UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Facoltà di Ingegneria
Corso di Laurea Magistrale in Biomedical Engineering

# Design and Implementation of Techniques for the Secure Authentication of Users Based on Electroencephalogram (EEG) Signals

Candidate:
**Nibras Abo Alzahab**

Advisor:
**Prof. Marco BALDI**

Coadvisor:
**Prof. Lorenzo SCALISE**

Academic Year 2020-2021


UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Facoltà di Ingegneria
Corso di Laurea Magistrale in Biomedical Engineering

# Design and Implementation of Techniques for the Secure Authentication of Users Based on Electroencephalogram (EEG) Signals

Candidate:
**Nibras Abo Alzahab**

Advisor:
**Prof. Marco BALDI**

Coadvisor:
**Prof. Lorenzo SCALISE**

Academic Year 2020-2021

Università Politecnica delle Marche
Facoltà di Ingegneria
Corso di Laurea Magistrale in Biomedical Engineering

*Dedication*

*To the heart that holds me with warm wings... My Mum Lina*

*To the friend, support, and partner... My brother Mamoon*

*To happiness and tenderness... My grandmother Nahla*

*To the smile of satisfaction... My grandparents Mamoon & Hind*

*To the breeze of friendship and motherhood... Nidaa & Yamam*

*To the warmth of love and giving... Tasneem & Riadh*

*To my teacher whom I trust in his advice... Mouhannad Al-Koussa*

*To a friend of knowledge... Hassan AlImam*

*To the support, supervisor, and friend Bilal Alchalabi.*

*To my friends who are far in countries... Close to hearts... Hassan & Maram*

*To My brother, To the man of difficult situations... Mohammad Shaker*

*To whom we shared laughter and sadness... Riham*

*To those who shared our sweat and success... Ali & Shafik*

*To my cheerful spirit and special friend... Dana*

*To the family who I share success dreams... Eyad & Tala*

*To partner of happy and difficult days... Mouaz Alshalak*

*To the strong smile in the history, present, and future... Bilel Baazaoui*

*To the spirit of success... Lama*

*To the laughter of life... Maryam*

*To the support and kind friend Sameh*

*To the Italian horsemen... Angelo, Luca, & Alessandro*

*To the friend with the constant support... Ismaela Avelino*

*To every friend with whom I share dreams and ambitions... To every friend I have learned from... To every teacher who is a friend...*

# Abstract

Biometrics are getting more popular in the field of security systems and authentication. This is because biometrics are less able to be lost and less able to be stolen or spoofed. EEG-based biometrics are getting more attention recently since they are more resistant to be hacked. This thesis aims to design and implement techniques for the secure authentication of users based on electroencephalogram (EEG) signals. The study was conducted using three datasets: Simultaneous Task EEG workload Dataset, EEG Alpha wave Dataset, and Local Dataset.

The first problem addressed is the curse of dimensionality. Four reduced feature sets were used to reduce the dimensions of the systems, namely, cluster map, ANOVA F-Value, logistic regression weights; the cluster map method reached the highest performance with an 82.37% reduction in computation time.

The second problem is to reduce the time required to record EEG signals. Different scenarios with different EEG recording durations were tested. The results reveal a temporal threshold, equals to 4 seconds, that balances between performance and implementability.

The third problem is the effect of the auditory stimuli. To do so, six experiments were conducted, native, non-native, and neutral songs. The three songs were conducted using In-Ear and Bone-Conducting headphones. The results show that an increase in the performance of authentication equals 9.27% when using auditory stimuli. Additionally, it shows that using In-Ear or Bone-Conducting auditory stimuli is based on the balance between performance and implementability. Finally, the performance of EEG-based authentication is independent of the language of auditory stimuli.

In conclusion, this thesis contributes to the development of EEG-biometrics by bridging some important gaps in the field.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Security of information is a critical issue that was a subject of interest for human beings for ages. This Security of data involves giving access to the protected information only for authorised users. This authentication is based on the three methods, 1) something the person has, like a card or a key, 2) something the person knows like a password, and 3) something the person is, like unique physiological traits. The last one is known as biometric. One promising biometric is the electroencephalography biometric (EEG-based biometric).

EEG-based authentication is still a fresh application of brain-computer interface (BCI) systems. There are still many questions to be answered before the system can implement in real-life applications. Recent studies have discussed EEG-based authentication. However, there is a lot of gaps to be bridged.

The work aims to develop new methods to enhance the efficiency of the EEG-based authentication systems. This enhancement is from performance and implementability points of view.

The importance of this work comes from the fact that it reveals unanswered questions. Furthermore, it conducts experiments that have never been performed before, such as the effect of EEG-based authentication on bone-conducting auditory stimuli.

Three main experiments were conducted in this work. The first one was to select a feature set for the system. This feature set aims to create a system with high performance and low computational effort—the results of the feature set used for the second experiment. The second experiment aims to find a temporal threshold that balances between minimizing the duration of recorded EEG signals. Those results were used in the last experiment to discover the effect of auditory stimuli on the performance and implementability of EEG-based authentication.

The rest of the thesis is organized as follows: Chapter 2 presents background information and a Literature Review of the topic. It provides basic information about the physiology of the brain and nervous system, neuroimaging techniques, Brain-Computer Interface (BCI) systems, Security and Authentication, and biometrics. Moving on to Chapter 3 where it discusses the problem statement and clearly presents the objectives and hypotheses of the work. Afterwards, Chapter 4 represents the experimental work by describing the materials and methods by describing the

datasets, prepossessing, feature selection, recording duration, and feature auditory stimuli experiments. Chapter 5 presents the results of the experiments, while Chapter 6 discusses those results. Finally, Chapter 7 conclude the work by listing the main contribution of this thesis.

# Chapter 2

# Background Study and Literature Review

## 2.1 Physiological background

As a starting point to discuss the brain biometric, it is important to understand the anatomy and physiology of the brain and nervous system. Therefore, this section will go through a general introduction about the nervous system, the brain, and the cortex as a macro view. Afterwards, the micro view covers the anatomy of the neuron.

### 2.1.1 Nervous system

The nervous system is considered one of the highest complexity systems in the human body. This is because of the huge number of tasks that the nervous system should perform. One of its main tasks is regulating the organs' physiological function. Additionally, it is responsible for rationality, memory, language, and all other metals, conscious and unconscious, functions. To sum up, the nervous system functions could be categorized into four main functions:

1. Sensory
2. Integrator
3. Effector
4. Internal regulator

#### The anatomy of the nervous system

Generally, the nervous system is anatomically divided into two main parts: The central nervous system (CNS) and the Peripheral Nervous system (PNS). The CNS consists of three parts, the Brain, the brainstem, and the Spinal Cord, while PNS is divided into the Somatic nervous system and the Autonomic nervous system. This classification is illustrated in Fig 2.1.

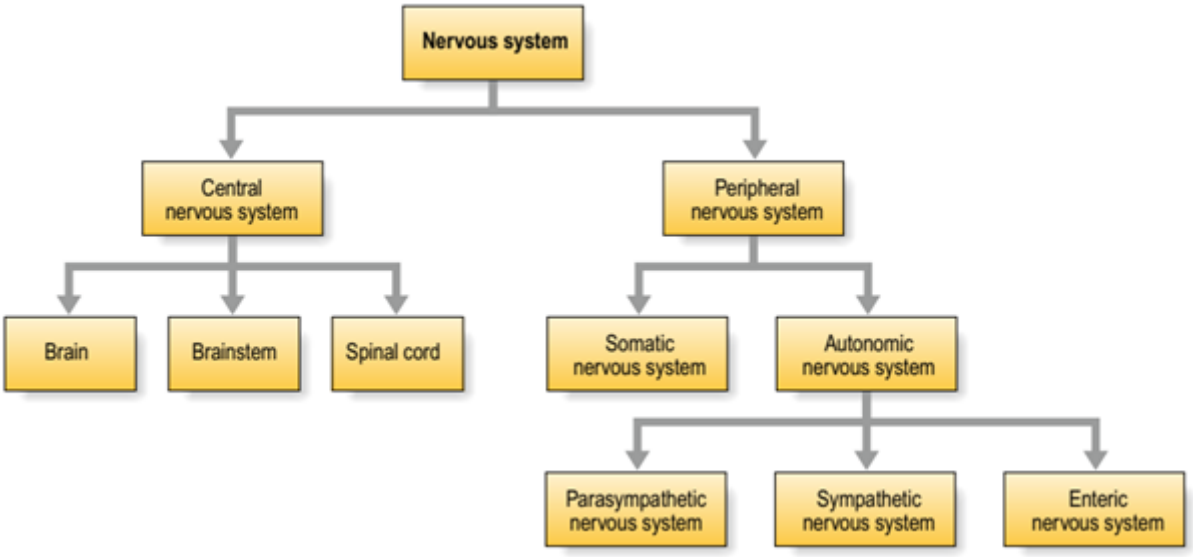


Figure 2.1: General structure of nervous system [Michael-Titus et al., 2010]

#### The forebrain

The forebrain constitutes 80% of the brain's volume. It comprises the cerebral cortex, the limbic system and the basal ganglia. The forebrain is responsible for high functions, such as cognition, memory, controlling the higher motor functions and emotions. There is a difference in architecture between the cerebral cortex and the limbic cortex. The limbic cortex consists of three to four layers; most are pyramidal cells, while six layers of neurons could be found in the cerebral cortex.
The cortex is divided into functional areas where each area has a single specific function, called primary areas. The function of the primary areas, illustrated in Fig 2.2, is to receive and perform the initial information processing. Those areas include:

1. Frontal lobe: responsible for movement control.
2. Parietal lobe: responsible for speech and sensation.
3. Temporal lobe: responsible for hearing and comprehension.
4. Occipital lobe: responsible for vision.
5. Limbic lobe: responsible for emotions, memory, motivation, and cognition.

### 2.1.2 Neurons

The nervous system consists mainly of two types of cells: neurons and glial cells. The neurons are responsible for the function of the nervous system, while the glial cells have the role of support, protection, and nutrition. The focus of this section will be on the neurons' anatomy and function.

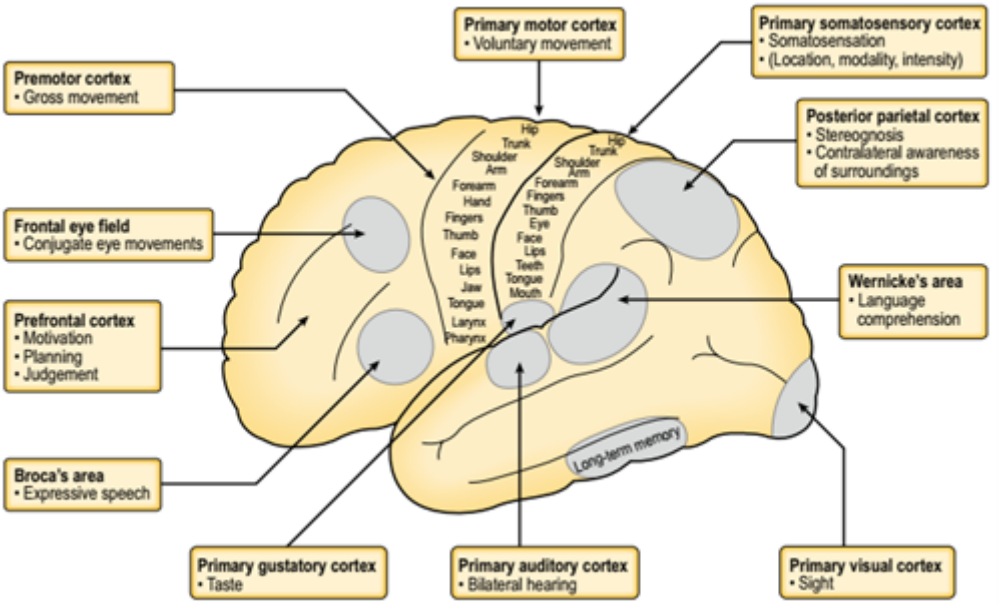


Figure 2.2: Forebrain and its primary functional areas [Michael-Titus et al., 2010]

### 2.1.3 The neuron anatomy

There are several shapes and sizes of the neurons, but they all share the same three elements:

1. Soma (cell body): It contains intracellular organelles such as the nucleus and ribosomes.

2. Dendrites: They are highly branched outgrowths in the cell. Their function is to receive the input from other cells and sensory units.

3. Axon: It is a single process that carries out the soma.

Fig 2.3 shows a typical illustration of the neurons.

