A_3} \tag{23}$$

Then, the RWE is defined as follows:

$$RWE_{C_i} = \frac{E_{C_i}}{E_T} \tag{24}$$

Entropy per Wavelet Coefficient:

$$EN_{C_i} = \sum_{j=1}^{N} C_{i_j}^2 \log_2(C_{i_j}^2) \tag{25}$$

3.3. Feature selection

An individual evaluation was carried out for each of the features detailed in Section 3.2 to select the best one. The entire process followed to evaluate and select the best feature is detailed below.

3.3.1. Hyperparameter Optimization

This study assessed five classifiers: Multilayer Perceptron (MLP), AdaBoost (AB), Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). For optimizing the hyperparameters of all classifiers, a greedy search optimization was applied. This algorithm uses a hyperparameter set and analyzes all possible combinations generated using the parameters contained in that set (Matuszyk et al., 2016). The optimization was applied individually for each feature in each dataset. Additionally, ten-fold cross-validation was used to increase the reliability of the selection of hyperparameters. The set of parameters for each classifier was the same throughout the optimization process and is shown in Table 1.

On the other hand, Table 2 shows the best parameters for each feature for the DEAP and BIOMEX-DB datasets. For each dataset, 20% of the available data was extracted to perform hyperparameter optimization. These data were divided into a training and testing set to carry out the greedy search optimization. This division was performed in a balanced way to avoid possible bias when selecting the best parameters. For this reason, 75% of the trials of each subject were randomly selected for the training set, and the remaining 25% made up the test set. For avoiding favoring any classifier, all of them were optimized using the same folds. Furthermore, it is worth mentioning that the data used at this stage was not reused at any later stage.

3.3.2. Classification

After obtaining the best combination of parameters for each classifier, these were independently trained for each dataset and feature. As mentioned in Section 3.1, this study followed a multi-class classification approach where each subject corresponds to a class, resulting in 32 classification categories for DEAP and 51 classification categories for BIOMEX-DB. For training the classifiers, a closed set strategy was followed, meaning that each classification category was presented during training.

3.3.3. Performance Evaluation and Best Feature Selection

As this study followed a multi-class classification approach, three of the multi-class performance metrics proposed by Sokolova & Lapalme (2009) were computed. The three computed metrics are Macro-averaging Sensitivity (Se), Macro-averaging Specificity (Sp),

Table 1: Set of values for hyperparameter optimization.

	Hyperparameter	Set of Values
MLP	1. Net specification (neurons per layer)	(106); (106,106); (106,106,106); (84,84); (127,127)
	2. Learning rate	1e-3; 5e-3; 0.01; 0.05; 0.1
	3. Batch normalization	True; False
	4. Dropout	True; False
	5. Dropout percentage	10; 20; 30; 40; 50
	6. L2 regularization	True; False
	7. L2 regularization value	0.01; 0.05; 0.1; 0.5
	8. Epochs	10; 100; 500; 1000
AB	1. Weak Classifier	SVM; RF
	2. Number of weak classifiers	5; 10; 50; 100; 500; 800
	3. Learning rate	0.1; 0.5; 1; 5
	4. Boosting algorithm	SAMME; SAMME.R
RF	1. Number of estimators	1; 10; 50; 100; 200; 500; 750; 1000
	2. Min. number of samples required to split an internal node	2; 5; 10; 50; 100
	3. Criterion	Gini; Entropy
SVM	1. Penalty Parameter	0.5; 1; 10; 50; 100; 200; 300
	2. Kernel	linear; rbf; sigmoid
	3. Tolerance	1e-7; 1e-6; 1e-5; 1e-3; 0.1; 1
	4. Kernel coefficient	scale; auto
KNN	1. Number of neighbors	1; 5; 10; 20; 50; 100
	2. Distance metric	Euclidean; Manhattan
	3. Leaf size	5; 10; 30; 50; 100

and Average Accuracy (Acc). All these metrics are based on the confusion matrix of the classifier, and their formulas are the following:

$$Se = \left(\sum_{i=1}^{l} \frac{Tp_i}{Tp_i + Fn_i} \right) / l, \tag{26}$$

$$Sp = \left(\sum_{i=1}^{l} \frac{Tn_i}{Tn_i + Fp_i} \right) / l, \tag{27}$$

$$Acc = \left(\sum_{i=1}^{l} \frac{Tp_i + Tn_i}{Tp_i + Fn_i + Fp_i + Tn_i} \right) / l, \tag{28}$$

where l is the number of participants (classes), Tp_i are the true-positive classifications, Tn_i are the true-negative classifications, Fp_i corresponds to the false-positive classifications and Fn_i corresponds to the false-negative classifications; all of them of the i^{th} subject.

Table 2: Best hyperparameters for each classifier and feature using DEAP and BIOMEX-DB datasets. The number at the left of the hyperparameters column correspond to the name of the hyperparameter specified on Table 1, (*) means that the hyperparameter was selected for all non explicitly specified features, (-) means that no features used the hyperparameter.

Classifier	**Hyperparameters**		**DEAP**	**BIOMEX-DB**
MLP	1	(106)	*	Max, Std
		(106,106)	HFD, Mean, TAP, TsEn	-
		(106,106,106)	-	ApEn, HFD, LZC SampEn, Skewness
		(84,84)	Entropy, Min	-
		(127,127)	ApEn, Median, SampEn	*
	2	1e-3	*	*
		5e-3	Max	Max, Median, MFreq, RelPow, Var
		0.01	Entropy, TAP	-
		0.05	Energy, PSD, Var	-
	3	True	*	ApEn, LZC, Median, Std
		False	Energy, Entropy, PSD, TAP, TsEn, Var	*
	4	True	*	*
		False	ApEn, HFD, Median, Min, SampEn, Skewness	Max, PSD, Skewness, SampEn, TsEn
	5	10	Entropy, LZC, Mean, RelPow, RWE	*
		20	TAP, TsEn	Entropy
		30	Max, MFreq	Var
		40	Var	-
		50	Energy, PSD, Std	-
		NA	*	Max, PSD, Skewness, SampEn, TsEn
	6	False	*	*
	7	NA	*	*
	8	500	-	SampEn
		1000	*	*
AB	1	SVM	RWE, Skewness	-
		RF	*	*
	2	5	*	*
		10	-	HFD, Median, MFreq, TAP, Var
		100	-	Mean, TsEn
		500	SampEn	Max
		800	TsEn	-
	3	0.1	*	*
		0.5	RWE	Entropy, TAP, Var
		1	-	SampEn

Classifier	Hyperparameters		DEAP	BIOMEX-DB
		5	-	LZC, Max, Median
	4	SAMME	*	*
		SAMME.R	SampEn, TsEn	Entropy, Max, Mean, Median, SampEn, Tap, TsEn, Var
RF		100	Energy, Var	-
		200	SampEn, Std	-
		500	ApEn, Entropy, Min, PSD, Skewness	HFD
		750	*	Entropy, Mean, MFreq, RWE, SampEn TsEn, Var
		1000	Median, RelPow, TAP	*
	2	2	*	*
		5	ApEn, Mean, SampEn, Skewness, Std	LZC, Mean, Skewness
		10	Entropy	-
	3	Gini	ApEn, Energy, Median, Min, RWE, Std, TAP	*
		Entropy	*	Energy, RWE
SVM	1	1	SampEn	LZC, Skewness
		10	LZC, Median, Skewness	Median
		50	ApEn, HFD, Mean, RelPow	ApEn, Min, Std
		100	Max, Min	Max, Mean
		200	Std, TsEn	HFD, PSD
		300	*	*
	2	linear	HFD, Max, Min, SampEn, Std	RWE, Std
		rbf	*	*
		sigmoid	TsEn	-
	3	1e-5	-	PSD
		1e-3	*	*
		0.1	HFD, Median, Min, RWE, TAP, TsEn	ApEn, Entropy, HFD, Mean, RelPow, SampEn, Std, Var
		1	LZC, SampEn	MFreq
	4	scale	*	*
		auto	ApEn, RWE, TsEn	HFD, SampEn
KNN	1	1	*	*
		5	ApEn	HFD, Median, RelPow
		10	-	ApEn, Mean, MFreq SampEn
		20	-	LZC
		100	-	Skewness

Classifier	Hyperparameters		DEAP	BIOMEX-DB
	2	Euclidean	ApEn, HFD, LZC, Mean, Median, MFreq, RelPow, RWE, SampEn	-
		Manhattan	*	*
	3	5	*	*

For increasing the reliability of the experimental results, a ten-fold-cross validation was used. The same process described in Section 3.3.1 was followed to obtain balanced folds. This stage used the data that were not used during optimization. Finally, the best feature was the one that produced the highest three performance metrics.

3.4. Channel selection

Channel selection is divided into two main stages. The first is responsible for evaluating each channel individually, while the second looks for the optimal number of channels.

3.4.1. Single channel evaluation

For evaluating each channel individually, the "good" features were first selected based on the results obtained after applying the process described in Section 3.3. A feature was considered "good" if at least three classifiers achieved a performance higher than 85% in all evaluation metrics. This selection was performed independently for the two datasets.