#### Neuron excitability

The function of neurons could be surmised as receiving input signals, processing and firing output signals. The input signals are received from the external environment through physiological sensors. The receiving parts are the dendrites and the soma. The received signals could be excitatory (epsps: excitatory postsynaptic potentials) or inhibitory (ipsps: inhibitory postsynaptic potentials) signals. Those signals are summed together through temporal and spatial summation to produce the output signal.
Neurons signals are in a shape of a fixed size electrical impulse called the action potential. Which mean it is an all-or-none phenomenon. The potential starts with a rapid depolarization from -75 mV (resting membrane potentials) to +40 mV, followed

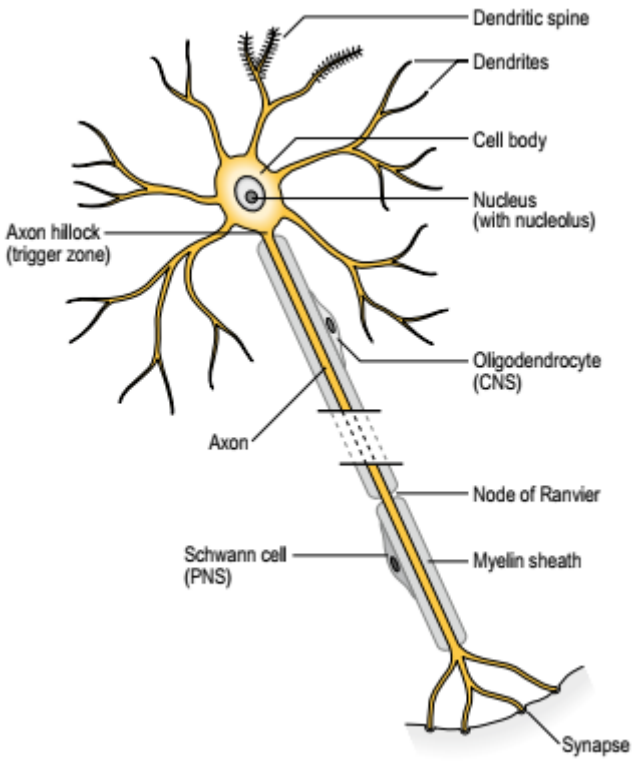


Figure 2.3: Anatomy of Neuron [Michael-Titus et al., 2010]

by repolarization to -95 mV—afterwards, the potential backs to the resting membrane potentials. Figure 2.4 represent a typical action potential signal.

## 2.2 Neuroimaging

There are several methods for signal acquisition from neurons which is known as neuroimaging. Those methods are classified into invasive and noninvasive methods. Each method has its own advantages and disadvantages in terms of spatial resolution, temporal resolution, cost, and complexity. Figure 2.5 demonstrate neuroimaging methods. [Nam et al., 2018]

Invasive methods are those where electrodes make direct contact with the brain. Despite its high accuracy, placing invasive electrodes requires a surgical operation [Nam et al., 2018]. This is impractical for brain-biometric applications; therefore, it will not be discussed further.

Noninvasive methods are those where electrodes do not make direct contact with the brain. The electrodes are placed on the skin or surrounding the head. We can distinguish between two types of noninvasive neuroimaging methods, indirect and direct. The indirect methods measure a reflection of the brain activity, such as brain metabolism or hemodynamic of the brain. Those methods include Functional Magnetic Resonance Imaging (fMRI), Functional Near-Infrared Spectroscopy (fNIRS),

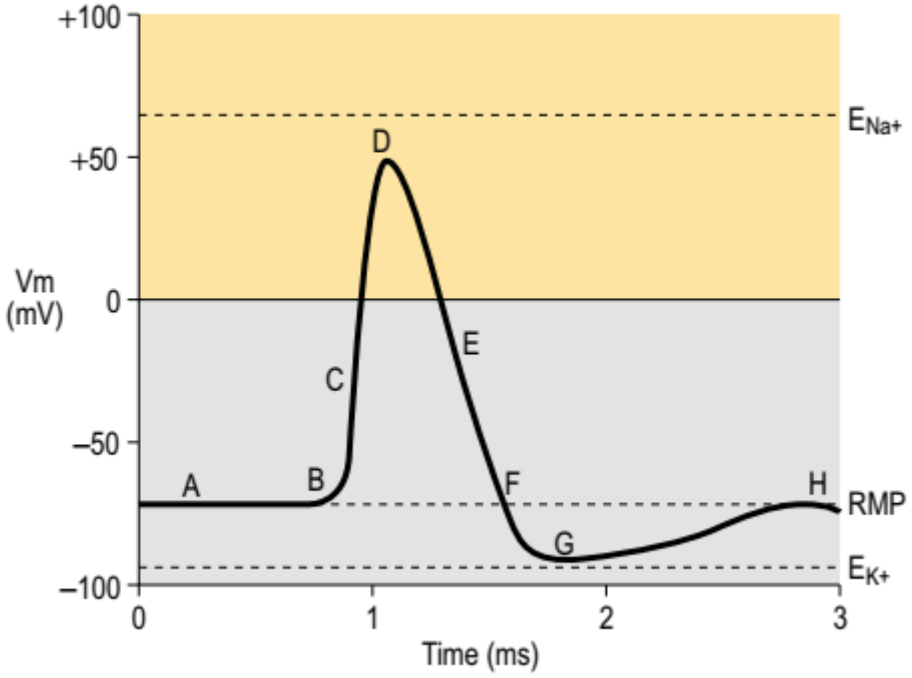


Figure 2.4: Action Potential waveform [Michael-Titus et al., 2010]

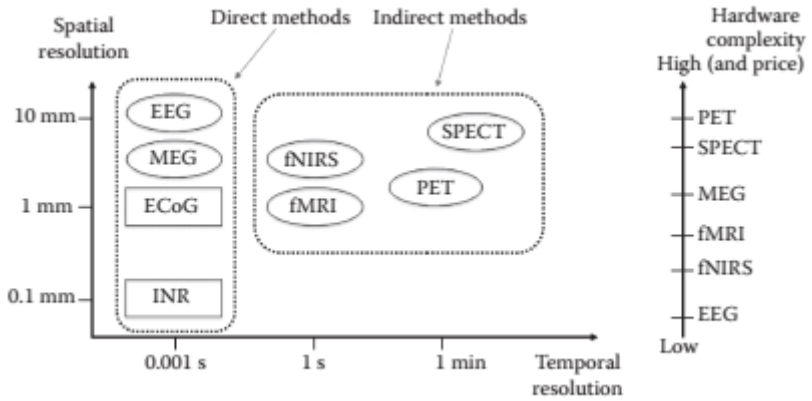


Figure 2.5: Comparision between neuroimaging methods [Nam et al., 2018]

Positron Emission Tomography (PET). On the other hand, the direct methods record directly the magnetic, such as Magnetoencephalography (MEG), or electric, such as Electroencephalography (EEG), activity [Bhattacharyya et al., 2017]. Those methods will be briefly explained with a focus on the EEG.

### 2.2.1 fMRI

fMRI is an indirect method to identify neural activity by measuring the blood-oxygen-level-dependent (BOLD). The principle of measuring this hemodynamic response is the coupling between the regional cerebral blood flow (rCBF) and regional cerebral blood oxygenation (rCBO). In other words, n decrease in oxygenated blood level is detected in the surrounding blood of firing neurons. This is due to the increase in the metabolism level of those neurons [Lindauer et al., 2010]. To evaluate the quality of neuroimaging method, two types of parameters are identified, spatial accuracy and temporal accuracy. The spatial accuracy of fMRI is equal to 1 millimetre while the temporal one equals 1 second. In a general context, fMRI machines are bulky and require a specific isolated room for data acquisition.

### 2.2.2 fNIRS

fNIRS is a non-invasive method to detect brain activity. It depends on the difference in light absorption between oxygenated and deoxygenated blood. As an example, light below 800 nm is likely to be absorbed by deoxygenated blood, while oxygenated blood absorbs more light in the range higher than 800 nm [Giardini et al., 2000, Wilcox et al., 2008]. Since fNIRS systems are vulnerable to different noise resources, there must be artefacts removal systems like independent component analysis (ICA) or wavelet. Generally, fNIRS is portable and inexpensive technology; therefore, it can use outside the lab environment [Castermans et al., 2014].

### 2.2.3 PET

PET is an indirect method that is based on nuclear medicine to record three-dimensional data about the human body activity [Stollfuss et al., 2015]. In simple words, PET systems rely on a positron-emitting tracer, such as fludeoxyglucose, that acts similar to the glucose in the blood. By recording the positron-emission activity, the neural activity is indirectly recorded. Despite its good spatial accuracy, it is considered a high-cost option that requires injection with radioactive materials [Townsend, 2008].

### 2.2.4 MEG

MEG is a direct method to record brain activity. It requires a superconducting quantum interface device (SQUID) as a magnetic field sensor. MEG provides great advantages such as high temporal resolution, high temporal resolution, and direct

measure of brain activity. On the other hand, recording MEG requires expensive tools in a magnetically shielded room [Uhlhaas, 2015].

### 2.2.5 EEG

EEG is a direct method to noninvasive record the activity of the brain. This is done by placing electrodes on the surface of the scalp [Da Silv, 2005]. This method measures the electrical activity of the brain directly. In particular, EEG measures the activity of the pyramidal neurons of the cortex [Cantor and Evans, 2013]. EEG is one of the best neuroimaging techniques in terms of temporal resolution. However, it is difficult from a mathematical perspective to detect the location of brain activity. Moreover, it is hard to determine the activity distribution that generates a specific signal. This is called the "inverse problem" [Castano-Candamil et al., 2015]. Due to the low cost and portability of EEG acquisition tools, it was used in 68% of Brain-Computer Interface (BCI) applications. Therefore, it will be the first candidate to be discussed in Brain-activity-based authentication systems [Nam et al., 2018].

#### EEG Generation

EEG signal is an electrical recording of the activity of the neurons. In other words, it measures the electrical current resulted from the synaptic excitation of the pyramidal neurons in the cerebral cortex. This difference in electrical potential generates a dipole between the neuron's soma and the dendrites.
This electrical signal cannot be recorded noninvasively without attenuation. There are three main layers between the electrodes and the dipole. The three layers are the scalp, skull, and brain, in addition to the number of thin layers. This attenuation is the main source of the internal noise combined with the EEG signal.

#### Electrode Placement

Electrode Placement is an essential aspect for recording EEG signals since the location of the electrode is correlated with cortical areas. The most common protocol for EEG-electrodes placement is the international 10/20 system. In this system, 21 electrodes are placed according to six locations with interval rates of 10%, 20%, 20%, 20%, 20%, 10% of the distance from the nasion to the inion. The six locations are: Fp: Frontopolar, F: Frontal, C: Central, P: Parietal, and O: Occipital, see Fig 2.6. The 10/20 system was expanded to more dense systems like the 10/10 system and 10/5. To accurately locate the correct position of electrodes, four points must be identified: 1) the nasion, the point between the nose and the forehead, 2) the inion, in the back of the head where the lowest point of the skull exists. 3,4) the pre-auricular points, the points anterior to the left and right ears [Homan et al., 1987, Jasper, 1958].

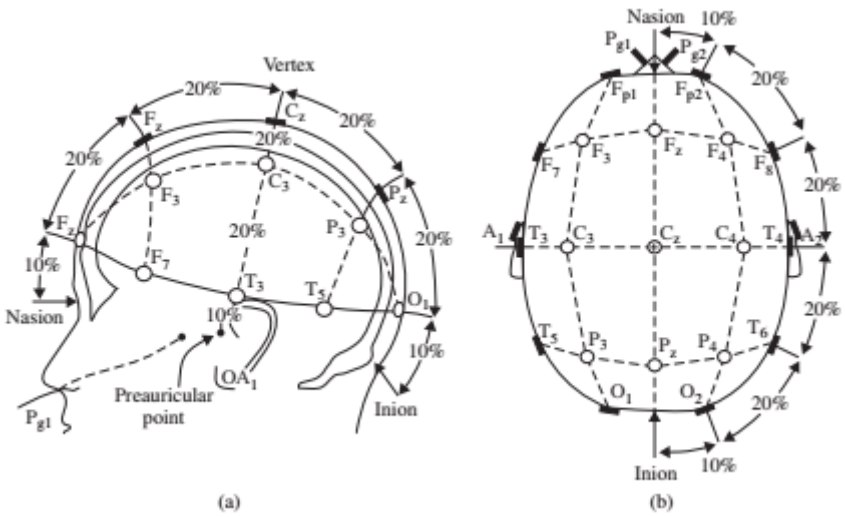


Figure 2.6: International 10-20 EEG electrodes placement system [Sanei and Chambers, 2007].

#### Brain Rhythms (Brainwaves)

EEG signals is a complex signal since it contains a number of sub-signals, which are called Brainwaves. Brainwaves are divided into five major signals according to their frequency range. Those signals are listed in table 2.1. Each signal dominates in specific mental states and has specific characteristics.

## 2.3 Brain-Computer Interface (BCI)

Brain-Computer Interface (BCI) is an artificial communication system that enables the brain to communicate with the world electronic system. BCI system bypasses the human communication systems, which includes verbal communication and sign language. This opens the gates for advanced application in the human-machine interfaces such as therapeutic, affective, artistic, entertainment, and security applications. The main principle of BCI systems is translating thought and intention into a command that communicates with the real world [Alzahab et al., 2021, Lotte et al., 2007,

Table 2.1: Categories of Brain Rhythms patterns [Nam et al., 2018, Sanei and Chambers, 2007].

| Name | Frequency Range (Hz) | Amplitude ($\mu$ V) |
|---|---|---|
| Delta ($\delta$) | 0.5 - 4 | 100 - 200 |
| Theta ($\theta$) | 4 - 8 | 5 - 10 |
| Alpha ($\alpha$) | 8 - 13 | 20 - 2080 |
| Beta ($\beta$) | 13 - 30 | 1 - 5 |
| Gamma ($\gamma$) | 30 - 60 | 0.5 - 2 |

Alchalabi and Faubert, 2019.

BCI systems could be classified according to different categories. Firstly, BCI systems according to the brain signal pattern are categorized into **Steady State Visually Evoked Potential (SSVEP), Steady-State Auditory Evoked Potential (SSAEP), Steady-State Somatosensory Evoked Potential (SSSEP)** are the brain activity in response to a certain type of stimulation, (P300) where the positive voltage peak is reached after 300 ms after the stimuli, **event-related desynchronization/synchronization (ERD/ERS)**, which is the activity of the brain when imagining a movement, and **slow cortical potentials (SCPs)** that is a shift from milliseconds to seconds in the cortical activity. Another classification is based on the stimulus modality. In this case, we have two categories, internal and external stimulus. The internal one is a self-regulated stimulus, which operates in cognitive efforts strategy. While the external could be visual, auditory, or tactile, which operates in a selective attention strategy. Another possible classification is according to the used neuroimaging method. Fig 2.7 demonstrate the classification of BCI systems.

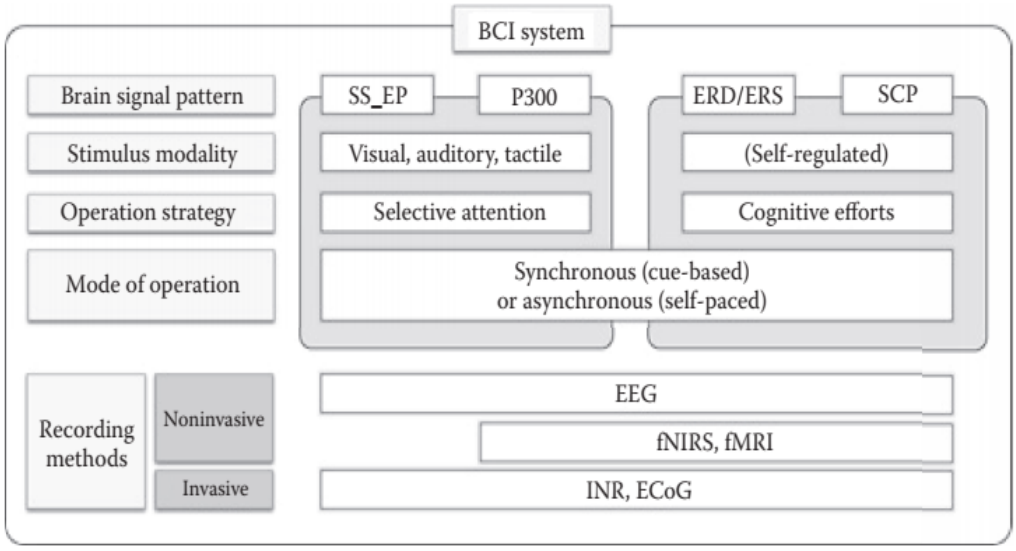


Figure 2.7: Classification of BCI Systems

### 2.3.1 BCI framework

BCI systems use a general framework, which is illustrated in Fig 2.8, consisting of many stages. Firstly, brain activity recording using a neuroimaging method. Secondly, pre-processing of the recorded signal. The importance of this step is to ensure the quality of the signal by removing the artefacts. Afterwards, extracting the signal's features which represent the useful information of brain activity. Finally, translating the extracted feature into commands using a classification algorithm. To

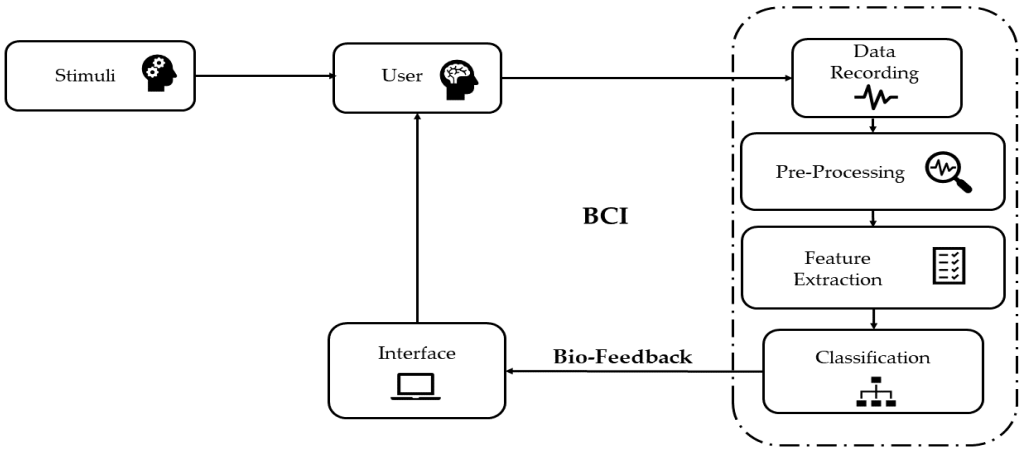


Figure 2.8: Brain-Computer Interface framework [Alzahab et al., 2019]

close the loop, feedback connects the resultant command with the user.

### 2.3.2 Pre-processing

Input brain activity is subject to many sources of artefacts and noise. This noise is generated from many sources, exogenous or endogenous noises. The internal noise is generated from other physiological signals like electrocardiogram (ECG), electrooculogram (EOG), and the effect of an eye blink. In comparison, the source of the external one is from the background power line network. Those sources add information to the signal that is not related to neural activity. Therefore it is important to eliminate the non-informative data from the signal.
To reduce the effect of noise and artefacts, many denoising techniques are implemented. Those techniques differ according to the degree of denoising and computational complexity. An example of simple denoising techniques is simple filters such as finite impulse response (FIR) and infinite impulse response (IIR) filters. These filters reduce the frequency range of the signal since neural information lies below 30 Hz in most BCI applications). More advanced techniques include Blind Source Separation (BSS) methods, semi-BSS, and Independent Component Analysis (ICA). More details about the algorithms implemented in this project will be demonstrated later.

### 2.3.3 Feature Extraction

In order to implement the BCI system in an optimal configuration, it is important to choose the features that represent the neural activity of a specific task. Therefore, the feature selection phase is essential. This project considers 17 features to represent the neural activity responsible for authentication systems extracted from EEG signals. The considered features are:

### Statistical Features

: Those features provide descriptive statistics features, including centre tendency statistics, dispersion statistics, and shape statistics.

- **Amplitude Mean**:
  The mean of the EEG signal is given in the following equation:

$$\mu = \sum_{n=1}^{N} \frac{x(n)}{N} \tag{2.1}$$

  $\mu$ is the amplitude mean, $x(n)$ is the EEG signal, and $N$ is the number of the samples in the signal.

- **Amplitude Standard Deviation**:
  The standard deviation of the EEG signal is given in the following equation:

$$\sigma = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - \mu)^2}{N - 1}} \tag{2.2}$$

  $\sigma$ is the amplitude Standard Deviation, $x(n)$ is the EEG signal, and $N$ the number of the samples in the signal.

- **Amplitude Variance**:
  The Variance of the EEG signal is given in the following equation:

$$S^2 = \frac{\sum_{n=1}^{N}(x(n) - \mu)^2}{N - 1} \tag{2.3}$$

  Where $S^2$ is the amplitude Variance, $x(n)$ is the EEG signal, and $N$ the number of the samples in the signal.

- **Amplitude Range**:
  The range of the EEG signal is given in the following equation:

$$Range = Max(x(n)) - Min(x(n)) \tag{2.4}$$

  Where $x(n)$ is the EEG signal

- **Amplitude Skewness**:
  This feature measures the non-Gaussianity of EEG signals. Skewness is a high-order moment to measure the lack of symmetry of the distribution. The Skewness of the EEG signal is given in the following equation:

$$Skewness = \frac{E[(x(n) - \mu^3]}{\sigma^3} \tag{2.5}$$

  where $E$ is the statistical expectation, $x(n)$ is the EEG signal, $\mu$ is the amplitude mean, and $\sigma$ is the amplitude standard deviation [Sanei and Chambers, 2007].

- **Amplitude Kurtosis**: This feature measures the flatness of the EEG signal. Kurtosis is a high-order moment to measure whether the data are flat or peaked relative to the normal distribution. The Kurtosis of the EEG signal is given in the following equation:

$$Kurtosis = \frac{m_4[x(n)]}{m_2^2[x(n)]} \tag{2.6}$$

where $m_i[x(n)]$ is the central moment of the EEG signal with the *ith* order. For example, $m_i[x(n)] = E[(x(n) - \mu)^i]$.
For a signal with normal distribution, the kurtosis equels to 3. Therefore, the normalized kurtosis, also called the excess kurtosis is given in the following equation:

$$ExKurtosis = \frac{m_4[x(n)]}{m_2^2[x(n)]} - 3 \tag{2.7}$$

In this case, the value 0 represent the normal distribution while positive or negative values represent the degree of flatness [Sanei and Chambers, 2007].

### Power Spectral Density (PSD) Features

The Power Spectral Density is a widely known feature for analysing signals. It represents the distribution of the signal's power as a function of frequency. PSD is computed as follows according to the Welch method [Rahi and Mehra, 2014, Lotte, 2008]:

1. Segmenting the EEG time series into successive, possibly overlapping, segments

2. Multiplying each segment by a Hamming window to reduce the variance [Podder et al., 2014].

$$w(n) = 0.54 + 0.46.cos(\frac{2\pi n}{N}), -\frac{N}{2} \leq n \leq +\frac{N}{2} \tag{2.8}$$

where $w(n)$ is the hamming window, and $N$ is the number of samples.

3. Using Fast Fourier Transformation (FFT) to compute the periodogram.

4. Squaring the periodogram to compute the PSD

$$P(f) = \frac{\Delta t}{N} |\sum_{n=0}^{N-1} h_n x(n) e^{-j2\pi f_s n}|^2 \tag{2.9}$$

Where

5. Reassigning each PSD segment to the centre of energy of its bin. This Provide exact localization for the impulses [Fulop and Fitz, 2006].

6. Calculating the average of the reassigned squared PSD segments to estimate

the PSD.

$$PSD = \frac{1}{K} \sum_{K=0}^{K-1} P(f) \tag{2.10}$$

Where $K$ is the number of the segments.