Once the good features were identified, new data matrices were created for each subject. The dimension of the matrices was $t \times n$, where t is the number of trials of each subject and n corresponds to the concatenation of the good features using a single channel. These new matrices were created for all available channels, resulting in matrices for each channel instead of each feature. Subsequently, the new dataset was divided in the same way as before, i.e., 20% for optimization and 80% for evaluation.

This study used a MLP to perform the channel evaluation due to its potential to find patterns in complex datasets. This classifier was optimized using the same approach described in Section 3.3.1. Note that a single MLP was optimized for all channels. For optimizing a single MLP, the optimization folds of each channel were joined into a single fold. Consequently, each fold contained the same amount of information from each channel and continued to have a balanced selection in terms of subjects. In this way, we maintained the impartiality of the classifier over the channels.

Table 3 contains the results of MLP optimization for each dataset. After the optimization, the channels were evaluated individually using the optimized MLP and the correct classification rate (CCR) as a performance metric. The CCR is defined as:

$$CCR = \frac{TP + TN}{TP + TN + FP + FN}. \tag{29}$$

This evaluation was used to find the best channel to develop a biometric system using a single EEG channel. It also allowed ordering the channels from best to worst based on the CCR to perform the subsequent analysis.

Table 3: Best hyperparameters for single channel evaluation using DEAP and BIOMEX-DB datasets.

	Hyperparameter	**Best for DEAP**	**Best for BIOMEX-DB**
MLP	Net specification	(127, 127)	(127, 127)
	Learning rate	1e-3	1e-3
	Batch normalization	False	False
	Dropout	False	True
	Dropout percentage	NA	10
	L2 regularization	False	False
	L2 reg. value	NA	NA
	Epochs	500	500

3.4.2. Optimal Number of Channels Assessment

This assessment worked only with the best feature and classifier found in Section 3.3, which was Standard Deviation. The assessment consisted of analyzing the performance of the classifier with the three metrics explained in Section 3.3.3 using a different number of channels starting from a single channel until reaching the total available channels for each dataset. The variation was made progressively, increasing one channel each time.

This study proposes to use the channel ordering method described in the previous section. For this reason, to demonstrate the efficiency of the proposed method, a comparison was made against a random ordering. For the random ordering method, the results of 10 executions (10 different ways to sort channels) were averaged. Subsequently, a multivariate analysis of variance (MANOVA) was conducted to determine the optimal number of channels for each method. The MANOVA is a method for testing statistical significance in differences among multivariate sample means and is the multivariate extension of the univariate analysis of variance (ANOVA). This method offers advantages over conducting ANOVA test for each dependent variable, at the cost of added complexity, as it reduces the likelihood of type 1 errors, i.e. wrongly rejecting a true null hypothesis, and captures correlations among combinations of independent variables (Warne, 2014). In this scenario, the optimum corresponds to the number of channels from which that number is no longer an influencing factor in the performance of the classifier.

The MANOVA analyzed the results varying the number of channels from all until a single channel. Consequently, the optimal number of channels is the least number of channels that maintain a p-value greater than 0.05. To verify if there was a significant difference between the proposed ordering method and the random ordering, the lowest performance metric among the Se, Sp, and Acc was selected to perform a Wilcoxon Test between both methods. This test was applied independently for each channel evaluation, i.e., 1 channel, 2 channels, $\ldots$, c channels, where c is the total amount of channels.

4. Experimental Results

The results and analysis are divided into feature selection, single-channel evaluation, and the optimal number of channels assessment. Moreover, the results for each dataset are presented separately.

4.1. Feature Selection

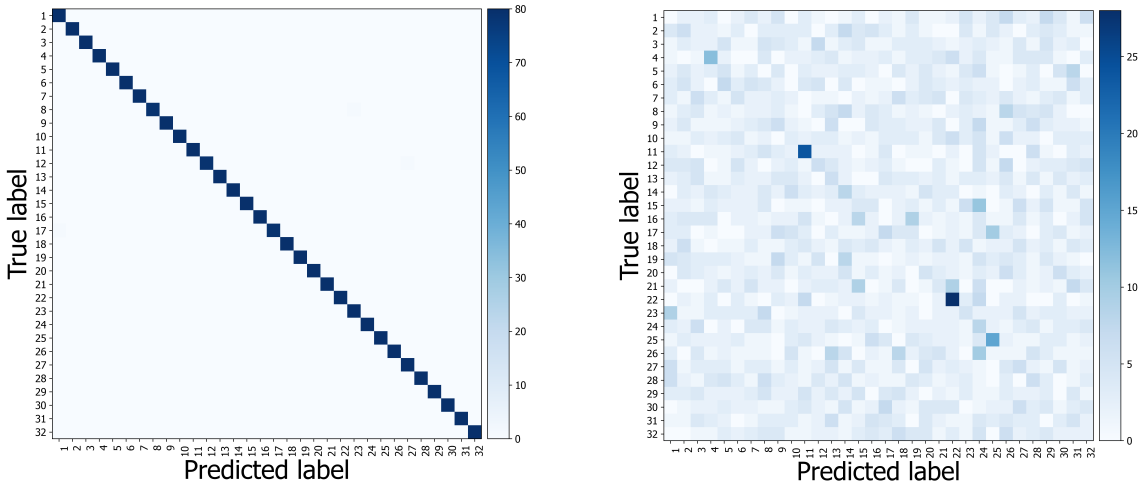
Table 4 shows the performance for each classifier using the 19 features mentioned in Section 3.2 for the case of DEAP dataset. The features are ordered from best to worst based on the best classification results. The best feature was Standard Deviation per Wavelet Coefficient using a MLP, while the worst was Skewness per Wavelet Coefficient using RF. Figure 1 illustrates clearly the difference between those two features by presenting the confusion matrix of the MLP using both of them.

Table 4: Classification results for each features using DEAP dataset. Bold values are the best for the specific feature.

Feature	**Classifier**	**Sensitivity (%)**	**Specificity (%)**	**Accuracy (%)**
Standard	**MLP**	**99.96±0.12**	**100.0±0.00**	**100.0±0.01**
Deviation	AB	99.14±0.29	99.97±0.01	99.95±0.02
per Wavelet	RF	98.87±0.32	99.96±0.01	99.93±0.02
Coefficient	SVM	99.77±0.26	99.99±0.01	99.99±0.02
	KNN	99.41±0.31	99.98±0.01	99.96±0.02
Energy	**MLP**	**99.61±0.30**	**99.99±0.01**	**99.98±0.02**
per Wavelet	AB	98.44±0.43	99.95±0.01	99.90±0.03
Coefficient	RF	98.83±0.43	99.96±0.01	99.93±0.03
	SVM	94.88±0.92	99.83±0.03	99.68±0.06
	KNN	98.83±0.35	99.96±0.01	99.93±0.02
PSD	**MLP**	**99.49±0.43**	**99.98±0.01**	**99.97±0.03**
	AB	98.55±0.50	99.95±0.02	99.91±0.03
	RF	98.52±0.57	99.95±0.02	99.91±0.04
	SVM	96.99±0.86	99.90±0.03	99.81±0.05
	KNN	98.98±0.40	99.97±0.01	99.94±0.02
Variance	**MLP**	**99.34±0.46**	**99.98±0.01**	**99.96±0.03**
per Wavelet	AB	99.10±0.25	99.97±0.01	99.94±0.02
Coefficient	RF	98.40±0.51	99.95±0.02	99.90±0.03
	SVM	94.88±0.92	99.83±0.03	99.68±0.06
	KNN	98.79±0.37	99.96±0.01	99.92±0.02
Max	**MLP**	**99.22±0.43**	**99.97±0.01**	**99.95±0.03**
per Wavelet	AB	98.52±0.42	99.95±0.01	99.91±0.03
Coefficient	RF	98.05±0.43	99.94±0.01	99.88±0.03
	SVM	98.67±0.50	99.96±0.02	99.92±0.03
	KNN	98.75±0.65	99.96±0.02	99.92±0.04
Entropy	MLP	98.52±0.62	99.95±0.02	99.91±0.04
per Wavelet	**AB**	**99.14±0.29**	**99.97±0.01**	**99.95±0.02**
Coefficient	RF	98.13±0.42	99.94±0.01	99.88±0.03
	SVM	90.08±2.10	99.68±0.07	99.38±0.13
	KNN	98.24±0.44	99.94±0.01	99.89±0.03
Min	MLP	83.05±26.7	99.45±0.86	98.94±1.67
per Wavelet	AB	98.52±0.49	99.95±0.02	99.91±0.03
Coefficient	RF	98.40±0.54	99.95±0.02	99.90±0.03