From the calculated PSD, the mean and standard deviation were considered as frequency-domain features:

- **PSD Mean**
- **PSD Standard Deviation**

#### Entropy Features

The entropy concept is a wide-separated tool to indicate the complexity of nonlinear time series. What makes this concept favourable is its relevance to a wide range of problems and its computational efficiency. Since EEG signal is a non-stationary, multidimensional, nonlinear signal, the study of its fluctuations reveals some of its hidden information [Plastino and Rosso, 2005]. Those fluctuations resulted from the global activity of the brain. Which mean the fluctuation analysis differs according to the architectures of the neurons layers in the cortex. In other words, the fluctuation analysis differs from one brain to another. This refers to consider entropy features as a promising candidate for EEG-based authentication. There are diffident

- **Permutation Entropy**
  Bandt and Pompe introduced the concept of permutation entropy in 2002 [Bandt and Pompe, 2002] to measure the complexity of physiological signals. The conceptual simplicity, robustness to noise and artefacts, and the low computational complexity of permutation entropy make it a desirable feature to be used for EEG signals analysis. [Morabito et al., 2012] Used permutation entropy studying Alzheimer's disease. However, it was never used before for EEG-based biometrics in the literature.
  The basics of permutation entropy are to calculate the frequency of appearance of ordinal patterns, also called motifs and denoted as $\pi_j$, see Fig 2.9. To compute permutation entropy, We define $d$ is the number of samples of an EEG segment, segmented from the EEG time series with $N$ sample, and $\tau$ as time-lag, which is the number of samples that a motif covers. The count of each motif in the EEG segment, known as the frequency of appearance of the motif j $f(\pi_j)$, is used to compute the relative frequency $P(\pi_j)$:

$$P(\pi_j) = \frac{f(\pi_j)}{N - d + 1} \tag{2.11}$$

  Afterwards, by fixing the embedding dimension $d > 2$, and the time-lag $\tilde{\tau}$, the

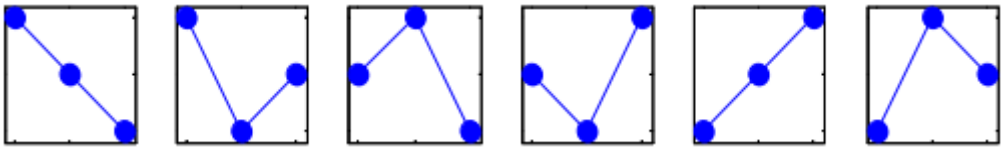

Figure 2.9: The possible types of motifs when $d = 3$ [Morabito et al., 2012]

permutation entropy. $H(d, \tilde{\tau})$ is calculating using the following equation:

$$H(d, \tilde{\tau}) = -\sum_{\pi_j=1}^{d!} p(\pi_j)\ log_2\ p(\pi_j) \tag{2.12}$$

The output of permutation entropy ranges between zero, a very regular time series with a repetition of the same motif, and $log_2(d!)$, where the probability of all motifs is equal. A more convenient representation of permutation entropy is introduced by normalizing $H(d, \tilde{\tau})$ by the highest value $log_2(d!)$:

$$0 \leq \frac{H(d, \tilde{\tau})}{log_2\ (d!)} \leq 1 \tag{2.13}$$

- **Spectral Entropy**
  Generally, entropy features to extract information representing the complexity of a signal. The spectral entropy takes advantage of the frequency domain. The spectral entropy showed an outstanding performance in in-person identification systems [Phung et al., 2014b] when extracted from alpha, beta, and gamma brainwaves. Additionally to the high performance, it showed a low feature dimension, which in role increases the classification speed by reducing the computational complexity.
  To compute the spectral entropy, the PSD is first computed using FFT. From the resultant PSD, each $P_f$, the power of every frequency component, was normalized by dividing it by the total power $\sum P_f$ [Mu et al., 2017, Kannathal et al., 2005].
  $$p_f = \frac{P_f}{\sum P_f} \tag{2.14}$$
  Thus, the spectral entropy could be calculated using the folwlong equation:
  $$SpectralEntropy = \sum_f p_f\ \log\ (\frac{1}{p_f}) \tag{2.15}$$

- **Singular value decomposition (SVD) entropy**
  Singular value decomposition (SVD) entropy measures the richness of the information in a given time series. Its advantage is the ability to use concise and non-stationary signals [Jelinek et al., 2019]. SVD entropy was used

[Krishnan et al., 2020] to extract information from EEG signals for schizophrenia detection.
SVD is a mathematical method in linear algebra to decompose a matrix $X_{m\times n}$ into its constituent eigenvectors. Since the embedded dimension is set to 3. SVD decomposes $X_{m\times n}$ into three matrices [Jelinek et al., 2019].

$$X_{m\times n} = U_{m\times m} \times \Sigma_{m\times n} \times V_{n\times n} \tag{2.16}$$

Where $U$ and $V$ are unitary matrices and $\Sigma$ is a diagonal matrix. The matrix $\Sigma$ contains the singular values $\Sigma = diag[\lambda_1, ..., \lambda_I]$ which will be used for calculating the SVD entropy using the following equation [Krishnan et al., 2020]:

$$SVDEntropy = -\sum_{i=1}^{I} \frac{\lambda_i}{\sum \lambda} ln\ \frac{\lambda_i}{\sum \lambda} \tag{2.17}$$

Where $\sum \lambda$ is the sum of the singular values in the diagonal of the $\Sigma$ matrix.

- **Approximate entropy**
  Pincus, in [Pincus, 1991], overcame the limitations of Kolmogorov–Sinai entropy by proposing the approximate entropy. Approximate entropy was originally developed to deal with medical data and medical time series to determine how the complexity of the systems is changing for both deterministic stochastic and chaotic processes. Its main advantage is the ability to extract useful information from small amounts of data and its immunity to noise and artefacts for a certain level. EEG analysis was used to determine the effect of epilepsy on EEG-based biometrics [Phung et al., 2014a]. Additionally, the identification of epilepsy seizures was studied using approximate entropy [Sharanreddy and Kulkarni, 2013, Acharya et al., 2009].
  Computing the approximate entropy of an EEG signal with $N$ samples is done by [Delgado-Bonal and Marshak, 2019, Mu et al., 2017, Richman and Moorman, 2000]:

  1. Denoting EEG signal as $x(i)$.
  2. Identifying the length of compared run of data, denoted $m$ where $0 \leq m \leq N$, and a a filtering level $r$, positive real number.
  3. Identifying a set of m-dimensional vectors:

$$X_i^m = [x(i), x(i+1), ..., x(i+m-1)]; 1 \leq i \leq N-m+1 \tag{2.18}$$

  4. Calculating the maximum distance between two vectors

$$d[X_i^m.X_j^m] = \max_{k\epsilon(0,m-1)} |x(i+k) - x(j+k)|;\ i,j = 1 \sim N-m+1, i \neq j \tag{2.19}$$

  5. Calculating the number of blocks, $X_i^m$, of consecutive values which are

similar to a given block with a resolution equal to $r$, denoted $C_i^m(r)$:

$$C_i^m(r) = \frac{1}{N-m+1} C_i \tag{2.20}$$

6. Defining a function $\phi^m(r)$

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} ln\ C_i^m(r) \tag{2.21}$$

7. Similarly, defining a function $\phi^{m+1}(r)$

$$\phi^{m+1}(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln\ C_i^{m+1}(r) \tag{2.22}$$

8. Calculating the approximate entropy

$$ApproximateEntropy = \phi^m(r) - \phi^{m+1}(r) \tag{2.23}$$

- **Sample entropy**
  Richman and Moorman [Richman and Moorman, 2000] proposed an optimised version of approximate entropy called sample entropy. The proposed method has more suitability for physiological signals since it is independent of the recording duration [Mu et al., 2017]. When sample entropy was extracted from an EEG signal in a biometric authentication system, it resulted in a high recognition rate [Thomas and Vinod, 2016].
  Computing the sample entropy is similar to computing the approximate entropy [Delgado-Bonal and Marshak, 2019, Mu et al., 2017, Richman and Moorman, 2000]:

  1. Denoting EEG signal as $x(i)$.
  2. Identifying the length of compared run of data, denoted $m$ where $0 \leq m \leq N$, and a a filtering level $r$, positive real number.
  3. Identifying a set of m-dimensional vectors:

$$X_i^m = [x(i), x(i+1), ..., x(i+m-1)]; 1 \leq i \leq N-m+1 \tag{2.24}$$

  4. Calculating the maximum distance between two vectors

$$d[X_i^m . X_j^m] = \max_{k\epsilon(0,m-1)} |x(i+k) - x(j+k)|;\ i,j = 1 \sim N-m+1, i \neq j \tag{2.25}$$

  5. Calculating the number of blocks, $X_i^m$, of consecutive values which are

similar to a given block with a resolution equal to $r$, denoted $B_i^m(r)$:

$$B_i^m(r) = \frac{1}{N-m+1} B_i \tag{2.26}$$

6. Defining a function $\phi^m(r)$

$$\phi^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \tag{2.27}$$

7. Similarly, defining a function $\phi^{m+1}(r)$

$$\phi^{m+1}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^{m+1}(r) \tag{2.28}$$

8. Calculating the sample entropy

$$SampleEntropy = \log \frac{\phi^m(r)}{\phi^{m+1}(r)} \tag{2.29}$$

### Fractional Dimension (FD) Features

Fractional dimension (FD) is a mathematical representation of complexity through statistical methods. There are many methods to compute the FD; however, three main methods are used with EEG analysis [Petrosian, 1995, Esteller et al., 1999]:

- **Petrosian's fractal dimension**
  Petrosian's method for FD estimation is considered a fast method using the following equation [Petrosian, 1995, Esteller et al., 1999]:

$$FD = \frac{\log_{10}(N)}{\log_{10}(N) + \log_{10}(\frac{N}{N+0.4N_\Delta})} \tag{2.30}$$

  Where $N$ is the number of samples in the time series, and $N_\Delta$ is the number of sign changes

- **Katz's fractal dimension**
  Calculating Katz's FD is slightly slow in comparison with Petrosian's FD. It is calculated using the following equation [Katz, 1988, Esteller et al., 1999]:

$$FD = \frac{\log_{10}(N)}{\log_{10}(\frac{d}{L}) + \log_{10}(N)} \tag{2.31}$$

  Where $L$ is the sum of the distances between successive points and $d$ is maximum of the vector containing distances between the first point and all other points $d = \max[distances(1, i); i = 2, 3, ..., N]$

- **Higuchi's fractal dimension**
  Higuchi's FD is calculated as follows: [Higuchi, 1988, Esteller et al., 1999]:

  1. Considering $x(N) = x(1), X(2), ..., x(N)$ as the EEG time series with $N$ sample, we consider $k$ time series:

  $$x_m^k = \{x(m), x(m+k), x(m+2K), ..., x(m + \lfloor \frac{N-m}{k} \rfloor k)\} \tag{2.32}$$

  for $m = 1, 2, ..., k$ where k is the time interval between points, m is the initial time value, and $\lfloor b \rfloor$ is the integer part of the number $b$.

  2. The length $L_m(k)$ for every k time series $x_m^k$ is given bt:

  $$L_m(k) = \frac{\sum_{i=1}^{\frac{N-m}{k}} |x(m+ik) - x(m+(i-1)k|\ (N-1)}{\lfloor \frac{N-m}{k} \rfloor k} \tag{2.33}$$

  Where $\frac{(N-1)}{\lfloor \frac{N-m}{k} \rfloor k}$ is a normalization factor.

  3. Computing the average of lengths $L_m(k)$ for $m = 1, 2, ..., k$.
  4. Repeating the previous steps for each $k = 1, 2, ..., K_max$ to obtain the mean length for each $k$.
  5. Computing the least-squares linear best fit, using linear regression, for the curve $\ln(L(k))$ versus $\ln(\frac{1}{k})$.
  6. Considering the slope of the least-squares linear best fit as an estimation of the fractal dimension

### Detrended Fluctuation Analysis (DFA)

Detrended Fluctuation Analysis (DFA) was introduced tepeng1995quantification to measure the self-affinity of a signal statistically. Its main advantage is the suitability for non-stationary time-series. It has been shown that the DFA provides robust performance when was extracted from EEG signals [Maity et al., 2015]. For Computing DFA [Sanyal et al., 2015, Hardstone et al., 2012]:

1. Considering $x(N) = x(1), X(2), ..., x(N)$ as the EEG time series with $N$ sample
2. Integrating the EEG single into new series $y = [y(1), ..., y(N)]$ where

$$y_k = \sum_{i=1}^{k} (x_i - \bar{x}) \tag{2.34}$$

   where $\bar{x}$ is the amplitude mean of the EEG signal.

3. Segmenting The integrated series $y_k$ into equal-length segments, with length $n$.
4. Computing the least-squares linear best fit for each segment.

5. Calculating the root-mean-square fluctuation o of the series using the follo=wing equation:

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k) - y_n(k)]} \tag{2.35}$$

where $[y(k) - y_n(k)]$ is named detrending.

6. Finding the slope of the plot between $\log[F(n)]$ versus $\log(n)$, denoted $\alpha$.

7. $\alpha$ represents a self-similarity parameter which reflects the auto-correlation properties of the EEG signal.

- **DFA feature**

### 2.3.4 Classification

Classification is the central part of any BCI system. This is the step where extracted features are translated into a decision. There are diff rent types of classification algorithms that analyse the input statistically to find the best reality [Lotte et al., 2007]. In the case of biometric systems, the output of the classification part is the identity of the input subject. In this work, three types of classification algorithms will be used; 1) Multilayer Perceptron, 2) K Nearest neighbours, and 3) eXtreme Gradient Boosting.