Feature	Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)
	SVM	**98.87±0.41**	**99.96±0.01**	**99.93±0.03**
	KNN	98.59±0.50	99.95±0.02	99.91±0.03
TAP	**MLP**	**98.44±0.49**	**99.95±0.02**	**99.9±0.03**
	AB	97.85±0.98	99.93±0.03	99.87±0.06
	RF	97.81±0.94	99.93±0.03	99.86±0.06
	SVM	88.79±1.30	99.64±0.04	99.30±0.08
	KNN	98.20±0.89	99.94±0.03	99.89±0.06
RWE	**MLP**	**90.27±3.37**	**99.69±0.11**	**99.39±0.21**
	AB	83.87±2.75	99.48±0.09	98.99±0.17
	RF	81.91±2.21	99.42±0.07	98.87±0.14
	SVM	89.41±0.83	99.66±0.03	99.34±0.05
	KNN	75.51±1.45	99.21±0.05	98.47±0.09
Relative Power	**MLP**	**89.77±2.29**	**99.67±0.07**	**99.36±0.14**
	AB	79.45±1.93	99.34±0.06	98.72±0.12
	RF	78.63±2.51	99.31±0.08	98.66±0.16
	SVM	87.97±1.37	99.61±0.04	99.25±0.09
	KNN	78.63±2.51	99.31±0.08	98.66±0.16
Mean	MLP	43.05±14.4	98.16±0.47	96.44±0.91
per Wavelet	AB	83.87±1.55	99.48±0.05	98.99±0.10
Coefficient	**RF**	**89.69±1.57**	**99.67±0.05**	**99.36±0.10**
	SVM	38.83±2.12	98.03±0.07	96.18±0.13
	KNN	63.55±3.08	98.82±0.10	97.72±0.19
TsEn	MLP	14.92±2.69	97.26±0.09	94.68±0.17
	AB	82.85±1.96	99.45±0.06	98.93±0.12
	RF	**88.52±1.84**	**99.63±0.06**	**99.28±0.12**
	SVM	11.95±1.39	97.16±0.04	94.50±0.09
	KNN	76.80±1.86	99.25±0.06	98.55±0.12
MFreq	**MLP**	**79.69±2.78**	**99.34±0.09**	**98.73±0.17**
	AB	75.94±2.59	99.22±0.08	98.50±0.16
	RF	76.29±3.27	99.24±0.11	98.52±0.20
	SVM	77.34±2.35	99.27±0.08	98.58±0.15
	KNN	74.26±2.33	99.17±0.08	98.39±0.15
Median	MLP	29.30±3.04	97.72±0.10	95.58±0.19
per Wavelet	**AB**	**69.53±2.21**	**99.02±0.07**	**98.1±0.14**
Coefficient	RF	68.95±2.03	99.00±0.07	98.06±0.13
	SVM	25.35±2.19	97.59±0.07	95.33±0.14
	KNN	26.21±2.25	97.62±0.07	95.39±0.14
HFD	MLP	62.97±18.4	98.81±0.59	97.69±1.15
	AB	66.84±1.77	98.93±0.06	97.93±0.11
	RF	**66.87±1.88**	**98.93±0.06**	**97.93±0.12**
	SVM	64.77±2.26	98.86±0.07	97.80±0.14
	KNN	63.71±2.02	98.83±0.07	97.73±0.13

Feature	Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)
LZC	MLP	48.24±1.80	98.33±0.06	96.77±0.11
	AB	54.14±2.11	98.52±0.07	97.13±0.13
	RF	54.10±3.07	98.52±0.10	97.13±0.19
	SVM	**57.19±1.73**	**98.62±0.06**	**97.32±0.11**
	KNN	47.30±2.48	98.30±0.08	96.71±0.15
ApEn	MLP	34.02±2.76	97.87±0.09	95.88±0.17
	AB	**36.95±2.06**	**97.97±0.07**	**96.06±0.13**
	RF	36.60±2.37	97.95±0.08	96.04±0.15
	SVM	33.87±1.74	97.87±0.06	95.87±0.11
	KNN	29.41±1.42	97.72±0.05	95.59±0.09
SampEn	MLP	12.11±1.63	97.16±0.05	94.51±0.10
	AB	11.76±1.96	97.15±0.06	94.48±0.12
	RF	**17.73±2.83**	**97.35±0.09**	**94.86±0.18**
	SVM	11.60±1.69	97.15±0.05	94.48±0.11
	KNN	8.36±1.33	97.04±0.04	94.27±0.08
Skewness per Wavelet Coefficient	MLP	7.15±1.69	97.00±0.05	94.20±0.11
	AB	8.98±1.17	97.06±0.04	94.31±0.07
	RF	**13.36±1.55**	**97.21±0.05**	**94.58±0.10**
	SVM	9.73±1.79	97.09±0.06	94.36±0.11
	KNN	12.03±1.90	97.16±0.06	94.50±0.12



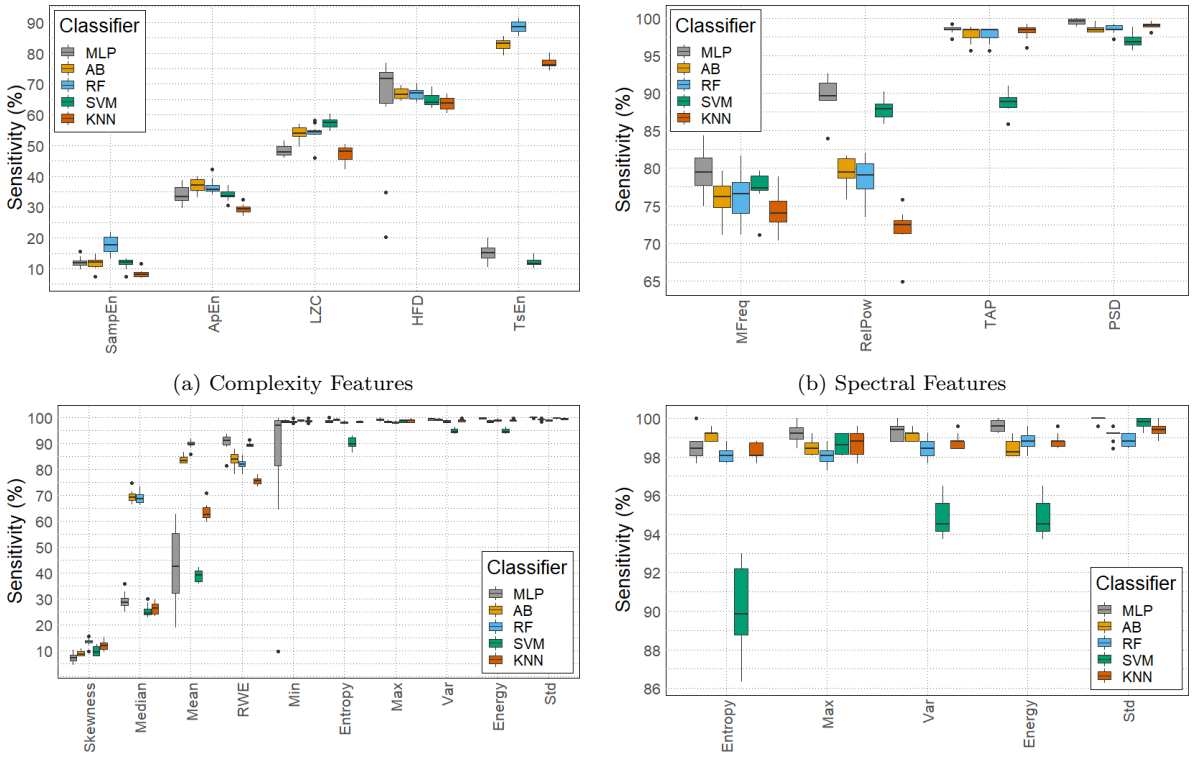
(a) Standard Deviation per Wavelet Coefficient

(b) Skewness per Wavelet Coefficient

Figure 1: Confusion matrices of the best and worst feature for DEAP dataset using the MLP classifier.

For having a better visualization of the results of Table 4, Figure 2 shows a graphic representation of the sensitivity of the classifiers with each feature. Only the sensitivity is shown because it was the lowest performance metric. This figure shows that even though complexity features can be useful for other applications such as medical applications, their not seem to be the most suitable choice for developing biometric systems. On the contrary, some of the spectral and wavelet features appeared to be the best option

for developing these systems. However, it is worth mentioning that not all wavelet-based features lead to high results because the worst feature (skewness) belongs to this category.



(a) Complexity Features

(b) Spectral Features

(c) Wavelet Features

(d) Best Wavelet Features

Figure 2: Boxplots of the feature categories grouped by classifier using DEAP dataset.

For BIOMEX-DB, Table 5 contains the classification results for each feature with each classifier. As in the case of DEAP, the best feature was Standard Deviation using a MLP and the worst feature was the Skewness using KNN. The difference between these two features using the MLP is depicted in Figure 3.

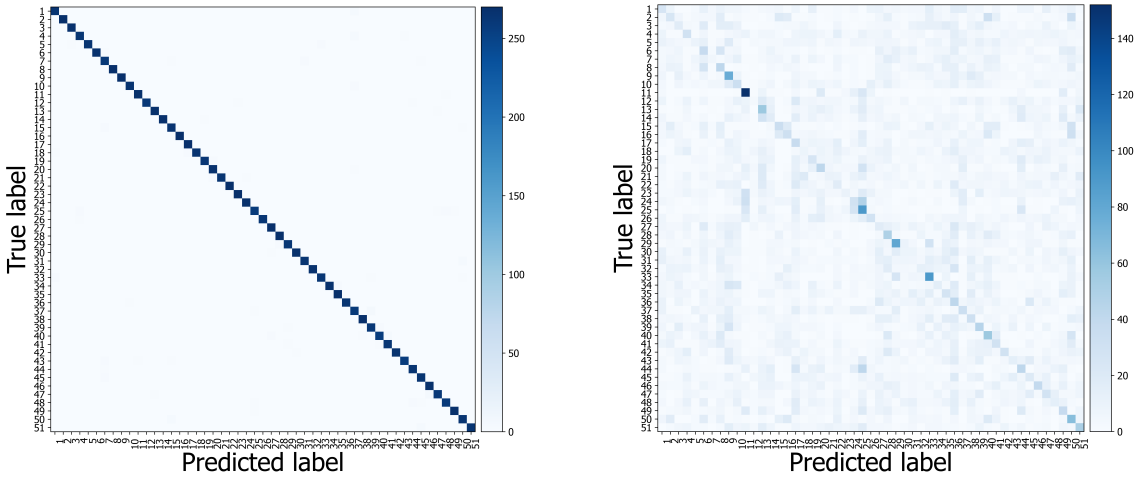
Table 5: Classification results for each features using BIOMEX-DB dataset. Bold values the best for the specific feature.

Feature	**Classifier**	**Sensitivity (%)**	**Specificity (%)**	**Accuracy (%)**
Standard	**MLP**	**98.21±0.30**	**99.96±0.01**	**99.93±0.01**
Deviation	AB	96.66±0.42	99.93±0.01	99.87±0.02
per Wavelet	RF	96.64±0.28	99.93±0.01	99.87±0.01
Coefficient	SVM	96.89±0.39	99.94±0.01	99.88±0.02
	KNN	95.99±0.15	99.92±0.00	99.84±0.01
Energy	MLP	95.34±0.52	99.91±0.01	99.82±0.02
per Wavelet	**AB**	**97.11±0.38**	**99.94±0.01**	**99.89±0.01**
Coefficient	RF	97.04±0.39	99.94±0.01	99.88±0.02
	SVM	83.89±1.10	99.68±0.02	99.37±0.04
	KNN	88.03±0.79	99.76±0.02	99.53±0.03
Entropy	MLP	92.68±0.63	99.85±0.01	99.71±0.02

Feature	Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)
per Wavelet	AB	96.86±0.33	99.94±0.01	99.88±0.01
Coefficient	**RF**	**96.92±0.34**	**99.94±0.01**	**99.88±0.01**
	SVM	70.28±2.41	99.41±0.05	98.83±0.09
	KNN	81.42±1.07	99.63±0.02	99.27±0.04
Variance	**MLP**	**96.86±0.29**	**99.94±0.01**	**99.88±0.01**
per Wavelet	AB	95.44±0.34	99.91±0.01	99.82±0.01
Coefficient	RF	96.65±0.29	99.93±0.01	99.87±0.01
	SVM	93.85±0.24	99.88±0.00	99.76±0.01
	KNN	93.53±0.88	99.87±0.02	99.75±0.03
RWE	MLP	94.17±0.70	99.88±0.01	99.77±0.03
	AB	95.81±0.46	99.92±0.01	99.84±0.02
	RF	**95.90±0.38**	**99.92±0.01**	**99.84±0.02**
	SVM	70.78±2.86	99.42±0.06	98.85±0.11
	KNN	93.26±1.02	99.87±0.02	99.74±0.04
Max	MLP	89.32±0.66	99.79±0.01	99.58±0.03
per Wavelet	**AB**	**93.73±0.65**	**99.87±0.01**	**99.75±0.03**
Coefficient	RF	93.70±0.63	99.87±0.01	99.75±0.02
	SVM	90.15±0.50	99.80±0.01	99.61±0.02
	KNN	86.71±0.98	99.73±0.02	99.48±0.04
Min	MLP	25.56±6.74	98.51±0.13	97.08±0.26
per Wavelet	AB	93.13±0.45	99.86±0.01	99.73±0.02
Coefficient	**RF**	**93.21±0.59**	**99.86±0.01**	**99.73±0.02**
	SVM	91.04±0.66	99.82±0.01	99.65±0.03
	KNN	86.27±0.80	99.73±0.02	99.46±0.03
Relative Power	MLP	90.44±0.94	99.81±0.02	99.62±0.04
	AB	91.55±0.88	99.83±0.02	99.67±0.03
	RF	**91.61±0.78**	**99.83±0.02**	**99.67±0.03**
	SVM	82.29±1.07	99.65±0.02	99.31±0.04
	KNN	83.07±0.94	99.66±0.02	99.34±0.04
TAP	MLP	45.24±2.06	98.90±0.04	97.85±0.08
	AB	70.34±1.18	99.41±0.02	98.84±0.05
	RF	**76.18±1.23**	**99.52±0.02**	**99.07±0.05**
	SVM	67.33±1.86	99.35±0.04	98.72±0.07
	KNN	60.43±1.30	99.21±0.03	98.45±0.05
TsEn	MLP	28.91±2.73	98.58±0.05	97.21±0.11
	AB	70.61±0.97	99.41±0.02	98.85±0.04
	RF	**76.06±0.75**	**99.52±0.02**	**99.06±0.03**
	SVM	66.52±1.19	99.33±0.02	98.69±0.05
	KNN	60.50±0.96	99.21±0.02	98.45±0.04
Median	MLP	26.14±3.48	98.52±0.07	97.10±0.14
per Wavelet	**AB**	**75.09±1.31**	**99.5±0.03**	**99.02±0.05**
Coefficient	RF	75.01±1.25	99.50±0.02	99.02±0.05

Feature	Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)
	SVM	17.48±1.02	98.35±0.02	96.76±0.04
	KNN	7.36±0.37	98.15±0.01	96.37±0.01
Mean	MLP	8.08±6.28	98.16±0.13	96.4±0.25
per Wavelet	AB	70.86±1.22	99.42±0.02	98.86±0.05
Coefficient	**RF**	**74.17±1.36**	**99.48±0.03**	**98.99±0.05**
	SVM	36.19±1.30	98.72±0.03	97.50±0.05
	KNN	14.26±0.66	98.29±0.01	96.64±0.03
HFD	**MLP**	**63.12±0.91**	**99.26±0.02**	**98.55±0.04**
	AB	49.32±1.23	98.99±0.02	98.01±0.05
	RF	58.26±0.76	99.17±0.02	98.36±0.03
	SVM	61.13±0.84	99.22±0.02	98.48±0.03
	KNN	49.64±0.92	98.99±0.02	98.03±0.04
PSD	MLP	32.67±2.71	98.65±0.05	97.36±0.11
	AB	**62.98±1.05**	**99.26±0.02**	**98.55±0.04**
	RF	62.78±0.91	99.26±0.02	98.54±0.04
	SVM	55.91±1.57	99.12±0.03	98.27±0.06
	KNN	48.83±1.17	98.98±0.02	97.99±0.05
ApEn	**MLP**	**61.05±1.32**	**99.22±0.03**	**98.47±0.05**
	AB	61.02±1.02	99.22±0.02	98.47±0.04
	RF	60.88±0.88	99.22±0.02	98.47±0.03
	SVM	60.09±1.31	99.20±0.03	98.44±0.05
	KNN	56.56±0.79	99.13±0.02	98.30±0.03
SampEn	MLP	45.90±1.37	98.92±0.03	97.88±0.05
	AB	49.24±0.93	98.98±0.02	98.01±0.04
	RF	**52.61±1.09**	**99.05±0.02**	**98.14±0.04**
	SVM	45.90±0.60	98.92±0.01	97.88±0.02
	KNN	38.63±0.76	98.77±0.02	97.59±0.03
MFreq	MLP	40.85±1.04	98.82±0.02	97.68±0.04
	AB	41.20±0.49	98.82±0.01	97.69±0.02
	RF	**42.03±0.75**	**98.84±0.02**	**97.73±0.03**
	SVM	32.02±1.30	98.64±0.03	97.33±0.05
	KNN	36.47±0.67	98.73±0.01	97.51±0.03
LZC	MLP	35.85±1.63	98.72±0.03	97.48±0.06
	AB	20.59±18.1	98.41±0.36	96.89±0.71
	RF	39.58±0.93	98.79±0.02	97.63±0.04
	SVM	**40.87±1.30**	**98.82±0.03**	**97.68±0.05**
	KNN	34.56±0.86	98.69±0.02	97.43±0.03
Skewness	MLP	11.55±0.88	98.23±0.02	96.53±0.03
per Wavelet	AB	18.53±0.64	98.37±0.01	96.81±0.03
Coefficient	**RF**	**18.62±0.73**	**98.37±0.01**	**96.81±0.03**
	SVM	13.27±0.88	98.27±0.02	96.60±0.03

Feature	Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)
	KNN	9.28±0.35	98.19±0.01	96.44±0.01



(a) Standard Deviation per Wavelet Coefficient

(b) Skewness per Wavelet Coefficient

Figure 3: Confusion matrices of the best and worst feature for BIOMEX-DB dataset using the MLP classifier.