#### Multi layer Perceptron

The most used Artificial Neural Networks type are those defined by [Rumelhart et al., 1986], which are called Multilayer Perceptron (MLP), MLP architecture consists of consecutive layers of neurons. Each layer contains a number of neurons, which are the core structure of the MLP. Neurons of each layer receive their input from the previous layer and feed their output into the next layer. The first layer is called the input layer, which receives the information of the extracted features. On the other hand, the last layer is the layer of decision making. The inner layers, known as the hidden layers, are responsible for analysing the input statistically, which is known as learning. The most common method of learning is the backpropagation method [Rumelhart et al., 1995].

#### K Nearest neighbors

This technique is based on distance measuring. It measures the distance between the input value and the data points in the data set. Afterwards, it assigns the class of the input value according to the class of the nearest K data points. There are diff rent methods to measure the distance [Alzahab et al., 2019, Lotte et al., 2007]. The most common one is the Mahalanobis distance. The Mahalanobis distance is given

in the following equating:

$$d_c(x) = \sqrt{(x - \mu_c)M_c^{-1}(x - \mu_c^T} \tag{2.36}$$

#### eXtreme Gradient Boosting

Thanks to parallel tree boosting, XGboost can solve machine learning problems in an accurate and fast way. The efficiency of XGboost comes from the fact that it integrates different methods, which is called ensembled learning. This method is so popular that it was implemented in many wining challenges on the Kaggle website. It started to be widely recognized in 2015 [Tiwari and Chaturvedi, 2019, Chen and Guestrin, 2016].

## 2.4 Security and Authentication

The main aim of authentication systems is to build an association between the person and his identity. Once the individual is authenticated, they are given access to his secret. To do so, three main methods are being used over the history of authentication. The first one is by possessing a specific key, which could be physical or digital. Whoever possesses the key is authenticated as a person who has access to the secret. This method requires a great effort to keep the key both available and hidden. The second method is by knowledge of a passcode. How has the knowledge of the pass-code had access to the secret? This requires a commitment to remember and secure the code. The final method is the intrinsic traits. The identity is determined by the inherent behavioural or physical traits, which is known as biometric. The biometric method is considered the most secure method because the biometric can not be lost or forgotten. Moreover, it is hard to be mimicked or spoofed[Jain et al., 2011].

## 2.5 Biometrics

### 2.5.1 Biometric Systems

The system of biometric security measures some physical or behavioural traits that represent the individual's identity. This trait could be a fingerprint, DNA, iris, signature, gait patterns, or face. The implementation of biometrics involves two main steps; *enrollment* and *recognition*. The enrollment stage is where the system is being built and calibrated, where the recognition stage is when the system is given the power to authenticate users. A Biometric system consists of four main blocks: sensor, feature extractor, database, and matcher. Those blocks are completely similar to what was discussed in the section: 2.3 [Jain et al., 2011]

### 2.5.2 Authentication and Identification

Generally, biometric-based security systems could be classified into two types, namely, authentication and identification. Firstly, authentication, which is also known as verification, tries to verify whether the person is really who he is calming. In other words, the system tries to answer, "Are you who you say you are?". Authentication system could be abstracted into a binary classification problem where the system tries to classify the input print into genuine and impostor. When the output is genuine, the system gives access. While when the output is impostor, the access is denied. The second type is identification, where the system tries to reveal the identity of the user. This is to say that the system tries to answer the question: "Are you someone who is known to the system?". The identification problem could be presented as a multi-class classification problem where the system tries to detect the identity of the input [Jain et al., 2011].

### 2.5.3 Performance measures

To optimize biometric systems, it is always desired to minimize the intra-user variations and the inter-user similarity. To be able to optimize the system effectively, some performance measures are desired to be assessed. The most common performance metrics for biometric systems are False Rejection Rate (FRR) and False Acceptance Rate (FAR). FRR refers to when the genuine is considered an impostor falsely; therefore, they cannot access his data. In contrast, the FAR represent when an imposter could fraud the system and considered genuine. When designing the biometric system, it is important to find a threshold where both FRR and FAR are acceptable since the two metrics work oppositely [Fairhurst, 2018, Gamassi et al., 2005, Sebastien et al., 2014].

# Chapter 3

# Problem statement, objectives and hypotheses

## 3.1 Problem statement

EEG-Biometric and Brain-based security systems still in the early stages. There still a lack of implementability and performance that limits real-life application. The problems to be challenged, according to [Del Pozo-Banos et al., 2014], [Gui et al., 2019], [Bidgoly et al., 2020], could be summarized by 9 points:

1. **Universality**: This means the ability to be used by almost all people. However, the work considered in the literature discussed the healthy subjects only. There is a lack of research that studied the differences between kinds of people, such as males and females, healthy and non-healthy, alcoholic and non-alcoholic, mother tongue, etc.

2. **Permanency**: This point discusses the stability of the signal over a period of time. The main concern here is to know whether an EEG-biometric system could be used over a period of time. From what was found in the literature, only a few papers recorded the data during the different sessions. The duration between sessions is limited to a couple of months.

3. **Uniqueness**: It is important for any biometric system is to identify an individual uniquely. This requires building systems that can distinguish between a vast number of individuals. Current studies develop their systems for a limited number of subjects, usual between 4  20. Only a small number of research reached a higher number, around 100. For better implementability, more subjects should be included in EEG-based biometric systems.

4. **Collectability**: Refers to how easy to collect the required traits. In EEG-based biometric, it is required to collect brain waves in a way that is easy and acceptable to the users. Therefore, it is required to minimize the number of channels required to collect reliable data.

5. **Security**: Security is the main goal for biometric systems. It is not enough to build high-accuracy systems that are able to identify the genuine successfully.

It is important also to prevent the imposters to attack the systems.

6. **Stability**: Aims to build a stable system from performance and implementability perspectives. One challenge to be addressed is the required duration for EEG biometric. This means to know what is the minimum duration of EEG data that produced the highest possible performance.

7. **User Friendly**: It may require performing a specific task to be authenticated by an EEG-biometric. User-Friendly means that the performed task is pleasant to the user. Therefore, it is important to inquire about the subjects' satisfaction.

## 3.2 objectives

This study has three main objectives:

- **Objective 1**: To develop and investigate the possibility of using an EEG-based biometric system that has low FAR, FRR values and low computational time.
- **Objective 2**: To find a temporal threshold that balances between EEG recording duration and performance.
- **Objective 3**: To discover the influence of different auditory stimuli on the performance of EEG-based authentication.

## 3.3 hypotheses

- **H1**: It is possible to reduce the computational cost of EEG-based biometric systems.
- **H2**: There is a temporal threshold that balances between performance and implementability.
- **H3**: Auditory stimuli can improve the authentication performance
- **H4**: Auditory stimuli using In-Ear or bone conducting stimuli can affect the performance of EEG-biometric.
- **H5**: The language of auditory stimuli can affect the performance of EEG-biometric.

# Chapter 4

# Materials and Methods

## 4.1 Data sets

### 4.1.1 STEW Data set

Simultaneous Task EEG Workload Data Set (STEW) is public data set. The data contains 2.5 minutes of recording sampled at 128 Hz using the Emotiv EPOC device. The device provides 14 channels, namely, AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4. The signals were recorded from 48 subjects in two sessions. In both sessions. The subject was asked to sit on a chair facing a monitor. In the first session, the EEG data was recorded in a resting state, while in the second one, the subject encountered a mental workload activity. The activity involves a multitask test according to the SIMKAP experiment. Afterwards, the subjects were asked to assess their mental workload on a 1-9 scale [Lim et al., 2018].

### 4.1.2 EEG Alpha Data set

The EEG Alpha Waves data set's main goal is to provide a simple data set to detect alpha waves. The data were recorded using a research-grade amplifier (g.USBamp, g.tec, Schiedlberg, Austria) and the EC20 cap. The instrumentation is able to record signals from 16 channels, namely, FP1, FP2, FC5, FC6, FZ, T7, CZ, T8, P7, P3, Pz, P4, P8, O1, Oz, and O2. The overall description of the dataset is that it contains two minutes of recording sampled at 512 Hz collected from 20 subjects [Grégoire Cattan, 2018].

### 4.1.3 Local Dataset

#### Determination of eligibility

The subjects have filled a questionnaire that contains subjects' data. Additionally, the subject read and confirmed a detailed informed consent.

#### Preparation and installation of equipment

The experiments start with the installation of the equipment needed. Those include four gold-cup electrodes with Ten20 Conductive Paste on the scalp. The electrodes

were placed on T7, F8, Cz, and P4 positions according to the 10/10 international EEG system [Jurcak et al., 2007], which were chosen according to [Altahat, 2017, Ravi and Palaniappan, 2007, Marcel and Millán, 2007]. Additional two electrodes were placed in the left and right ears as a reference electrode and ground electrode. Once all electrodes have been installed, a simple calibration was performed to ensure that everything worked correctly. The calibration includes checking the connectivity of the electrodes by measuring the skin impedance.

### Data recording

The subject was asked to sit down and relax on a comfortable chair. The recording was performed in a single day per subject, and it involves the acquisition of the electroencephalographic signal as follows:

1. Three minutes of resting-state, eyes open for three sessions.
2. Three minutes of resting-state, eyes closed for three sessions.
3. Non-related experiment (Not provided in the dataset).
4. Non-related experiment (Not provided in the dataset).
5. Recording EEG signal while hearing a song in the native language using in-ear headphones.
6. Recording EEG signal while hearing a non-native language song using in-ear headphones.
7. Recording EEG signal while hearing neutral music using in-ear headphones.
8. Recording EEG signal while hearing a song in your native language using bone-conducting headphones.
9. Recording EEG signal while hearing a non-native language song using bone-conducting headphones.
10. Recording EEG signal while hearing neutral music using bone-conducting headphones.

Note 1: If the person is Italian: the Arabic song was used for the non-Native song.
Note 2: If the person is not Italian: the Italian song was used for the non-Native song.

### Recording Tools

1. OpenBCI Ganglion Board, 200 Hz sampling rate, four channels.
2. Gold Cup Electrodes.

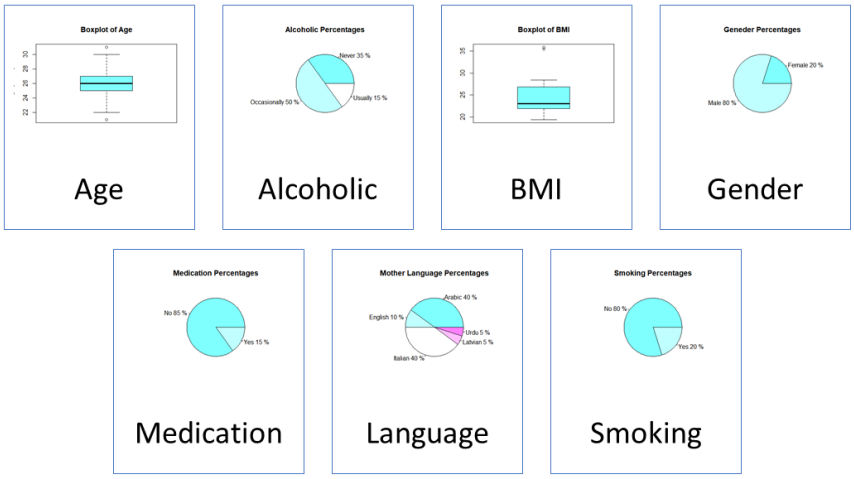


Figure 4.1: The metadata of the locally recorded dataset

3. Ten20 Conductive Paste.

4. Software: OpenBCI GUI · v5.0.3

The Metadata of the recorded dataset are presented in Fig 4.1. The detailed information is presented in Appendix 1.

## 4.2 Preprocessing

### 4.2.1 STEW Data set

The preprocessing of the STEW Data set involves three main procedures:

1. Segmentation: The Data was segmented into 60 segments. Each segment has a 4-seconds length and without overlapping

2. Filtering: The Data was filtered using a 1-40 Hz first-order Butterworth filter.

### 4.2.2 EEG Alpha Data set

The preprocessing of the EEG Alpha Data set involves two main procedures:

1. Segmentation: The Data was segmented into 36 segments. Each segment has a 4-seconds length and without overlapping

2. Filtering: The Data was filtered using a 1-40 Hz first-order Butterworth filter.

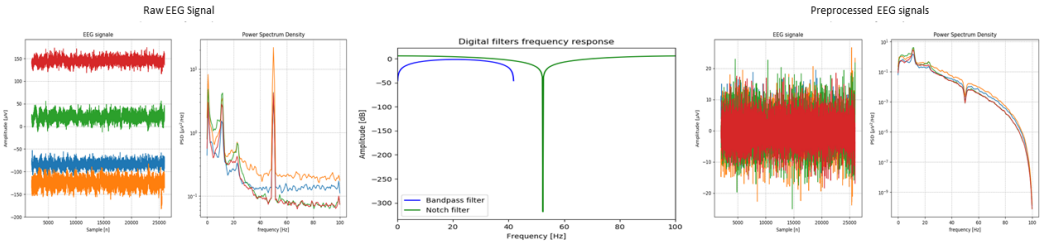


Figure 4.2: preprocessing of the local dataset.

### 4.2.3 Local Data set

The preprocessing of the EEG Alpha Data set involves two main procedures:

1. Segmentation: The Data was segmented into 45 segments. Each segment has a 4-seconds length and without overlapping

2. Filtering: The Data was filtered using a 1-40 Hz first-order Butterworth filter. Additionally, a 50 Hz notch filter was used to eliminate the power line interference with a quality factor equals to 30.

## 4.3 Efficient feature selection for electroencephalogram-based authentication

### 4.3.1 feature Selection

This part will discuss the methods of feature selection. In this work, 17 features were extracted. The features are:

1. Amplitude Mean

2. Amplitude standard deviation

3. Amplitude Variance

4. Amplitude Range

5. Amplitude kurtosis

6. Amplitude skewness
7. PSD Mean
8. PSD standard deviation
9. Permutation Entropy
10. Spectral Entropy
11. Singular value decomposition Entropy
12. Approximate Entropy
13. Sample Entropy
14. Petrosian Fractal Dimension
15. Katz Fractal Dimension
16. Higuchi Fractal Dimension
17. Detrended fluctuation analysis

Profound mathematical details about those features were mentioned in the section2.3.3. The rest of the section will select the best feature set. Four Datasets were used in this section: STEW: Resting-State, STEW: Mental workload, EEG Alpha Dataset, and the locally recorded dataset. Additionally, the features extraction time (FE time) was calculated. To do so, the FE Time of all subjects and all segments were recorded. Afterwards, the recorded time was divided by the number of subjects and the number of segments. The FE time was calculated for the full feature set as well as the reduced feature sets.

### CLuster Map

The cluster map, also known as the cluster heat map, is a simple way to present the relation between variables. It is represented as a rectangular tiling of the data's variables as a colour-shaded matrix. On its margin, cluster trees are appended [Wilkinson and Friendly, 2009]. In order to study the relationship between the features, a cluster map is presented in 4.3. The goal is to cluster the features together where clustered features represent the same information. This means features that belong to the same cluster represent redundant information. Therefore, all clustered features could be replaced by a single feature. The criteria to decide whether a feature belongs to a cluster is when the correlation coefficient of the is higher than 0.5 with four other features. As seen in Fig 4.3, three clusters are presented while detrended fluctuation analysis does not belong to any cluster. The first cluster includes:

1. Katz Fractal Dimension
2. Sample Entropy
3. Singular value decomposition Entropy
4. Higuchi Fractal Dimension
5. Spectral Entropy
6. Approximate Entropy
7. Permutation Entropy
8. Petrosian Fractal Dimension

Where sample entropy was chosen as representative for the first cluster since it contains higher correlations with other features in the cluster. The second cluster includes:

1. Amplitude Mean
2. PSD mean
3. PSD standard deviation
4. Amplitude Variance
5. Amplitude standard deviation
6. Amplitude Range

Where Amplitude Variance was chosen as representative for the second cluster since it contains higher correlations with other features in the cluster. The third cluster includes:

1. Amplitude kurtosis
2. Amplitude skewness

Where Amplitude Variance was chosen arbitrarily, the final cluster was presented by Detrended fluctuation analysis. Overall, the cluster map method resulted in four features in the first reduced feature set:

1. Sample Entropy
2. Amplitude Variance
3. Amplitude kurtosis
4. Detrended fluctuation analysis

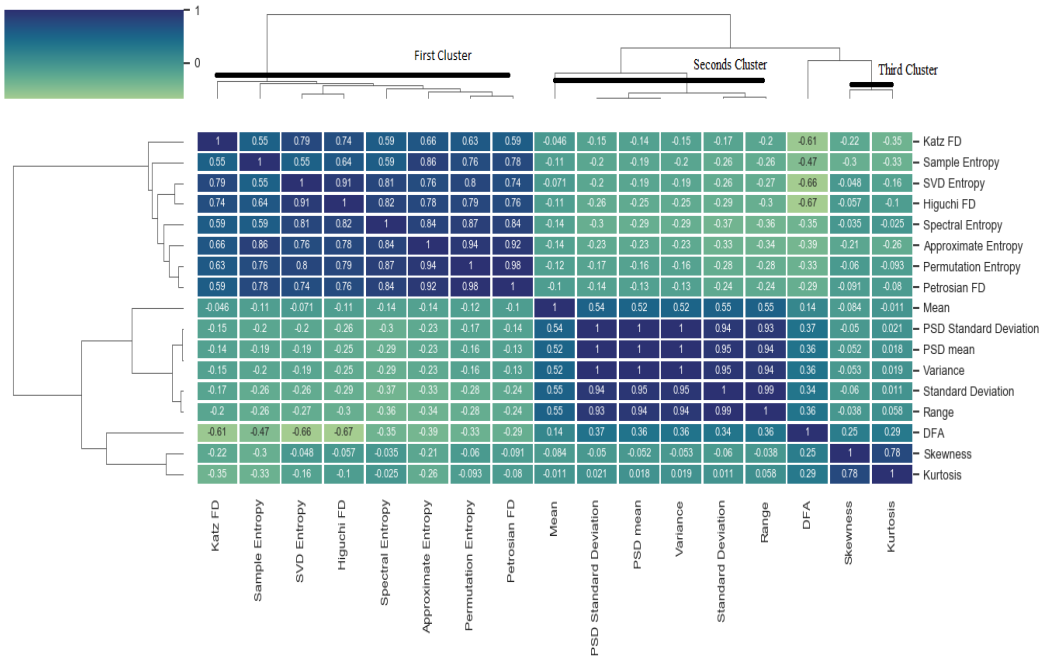


Figure 4.3: Cluster map method for feature selection

### Anova F-Value

It is considered a statistical method to calculate the variance between features and labels. The principle of ANOVA (analysis of variance) is determining the difference between means of each group. In other words, determining whether the mean of each group is equal. The output of int rest in ANOVA is the F-Value, which reveals the distance between features [Elssied et al., 2014]. As seen in Fig 4.4, There are five features that are more relevant than others. Those features are:

1. Amplitude Range
2. Amplitude Variance
3. PSD mean
4. Amplitude standard deviation
5. PSD standard deviation

### Logistic Regression Weights

On the contrary to the previous methods, Logistic Regression Weights is considered a forward feature selection method. It is based on training a classifier and find out which features are mostly contributing to making decisions. Accordingly, each feature is given a Weight. There are three features that are more relevant than others. Those features are:

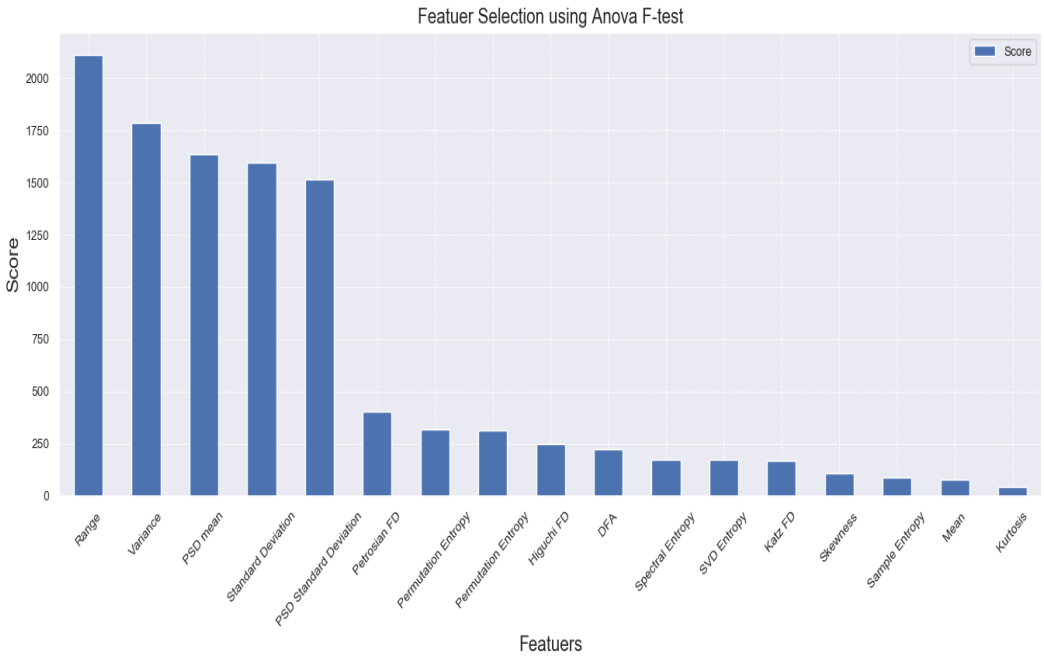


Figure 4.4: ANOVA F-Score method for feature selection

1. Katz Fractal Dimension
2. Amplitude Range
3. Amplitude Variance

### Fused feature set

Each method of the three previous methods resulted in many features that seem to be the most relevant to the task. The last feature set, which is called the fused feature set, contains their union. There are nine features as a result. Those features are:

1. Katz Fractal Dimension
2. Amplitude Range
3. Amplitude Variance
4. Amplitude kurtosis
5. Sample Entropy
6. PSD standard deviation
7. PSD mean
8. Amplitude standard deviation

9. Detrended fluctuation analysis

### 4.3.2 Classification

Each feature set of the previous four feature sets will be used as an input for a classification algorithm. Three main classification algorithms will be used, namely, MLP, KNN, and XGBoost.

## 4.4 Electroencephalography recording duration in EEG-based authentication systems

The aim of this part is to find a temporal threshold for EEG-biometric. This threshold meets two conditions to maximize the implementability of EEG-biometric. The two conditions are minimizing the EEG recording duration and maximizing the performance metrics of the system. Four Datasets were used in this section: STEW: Resting-State, STEW: Mental workload, EEG Alpha Dataset, and the locally recorded dataset.

### 4.4.1 Data segmenting

EEG data were segmented according to 10 scenarios. The segments lengths are: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2 ,3, 4, 5, 6, 7, 8, 9, 10 s, as seen in Fig 4.5. From each segment, features were extracted based on the results of section 4.3.

### 4.4.2 classification

Each scenario was classified using three different classifiers, namely, MLP, KNN, and XGBoost. The classification process was repeated three times for better robustness against classifiers randomness. For each classification time, the average and standard deviation of performance metrics were recorded.

## 4.5 Auditory stimuli

This experiment aims to answer three questions:

1. Dose the auditory stimuli affects the performance of EEG-biometric?
2. Dose the auditory conduction method affects the performance of EEG-biometric?
3. Does the EEG-biometric performance differs between native, non-native, and neutral music?

To answer those questions, the locally recorded dataset was used. Eight EEG-biometric systems were designed according to the eight experiments of the dataset as described in section 4.1.3. The extracted features were considered as the results

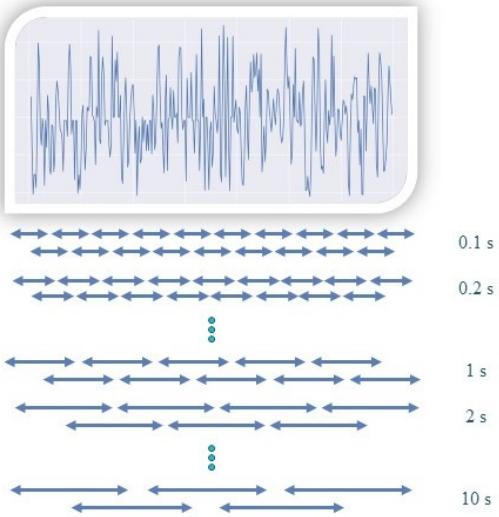


Figure 4.5: Caption

of 4.3while the signal segment length according to the results of section 4.4. Each system was trained three different times. Each time by using a specific classification algorithm, namely, MLP, KNN, and XGBoost. Additionally, subjects of the Local dataset were asked to assess their satisfaction with the experiments. They were asked to order the four types of experiments: Eyes Open Resting State, Eyes Closed Resting State, In-Ear Auditory Stimuli, and Bone-Conducting Auditory Stimuli.