As with DEAP, Figure 4 shows the sensitivity boxplots of the classifiers with each feature, but for the BIOMEX-DB dataset. For this dataset, complexity and spectral features did not lead to results as good as those obtained with some wavelet features. Besides, this figure helped to verify that the performance of the system heavily relies on the classifier. For example, the performance using the minimum of the DWT coefficients varies a lot using MLP regarding RF.

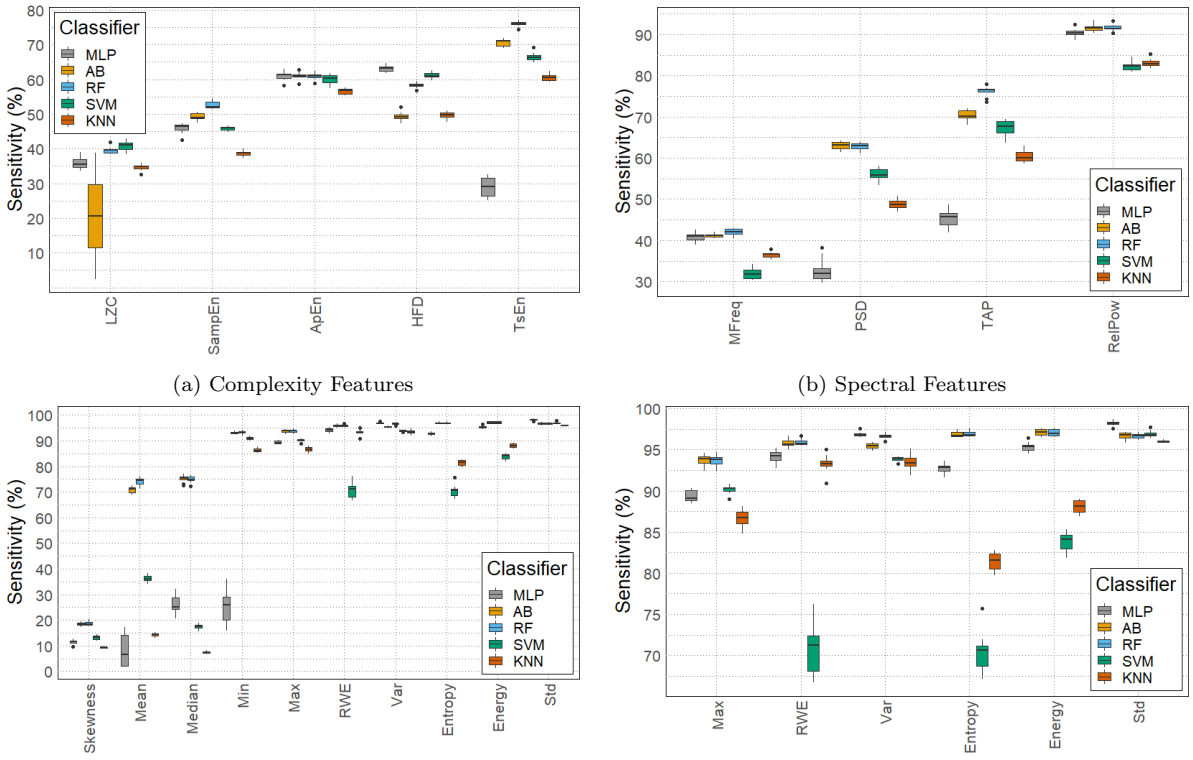
4.2. Single-channel Evaluation

For evaluating each channel of the DEAP dataset, the features that fulfilled the requirements of a "good" feature, explained in Section 3.4.1, were mainly wavelet-based features: Std, Energy, Var, Max, Entropy, and Min, but there were also two spectral features: PSD and TAP. Using these features and the optimized MLP (Table 3), the single-channel evaluation was performed. Table 6 exhibits the 32 available channels ordered from best to worst based on their correct classification rate. Figure 5 aids to visualize how the channels were placed and the order found with the MLP. By analyzing this figure, the best channels seem to be those located in the central area of the scalp.

On the other hand, the "good" features for the BIOMEX-DB dataset were one spectral feature (RelPow) and seven wavelet-based features: Std, Energy, Entropy, Var, RWE, Max, and Min. Table 7 presents the 14 available channels ordered descending by their CCR. As with DEAP, for better visualization of the results, Figure 6 indicates the location of the channels and their corresponding order for the BIOMEX-DB dataset. For this dataset, the best channels tend to be located on the right side of the scalp.

4.3. Optimal Number of Channels

Figure 7 illustrates the results of the analysis of the number of channels starting with one until reaching all available channels following the order presented in Table 6 and a



(a) Complexity Features

(b) Spectral Features

(c) Wavelet Features

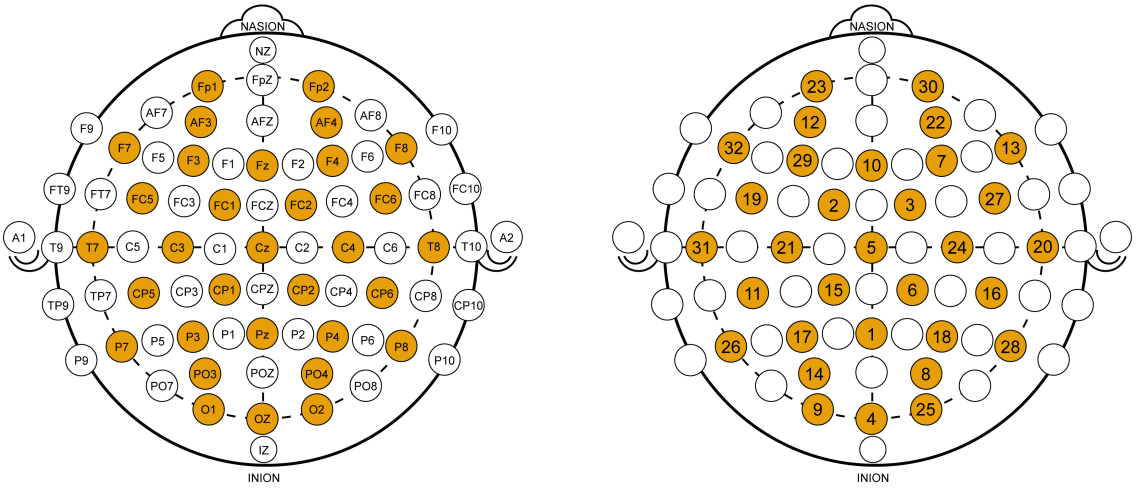
(d) Best Wavelet Features

Figure 4: Boxplots of the feature categories grouped by classifier using BIOMEX-DB dataset.

Table 6: DEAP dataset channels ordered descending by their correct classification rate.

Channel	**Correct Classification Rate (%)**	**Channel**	**Correct Classification Rate (%)**
Pz	39.06±3.40	P3	32.89±2.73
FC1	36.05±4.04	P4	32.58±2.76
FC2	35.90±3.52	FC5	32.58±2.10
Oz	35.16±2.64	T8	32.46±2.29
Cz	35.12±3.31	C3	32.15±2.10
CP2	35.04±1.26	AF4	31.84±2.71
F4	35.00±2.76	Fp1	31.64±2.49
PO4	34.88±1.86	C4	31.17±2.12
O1	34.49±2.68	O2	31.05±3.19
Fz	34.34±1.94	P7	31.02±3.40
CP5	34.06±3.01	FC6	30.94±1.49
AF3	33.98±1.77	P8	30.78±2.34
F8	33.67±1.70	F3	30.51±2.27
PO3	33.48±2.61	Fp2	30.08±3.05
CP1	33.24±2.52	T7	29.10±1.98
CP6	33.12±2.07	F7	26.45±1.65

random ordering. As before, only the sensitivity is presented due to this metric was the lowest. This figure encouraged the subsequent analysis to find the optimum number of



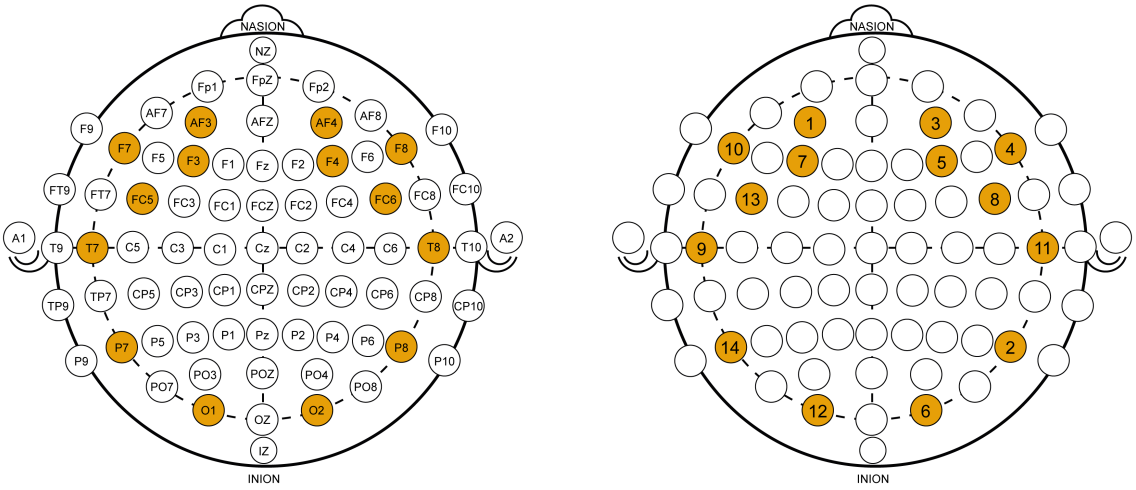
(a) Channels position

(b) Channels order

Figure 5: Channels position and order for the DEAP dataset. In (b), 1 corresponds to the best channel while 32 to the worst.