# Chapter 5

# Results

## 5.1 Feature selection

The results of the feature selection are presented in Fig 5.1, Fig 5.2, and Fig 5.3. The results represent the accuracy, FAR, and FRR, respectively. Each figure shows the results of the five datasets, namely, STEW: Resting-State, STEW: Mental Workload, EEG Alpha, Local Dataset Ex02, and Local Dataset Ex07. For each dataset, the results show the performance metrics for the five feature sets applied to the three classifiers: MLP, KNN, and XGBoost. Table 5.1 shows the numeric results as a better representation. Additionally, it shows the FE time as described in the section 4.3. Table 5.2shows the percentage of time reduction while extracting the features. Finally, Table 5.3 shows a comparison with the literature.

## 5.2 Recording Duration

This part aims to investigate what is the temporal threshold of EEG recording that meets the desired criteria. Fig 5.4 shows the relationship between EEG recording and authentication accuracy. The figure shows this relationship for three data sets, STEW: Resting-State, STEW: Mental Workload, and EEG Alpha dataset. Each one was tested using the three classifiers: MLP, KNN, and XGBoost. Additionally, the results were compared with the findings of [Carrión-Ojeda et al., 2019]. Fig 5.5 the correlation coefficients between the literature and the results of the STEW dataset.

## 5.3 Auditory stimuli

To evaluate the effect of the auditory stimuli. The results of the previous two experiments were used on the Local Dataset. The results answering whether the auditory stimuli affects the performance of EEG-Biometric are presented in Fig 5.6. The figure shows the accuracy, False Acceptance Rate, and Falser Rejection Rate of the system. It shows that the accuracy is increased by 9.27% when using auditory stimuli. To compare the effect of the method of conducting and the language of the auditory stimuli, Fig 5.7, Fig 5.8, and Fig 5.9 show the accuracy, False Acceptance Rate, and Falser Rejection Rate, respectively, of the three classifiers: MLP, KNN,

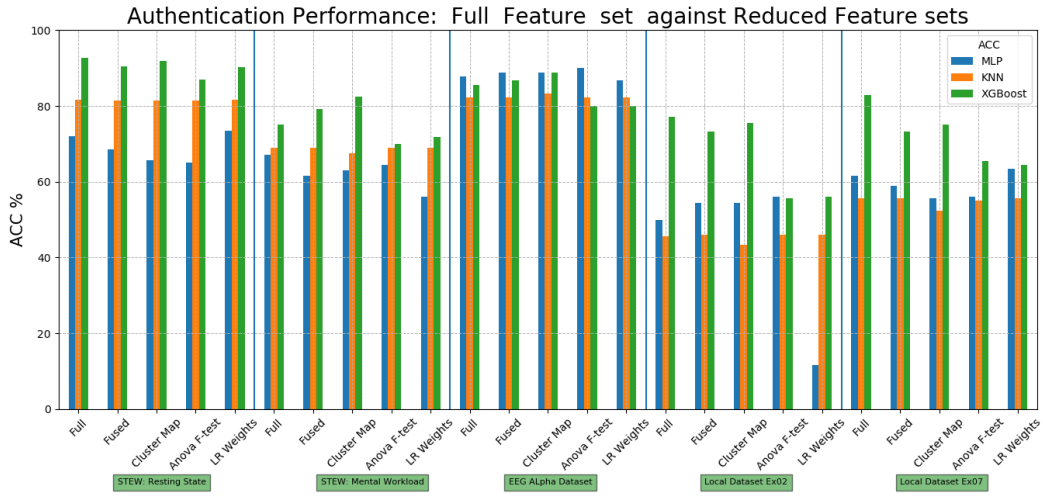


Figure 5.1: Authentication Performance (Accuracy): Full feature set against reduced feature sets.

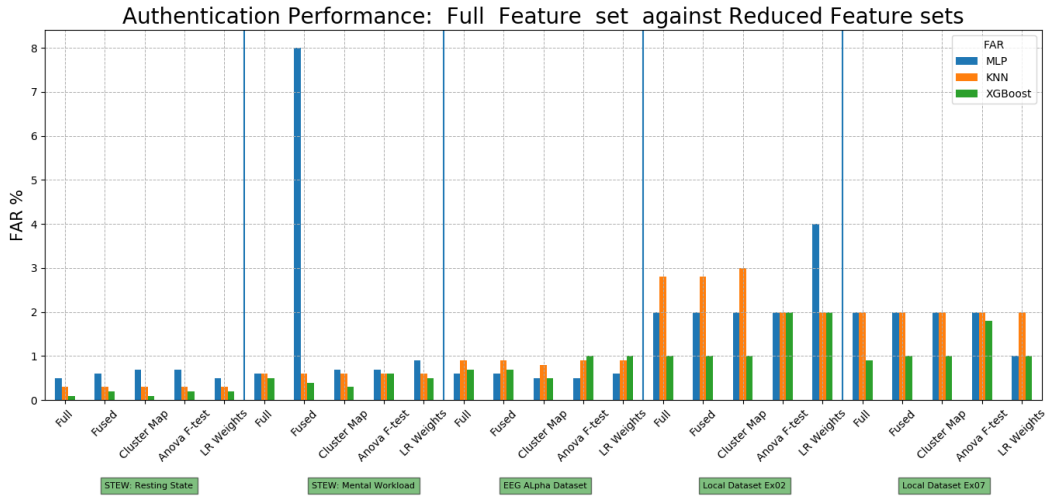


Figure 5.2: Authentication Performance (False Acceptance Rate): Full feature set against reduced feature sets.

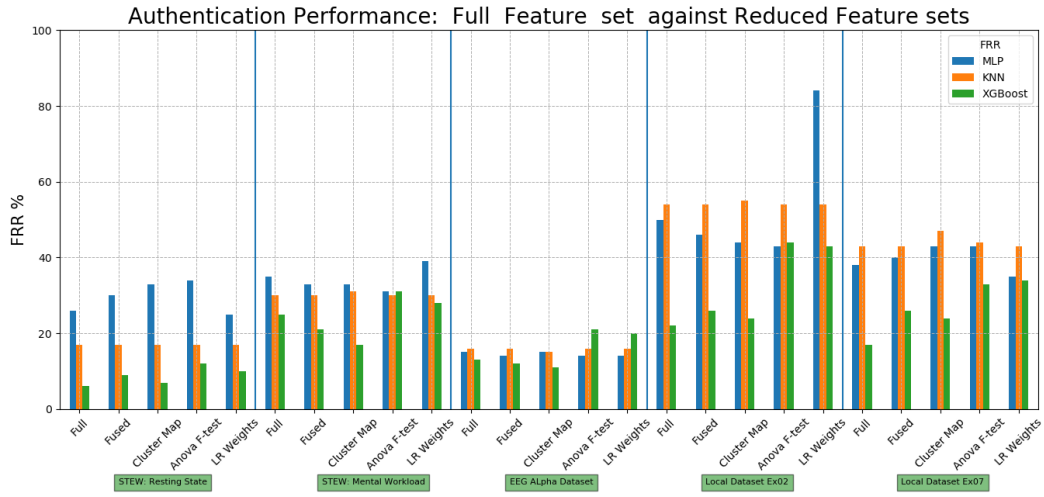


Figure 5.3: Authentication Performance (False Rejection Rate): Full feature set against reduced feature sets.

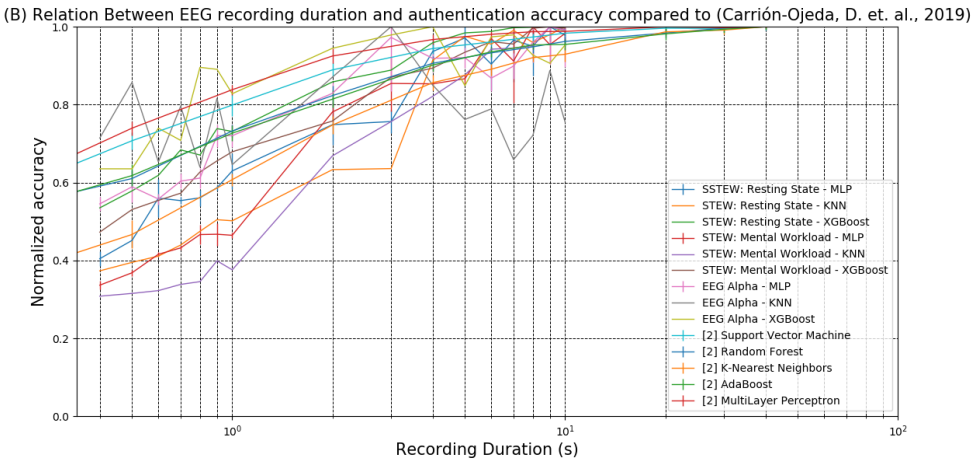


Figure 5.4: Relation between EEG recording duration and authentication accuracy.

Table 5.1: Performance metrics: full feature set against reduced feature set.

| | **Full** | **Fused** | **Cluster Map** | **Anova F-test** | **LR Weights** |
|---|---|---|---|---|---|
| | | | ***STEW: Resting State*** | | |
| **MLP** | 71.99 | 68.63 | 65.62 | 65.04 | 73.49 |
| **KNN** | 81.59 | 81.48 | 81.48 | 81.48 | 81.59 |
| **XGBoost** | 92.7 | 90.39 | 91.89 | 87.03 | 90.22 |
| **FE Time** *[s] | 0.198 | 0.054 | 0.020 | 0.019 | 0.050 |
| | | | ***STEW: Mental Load*** | | |
| **MLP** | 67.12 | 61.57 | 62.96 | 64.35 | 56.01 |
| **KNN** | 68.98 | 68.98 | 67.59 | 68.98 | 68.98 |
| **XGBoost** | 75 | 79.16 | 82.4 | 69.9 | 71.75 |
| **FE Time** *[s] | 0.198 | 0.054 | 0.020 | 0.019 | 0.050 |
| | | | ***EEG Alpha Wave Dataset*** | | |
| **MLP** | 87.77 | 88.88 | 88.88 | 90 | 86.66 |
| **KNN** | 82.22 | 82.22 | 83.33 | 82.22 | 82.22 |
| **XGBoost** | 85.55 | 86.66 | 88.88 | 80 | 80 |
| **FE Time** *[s] | 1.440 | 0.433 | 0.245 | 0.036 | 0.029 |
| | | | ***Local Dataset Ex02*** | | |
| **MLP** | 50 | 54.33 | 54.44 | 56.1 | 11.66 |
| **KNN** | 45.55 | 46.11 | 43.38 | 46.11 | 46.11 |
| **XGBoost** | 77.22 | 73.33 | 75.55 | 55.55 | 56.11 |
| **FE Time** *[s] | 0.0038 | 0.00113 | 0.00098 | 0.000362 | 0.000262 |
| | | | ***Local Dataset Ex07*** | | |
| **MLP** | 61.66 | 58.88 | 55.55 | 56.11 | 63.33 |
| **KNN** | 55.55 | 55.55 | 52.27 | 55 | 55.55 |
| **XGBoost** | 82.77 | 73.33 | 75 | 65.55 | 64.44 |
| **FE Time** *[s] | 0.0038 | 0.00113 | 0.00098 | 0.000362 | 0.000262 |

* *All experiments were conducted using PC with Core i7-3537U, 2.00 GHz Processor and 8.00 GB RAM.*

Table 5.2: Feature extraction time reduction for the reduced Feature sets

| | **Fused** | **Cluster Map** | **Anova F-test** | **LR Weights** |
|---|---|---|---|---|
| | | ***STEW: Resting State*** | | |
| **MLP** | 72.73% | 89.90 % | 90.40 % | 74.75 |
| **KNN** | 69.93% | 82.99 % | 97.50 % | 97.99 |
| **XGBoost** | 70.26% | 74.21 % | 90.47 % | 93.11 |
| **Average ± std** | 70.97 ± 1.53 % | 82.37 ± 7.86 % | 92.79 ± 4.08 % | 88.61 ± 12.25 % |

Table 5.3: Performance Metrics: Comparison with literature. The values represent the highest accuracy obtained using the cluster map feature set.

| | **Accuracy** |
|---|---|
| *STEW: Resting State* | 91.89% |
| *STEW: Mental Load* | 82.4% |
| *EEG Alpha* | 88.88% |
| *Local Dataset* | 75.55% |
| *[Tsai et al., 2020]* | 96.80% |
| *[Marino et al., 2020]* | 88.43% |
| *[Chen and Yin, 2020]* | 87.30% |