Table 7: BIOMEX-DB dataset channels ordered descending by their correct classification rate.

Channel	**Correct Classification Rate (%)**	**Channel**	**Correct Classification Rate (%)**
AF3	50.60±1.70	FC6	32.69±1.26
P8	38.55±1.58	T7	32.42±1.55
AF4	37.52±1.15	F7	31.87±1.10
F8	34.73±1.63	T8	31.73±0.49
F4	33.46±1.34	O1	30.77±1.89
O2	33.12±1.79	FC5	30.64±1.28
F3	32.92±1.14	P7	27.84±1.05



(a) Channels position

(b) Channels order

Figure 6: Channels position and order for the BIOMEX-DB dataset. In (b), 1 corresponds to the best channel while 14 to the worst.

channels since it appears that there is a number of channels from which no matter if this number increases, the performance remains the same.

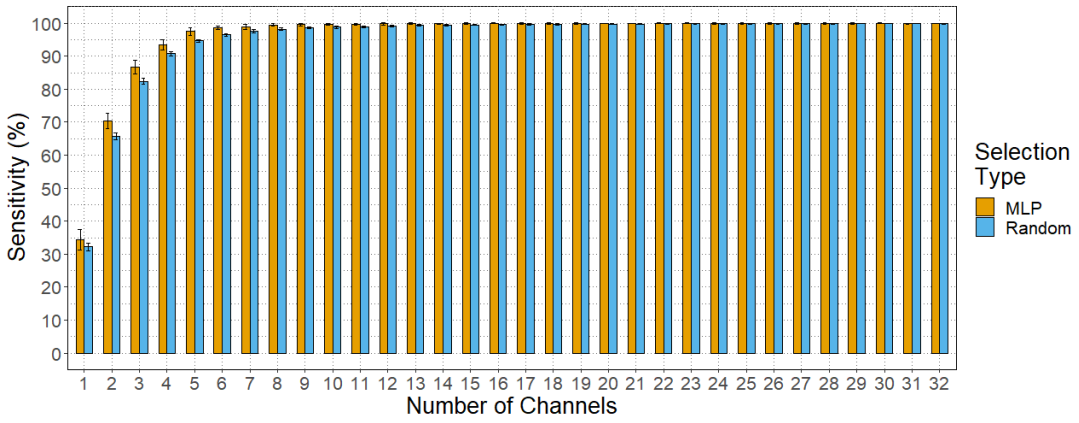


Figure 7: Comparison between proposed method for ordering channels (MLP) with respect to a random ordering using DEAP dataset.

Table 8 contains the results of the statistical analyzes applied in this research, i.e., MANOVA and Wilcoxon Test. In this table, the "Included Results" column has a different meaning for each analysis. For example, for the MANOVA, the row with 29 channels means that the results of using 32, 31, 30, and 29 channels were included during the analysis. In contrast, the same row means that the results of 29 channels using the proposed MLP were compared to random selection results with the Wilcoxon Test.

For the MANOVA, a p-value less than 0.05 means that the number of channels impacts the overall performance of the system. For this reason, the results of this analysis demonstrate that the proposed MLP was better than randomly selecting the order of the channels because by using the proposed MLP, the optimal number of channels was 11, while by using the random selection, this number increases to 25. On the other hand, for the Wilcoxon Test, a p-value less than 0.05 is interpreted as a statistically significant difference between the evaluated methods. These results verified that there is a difference between the results of the two methods in almost all the cases proving the efficiency of the proposed selection method.

Due to the results mentioned above, the suggested biometric system for the DEAP dataset has the following characteristics: EEG recordings of 1.75 seconds using the best 11 channels selected by the proposed MLP, and standard deviation extracted from a three-level DWT as a feature. Table 9 presents a comparison between the results of all classifiers using the proposed number of channels regarding using all available channels. In both cases, the the classifiers were evaluated using the standard deviation per wavelet coefficient. Despite the difference in the classification performance between using all the available channels and the proposed optimum, the previously mentioned MANOVA demonstrated that this difference is not statistically significant. Figure 8 illustrates the difference in the sensitivity of the classifies using both systems.

Table 8: Resulting p-values from statistical analyzes to identify the optimal number of channels and the difference between the two channel selection methods with the standard deviation per wavelet coefficient as feature for DEAP dataset. Bold values are p-values less than 0.05. In the case of MANOVA, the values with * correspond to the optimal number of channels.

	MANOVA		**Wilcoxon Test**
Selection Method	**MLP**	**Random**	
Included Results	**p-value**	**p-value**	**p-value**
31 channels	0.192	0.897	0.236
30 channels	0.981	0.801	**0.032**
29 channels	0.903	0.423	0.735
28 channels	0.969	0.710	0.138
27 channels	0.982	0.993	0.483
26 channels	0.987	0.781	0.211
25 channels	0.976	0.423*	0.481
24 channels	0.936	**0.029**	0.107
23 channels	0.999	**0.018**	**0.02**
22 channels	0.991	**0.005**	**0.006**
21 channels	0.515	**3.19e-04**	**0.002**
20 channels	0.577	**2.06e-05**	**0.012**
19 channels	0.737	**2.65e-07**	**0.01**
18 channels	0.873	**3.74e-10**	**0.004**
17 channels	0.924	**1.29e-13**	**0.004**
16 channels	0.902	**2.2e-16**	**0.002**
15 channels	0.344	**2.2e-16**	**0.02**
14 channels	0.556	**2.2e-16**	**0.002**
13 channels	0.422	**2.2e-16**	**0.002**
12 channels	0.221	**2.2e-16**	**0.002**
11 channels	0.076*	**2.2e-16**	**0.002**
10 channels	**0.004**	**2.2e-16**	**0.002**
9 channels	**2.20e-05**	**2.2e-16**	**0.002**
8 channels	**1.06e-11**	**2.2e-16**	**0.002**
7 channels	**1.03e-11**	**2.2e-16**	**0.002**
6 channels	**2.2e-16**	**2.2e-16**	**0.002**
5 channels	**2.2e-16**	**2.2e-16**	**0.002**
4 channels	**2.2e-16**	**2.2e-16**	**0.002**
3 channels	**2.2e-16**	**2.2e-16**	**0.002**
2 channels	**2.2e-16**	**2.2e-16**	**0.002**
1 channel	**2.2e-16**	**2.2e-16**	0.084

In the case of the BIOMEX-DB dataset, Figure 9 illustrates the results of the analysis of the number of channels using the proposed order selected by a MLP and a random selection method. Although an optimal number of channels less than the total available channels is hard to identify in this figure, the same statistical analyzes used with the DEAP dataset were applied.

Table 10 exhibits the resulting p-values of the statistical analyzes. In this table, the "Included Results" column has the same meaning explained above. For this dataset, the

Table 9: Comparison between the classifiers performance using the proposed optimal number of channels and the total available channels with the standard deviation per wavelet coefficient as feature for DEAP dataset.

		MLP	AB	RF	SVM	KNN
Sensitivity (%)	All Channels	100.0±0.0	99.1±0.39	98.91±0.38	99.77±0.26	99.41±0.31
	11 Channels	99.77±0.26	98.28±0.63	98.01±0.9	98.98±0.5	98.44±0.65
Specificity (%)	All Channels	100.0±0.0	99.97±0.01	99.96±0.01	99.99±0.01	99.98±0.01
	11 Channels	99.99±0.01	99.94±0.02	99.94±0.03	99.97±0.02	99.95±0.02
Accuracy (%)	All Channels	100.0±0.0	99.94±0.02	99.93±0.02	99.99±0.02	99.96±0.02
	11 Channels	99.99±0.02	99.89±0.04	99.88±0.06	99.94±0.03	99.9±0.04

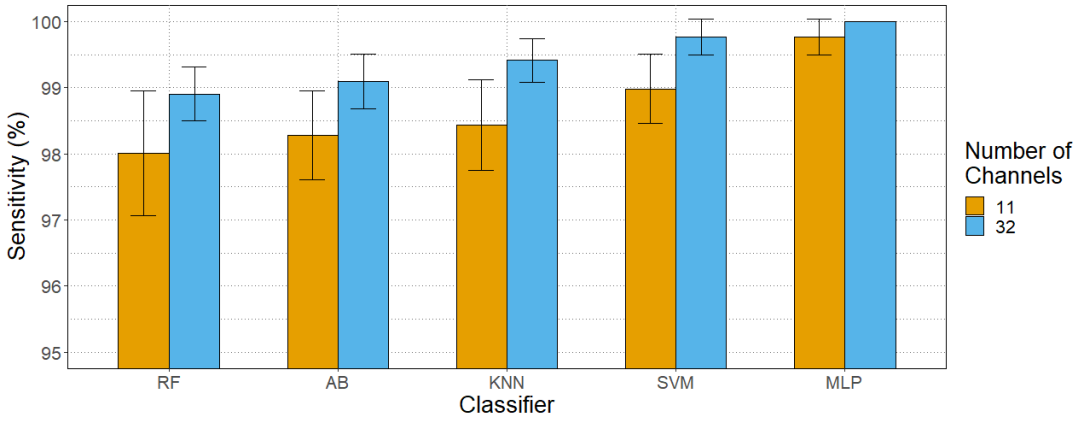


Figure 8: Comparison between the classifiers performance using the the proposed optimum number of channels (11) with respect to all available channels (32) using DEAP dataset.

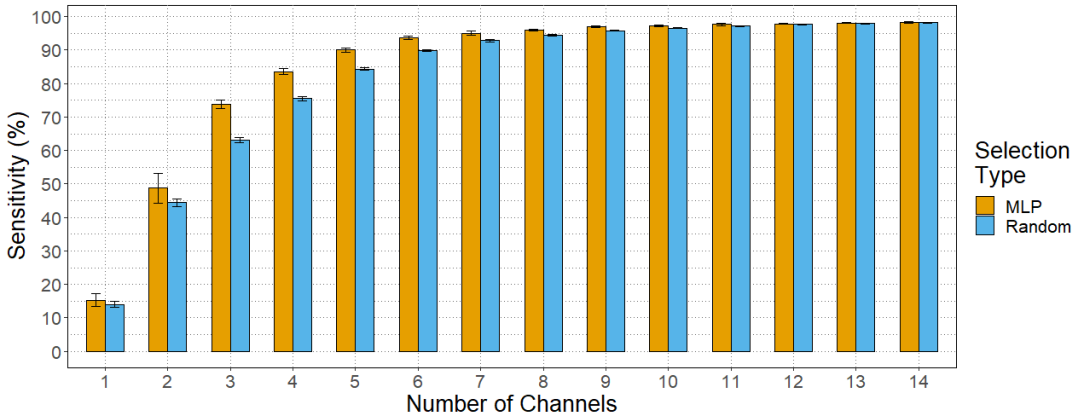


Figure 9: Comparison between proposed method for ordering channels (MLP) with respect to a random ordering using BIOMEX-DB dataset.

results of the MANOVA indicate that for both methods, the optimal number of channels is 13. Despite this result, in general, the p-values of the proposed MLP were less than the ones of the random selection. Moreover, the equality in the optimum could be because, in this dataset, the channels used for recording the EEGs were located at the left and right zones of the scalp. However, the results presented in Section 4.2 for the DEAP dataset, which uses more EEG channels distributed over a wider area, indicate that the

best channels are located on the center of the scalp.

Besides, the Wilcoxon test results verified that the two evaluated selection methods are statistically different in almost all scenarios. The only scenario where these methods were not statistically different was when working with a single channel. This result was expected since when working with a single-channel biometric system, it is extremely difficult for the results of using a specific channel to be better than the results of using any other channel.

Table 10: Resulting p-values from statistical analyzes to identify the optimal number of channels and the difference between the two channel selection methods with the standard deviation per wavelet coefficient as feature for BIOMEX-DB dataset. Values in bold are p-values less than 0.05. In the case of MANOVA, the values with * correspond to the optimal number of channels.

	MANOVA		**Wilcoxon Test**
Selection Method	**MLP**	**Random**	
Included Results	**p-value**	**p-value**	**p-value**
13 channels	0.064*	0.068*	**0.037**
12 channels	**0.001**	**2.5e-05**	**0.02**
11 channels	**6.37e- 06**	**5.87e-12**	**0.01**
10 channels	**4.84e-09**	**2.2e-16**	**0.002**
9 channels	**1.27e-15**	**2.2e-16**	**0.003**
8 channels	**2.2e-16**	**2.2e-16**	**0.004**
7 channels	**2.2e-16**	**2.2e-16**	**0.005**
6 channels	**2.2e-16**	**2.2e-16**	**0.006**
5 channels	**2.2e-16**	**2.2e-16**	**0.007**
4 channels	**2.2e-16**	**2.2e-16**	**0.008**
3 channels	**2.2e-16**	**2.2e-16**	**0.009**
2 channels	**2.2e-16**	**2.2e-16**	**0.027**
1 channels	**2.2e-16**	**2.2e-16**	0.375

Resulting from the above mentioned results, the proposed biometric system for the BIOMEX-DB dataset has the following characteristics: EEG recordings of 1.75 seconds using the best 13 channels selected by the proposed MLP, and standard deviation extracted from a three-level DWT as a feature. As with DEAP, Table 11 contains the results of the comparison between the proposed biometric system regarding the biometric developed using the 14 EEG channels. A graphic representation of this comparison is shown in Figure 10.

Table 11: Comparison between the classifiers performance using the proposed optimal number of channels and the total available channels with the standard deviation per wavelet coefficient as feature for BIOMEX-DB dataset.

		MLP	**AB**	**RF**	**SVM**	**KNN**
Sensitivity (%)	All Channels	98.23±0.26	96.7±0.35	96.68±0.27	96.89±0.39	95.99±0.15
	13 Channels	98.08±0.15	96.3±0.37	96.29±0.48	96.59±0.34	95.59±0.22
Specificity (%)	All Channels	99.96±0.01	99.93±0.01	99.93±0.01	99.94±0.01	99.92±0.0
	13 Channels	99.96±0.0	99.93±0.01	99.93±0.01	99.93±0.01	99.91±0.0
Accuracy (%)	All Channels	99.96±0.01	99.87±0.01	99.87±0.01	99.88±0.02	99.84±0.01
	13 Channels	99.92±0.01	99.86±0.01	99.85±0.02	99.87±0.01	99.83±0.01

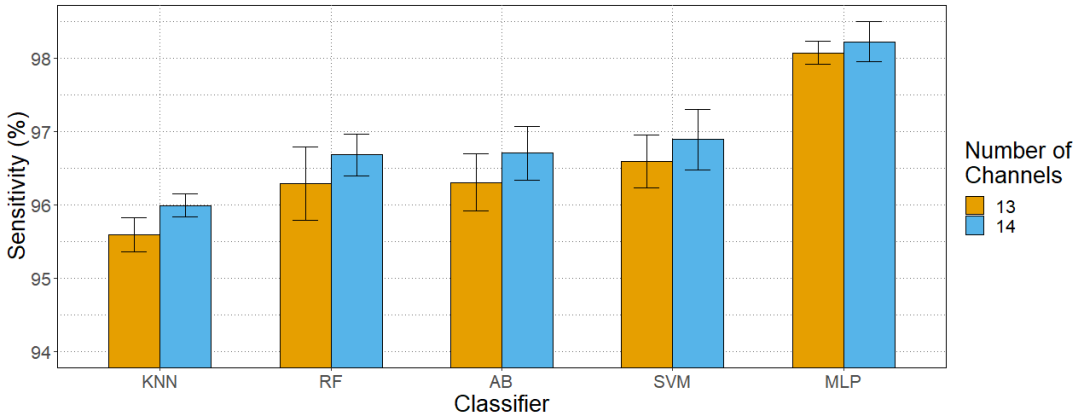


Figure 10: Comparison between the classifiers performance using the the proposed optimum number of channels (13) with respect to all available channels (14) using BIOMEX-DB dataset.

5. Conclusions and Future Work

This research proposed a methodology to select the best feature and the optimal number of channels for developing a biometric system based on EEG analysis. For proving the efficiency of the proposed methodology, this study worked with two datasets independently. Moreover, nineteen features belonging to three categories (spectral, complexity, and wavelet-based) were evaluated in this investigation.

The experimental results using two different datasets demonstrated that wavelet-based features might be the best option for developing biometric systems based on EEG signals. Additionally, the standard deviation per wavelet coefficient proved to be the most efficient feature, among all the analyzed features in this study, for effectively represent the differences between the subjects. On the contrary, although the complexity features are commonly used in medical applications, in general, they led to the lowest performance in this study.

Additionally, the evaluation of the EEG channels showed that it is not necessary to use all available ones. For instance, in the case of the DEAP dataset, using approximately one-third of the available channels produced results statistically equal to those obtained with all channels. This reduction in the number of channels is highly beneficial since it decreases the computational cost and increases the portability of the final system. Moreover, the experimental results seem to indicate that the best channels for developing biometric systems are located on the center of the scalp.

As future work, an in-depth analysis of the standard deviation should be performed to understand why this feature can effectively represent the differential factors of each subject. Besides, further analysis of the impact of the EEG channels' location over the overall performance of the system can significantly increase the understanding of the factors that influence the performance of EEG-based biometric systems. Finally, as the MLP was the best classifier for both datasets, different architectures may be evaluated, including Deep Neural Networks.

References

Abásolo, D., Escudero, J., Hornero, R., Gómez, C., & Espino, P. (2008). Approximate entropy and auto mutual information analysis of the electroencephalogram in alzheimer's disease patients. *Medical & Biological Engineering & Computing*, *46*, 1019–1028. doi:10.1007/s11517-008-0392-1.

Alotaiby, T., El-Samie, F. E., Alshebeili, S. A., & Ahmad, I. (2015). A review of channel selection algorithms for EEG signal processing. *EURASIP Journal on Advances in Signal Processing*, *2015*. doi:10.1186/s13634-015-0251-9.

Auger, F., & Flandrin, P. (1995). Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions on Signal Processing*, *43*, 1068–1089. doi:10.1109/78.382394.

Bashar, K. (2018). ECG and EEG Based Multimodal Biometrics for Human Identification. In *IEEE International Conference on Systems, Man, and Cybernetics* (pp. 4345–4350). IEEE. doi:10.1109/SMC.2018.00734.

Carrión-Ojeda, D., Fonseca-Delgado, R., & Pineda, I. (2021a). Analysis of factors that influence the performance of biometric systems based on EEG signals. *Expert Systems with Applications*, *165*, 113967. doi:10.1016/j.eswa.2020.113967.

Carrión-Ojeda, D., Martínez-Arias, P., Fonseca-Delgado, R., & Pineda, I. (2021b). EBAPy: A Python framework for analyzing the factors that have an influence in the performance of EEG-based applications. *Software Impacts*, *8*, 100062. doi:10.1016/j.simpa.2021.100062.

Carrión-Ojeda, D., Martínez-Arias, P., Fonseca-Delgado, R., Pineda, I., & Mejía-Vallejo, H. (2021c). Evaluation of features and channels of electroencephalographic signals for biometric systems - source code. https://www.codeocean.com/. doi:10.24433/CO.1541880.v2.

Carrión-Ojeda, D., Mejía-Vallejo, H., Fonseca-Delgado, R., Gómez-Gil, P., & Ramírez-Cortés, M. (2019). A method for studying how much time of EEG recording is needed to have a good user identification. In *IEEE Latin American Conference on Computational Intelligence* (pp. 1–6). IEEE. doi:10.1109/LA-CCI47412.2019.9037054.

Coronel, C., Garn, H., Waser, M., Deistler, M., Benke, T., Dal-Bianco, P., Ransmayr, G., Seiler, S., Grossegger, D., & Schmidt, R. (2017). Quantitative EEG markers of entropy and auto mutual information in relation to MMSE scores of probable alzheimer's disease patients. *Entropy*, *19*, 130. doi:10.3390/e19030130.

Dauwels, J., Vialatte, F., & Cichocki, A. (2010). Diagnosis of alzheimers disease from EEG signals: Where are we standing? *Current Alzheimer Research*, *7*, 487–505. doi:10.2174/156720510792231720.

Gómez, C., Mediavilla, Á., Hornero, R., Abásolo, D., & Fernández, A. (2009). Use of the higuchi's fractal dimension for the analysis of MEG recordings from alzheimer's disease patients. *Medical Engineering & Physics*, *31*, 306–313. doi:10.1016/j.medengphy.2008.06.010.

Gui, Q., Ruiz-Blondet, M. V., Laszlo, S., & Jin, Z. (2019). A Survey on Brain Biometrics. *ACM Computing Surveys*, *51*. doi:10.1145/3230632.

Hamad, A., Houssein, E., Hassanien, A., & Fahmy, A. (2016). Feature Extraction of Epilepsy EEG using Discrete Wavelet Transform. In *12th International Computer Engineering Conference* (pp. 190–195). IEEE. doi:10.1109/ICENCO.2016.7856467.

Kang, J.-H., Jo, Y. C., & Kim, S.-P. (2018). Electroencephalographic feature evaluation for improving personal authentication performance. *Neurocomputing*, *287*, 93–101. doi:10.1016/j.neucom.2018.01.074.

Kang, J.-H., Lee, C. H., & Kim, S.-P. (2016). EEG feature selection and the use of Lyapunov exponents for EEG-based biometrics. In *IEEE-EMBS International Conference on Biomedical and Health Informatics* (pp. 228–231). IEEE. doi:10.1109/BHI.2016.7455876.

Koelstra, S., Mühl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP : A Database for Emotion Analysis using Physiological Signals. *IEEE Transactions on Affective Computing*, *3*, 18–31. doi:10.1109/T-AFFC.2011.15.

Lee, C., Kang, J.-H., & Kim, S.-P. (2016). Feature slection using mutual information for EEG-based biometrics. In *39th International Conference on Telecommunications and Signal Processing* (pp. 673–676). IEEE. doi:10.1109/TSP.2016.7760968.

Liu, S., Bai, Y., Liu, J., Qi, H., Li, P., Zhao, X., Zhou, P., Zhang, L., Wan, B., Wang, C., Li, Q., Jiao, X., Chen, S., & Ming, D. (2014). Individual Feature Extraction and Identification on EEG Signals in Relax and Visual Evoked Tasks. *Communications in Computer and Information Science*, *404*, 305–318. doi:10.1007/978-3-642-54121-6_29.

Matuszyk, P., Castillo, R., Kottke, D., & Spiliopoulou, M. (2016). A Comparative Study on Hyperparameter Optimization for Recommender Systems. In *Workshop on Recommender Systems and Big Data Analytics* (pp. 13 – 21).

McBride, J. C., Zhao, X., Munro, N. B., Smith, C. D., Jicha, G. A., Hively, L., Broster, L. S., Schmitt, F. A., Kryscio, R. J., & Jiang, Y. (2014). Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early alzheimer's disease. *Computer Methods and Programs in Biomedicine*, *114*, 153–163. doi:10.1016/j.cmpb.2014.01.019.

Phinyomark, A., Thongpanja, S., Hu, H., Phukpattaranont, P., & Limsakul, C. (2012). The usefulness of mean and median frequencies in electromyography analysis. *Computational intelligence in electromyography analysis-A perspective on current applications and future challenges*, (pp. 195–220).

Richman, J. S., Lake, D. E., & Moorman, J. (2004). Sample entropy. In *Methods in Enzymology* (pp. 172–184). Elsevier. doi:10.1016/s0076-6879(04)84011-4.

Sarlabous, L., Torres, A., Fiz, J. A., Gea, J., Martínez-Llorens, J. M., Morera, J., & Jané, R. (2010). Interpretation of the approximate entropy using fixed tolerance values as a measure of amplitude variations in biomedical signals. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (pp. 5967–5970). IEEE.

Schröder, M., Bogdan, M., Hinterberger, T., & Birbaumer, N. (2003). Automated EEG feature selection for brain computer interfaces. In *International IEEE EMBS Conference on Neural Engineering* (pp. 626–629). IEEE. doi:10.1109/CNE.2003.1196906.

Shedeed, H. (2011). A new method for person identification in a biometric security system based on brain EEG signal processing. In *World Congress on Information and Communication Technologies* (pp. 1205–1210). IEEE. doi:10.1109/WICT.2011.6141420.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*, 427–437.

Stoica, P., Moses, R. L. et al. (2005). *Spectral analysis of signals*. Pearson Prentice Hall Upper Saddle River, NJ.

Thongpanja, S., Phinyomark, A., Phukpattaranont, P., & Limsakul, C. (2013). Mean and median frequency of emg signal to determine muscle force based on time-dependent power spectrum. *Elektronika ir Elektrotechnika*, *19*, 51–56. doi:10.5755/j01.eee.19.3.3697.

Tsallis, C. (1988). Possible generalization of boltzmann-gibbs statistics. *Journal of statistical physics*, *52*, 479–487.

Warne, R. T. (2014). A primer on multivariate analysis of variance (manova) for behavioral scientists. *Practical Assessment, Research & Evaluation*, *19*.

Yang, S., Hoque, S., & Deravi, F. (2019). Improved Time-Frequency Features and Electrode Placement for EEG-Based Biometric Person Recognition. *IEEE Access*, *7*, 49604–49613. doi:10.1109/ACCESS.2019.2910752.

Zhang, D., Jia, X., Ding, H., Ye, D., & Thakor, N. V. (2009). Application of tsallis entropy to eeg: quantifying the presence of burst suppression after asphyxial cardiac arrest in rats. *IEEE transactions on biomedical engineering*, *57*, 867–874.

Zhang, Y., Wei, S., Maria, C. D., & Liu, C. (2016). Using lempel–ziv complexity to assess ECG signal quality. *Journal of Medical and Biological Engineering*, *36*, 625–634. doi:10.1007/s40846-016-0165-5.