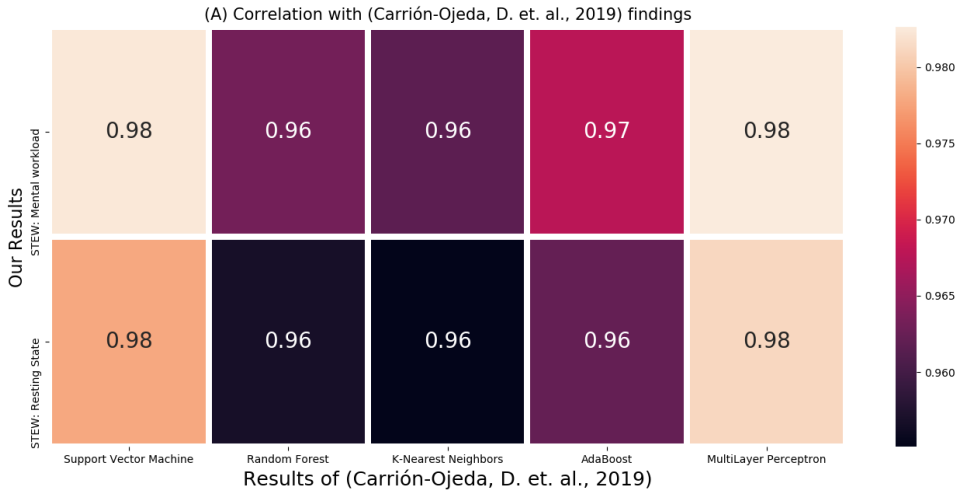


Figure 5.5: Correlation Coefficients with literature

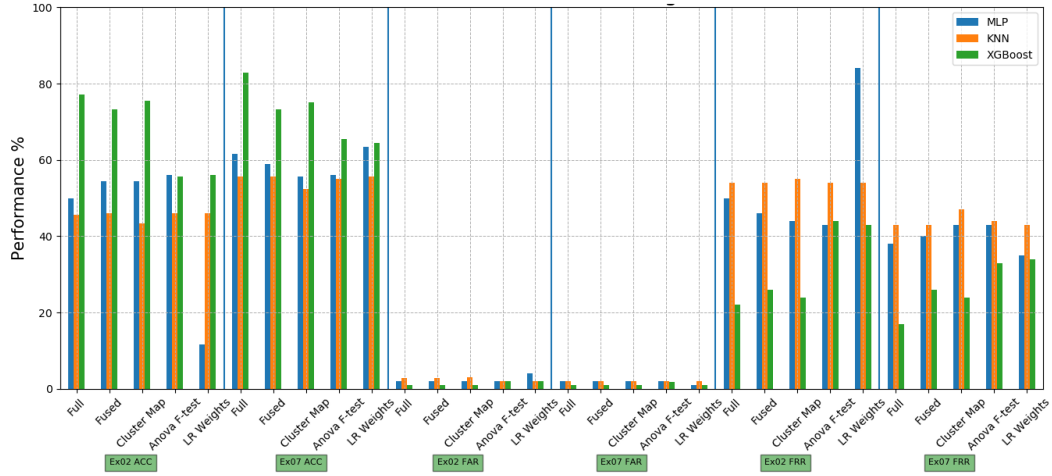


Figure 5.6: Authentication performance (Accuracy, FAR, and FRR): Ex02 (Resting State Closed eyes) versus Ex07 (Auditory Stimuli)

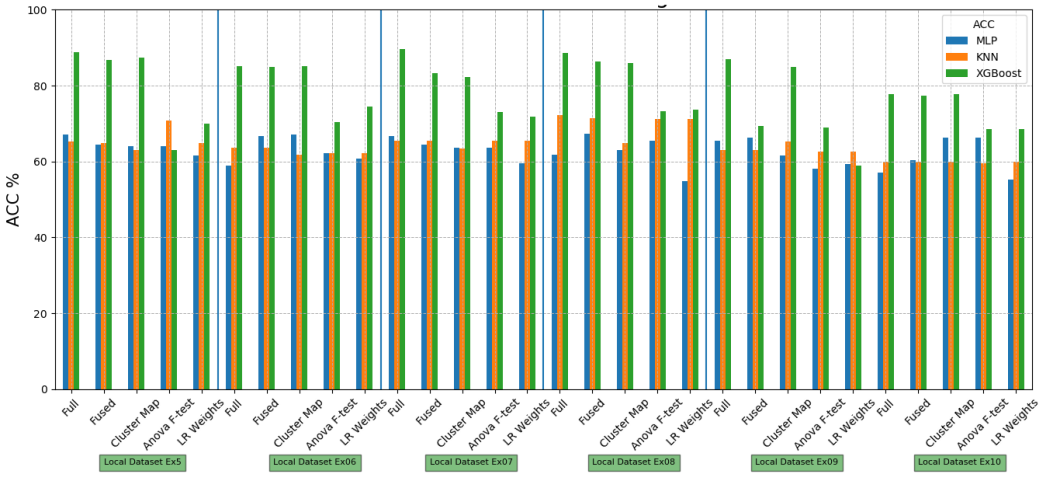


Figure 5.7: Authentication Performance with different auditory stimuli (Accuracy).

and XGBoost. Additionally, Fig 5.10 shows the stratification of the subjects when in four cases: Open eyes, Close eyes, In-Ear auditory stimuli, Bone-conducting auditory stimuli. The results of the subjects Satisfaction is shown in Fig 5.10.

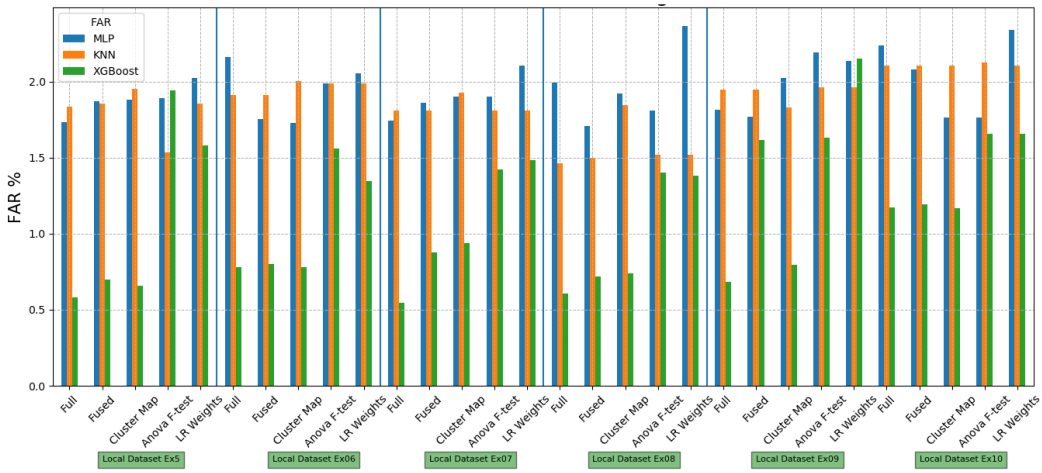


Figure 5.8: Authentication Performance with different auditory stimuli (FAR).

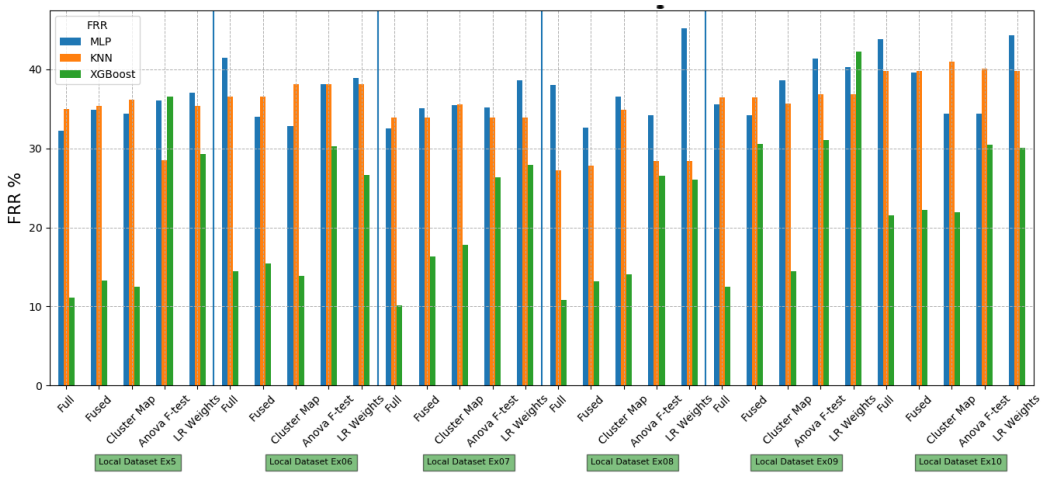


Figure 5.9: Authentication Performance with different auditory stimuli (FRR).

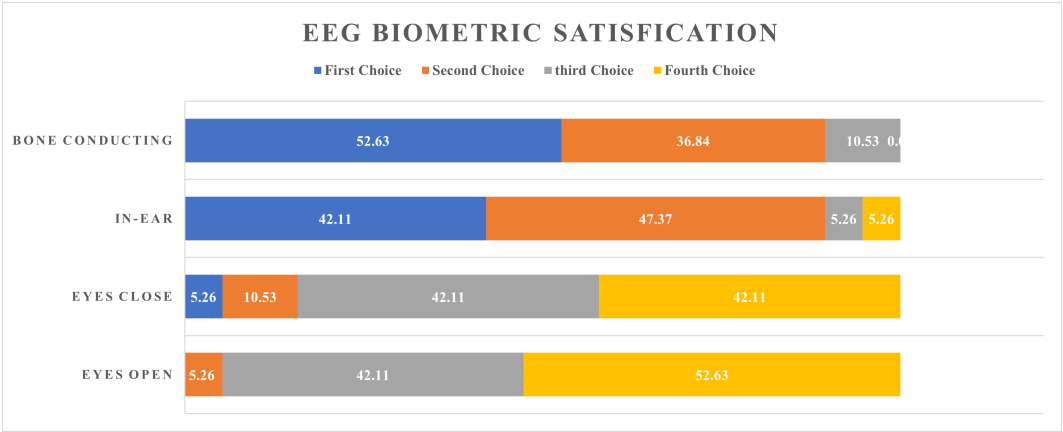


Figure 5.10: Subject's Satisfaction of EEG-Biometric

# Chapter 6

# Discussion

## 6.1 Feature Selection

From what could be observed from Fig 5.1, Fig 5.2, Fig 5.3, and Table 5.1, the results show a promising feature extraction. The reduced feature sets could ac hive similar or even higher performance than the full feature set. This could be explained that the reduced feature set contains only the information required for identity from the brain. In other words, those features represent how the brains of different individuals are different. The major benefit for the reduced sets is the decrease of information redundancy, which in roll dramatically decreases the time required for feature extraction, as seen in Table 5.1 and Table 5.2. We can conclude that the cluster map feature set is the reduced feature set that makes a balance between performance and computation efficiency. It performed best while using the XGBoost classifier, reaching accuracy over 91% and FAR is below 0.1% while FRR is around 7%. This means 1 out of 1000 times an imposter could have unauthorised access while 1 out of 14 tries of a genuine will be failed. This feature set grantee 82.37 $\pm$ 7.86 % computation time reduction, which in roll improves the implementability.

To compare this work with literature, as seen in Table 5.3, it is noticeable that it only [Tsai et al., 2020] outperformed this work. This is explained because they tend to use the deep convolutional neural network, which is computationally costly. It [Thompson et al., 2020], and require a lot of time to be trained and implemented. This will reduced implementability of thier system, which they did not discuss. Moving on to [Marino et al., 2020], they have ac hived the highest performance equals to 88.43%. However, their dataset suffers from high redundancy where they used 75% overlapping when they extracted their feature, compared with no overlapping in this work. Finally, [Chen and Yin, 2020] has a major drawback: the number of electrodes, which is equal to 40 electrodes. This might be a good solution to improve the performance. However, it make it useless from practical point of view.

## 6.2 Recording Duration

The goal is to find a temporal threshold for the EEG recording duration that gives a balance between performance and implementability. As seen in Fig 4.5, It shows an

increasing trend from 0.1 s until 4 s. After 4 s, no significant increase was observed, meaning that 4 s shall be considered a measure of the minimum recording duration needed to achieve optimal performance. The EEG-based authentication system will be less effective for shorter recordings. At the same time, the longer time duration does not offer sufficient benefits in terms of performance to justify the corresponding increased requirement of computational and memory resources.

To compare this work with literature, specifically [Carrión-Ojeda et al., 2019], Results show correlation coefficients in the range: 0.96 – 0.98 with their work, as in Fig. Hence, these are in line with their findings.

## 6.3 Auditory stimuli

This part aims to answer three questions:

1. **Dose the auditory stimuli affects the performance of EEG-biometric?**
   From Fig 5.6 we can see a comparison between Ex02 and Ex07, which were conducting in resting state and auditory stimuli respectively, as described in the section 4.1.3, we can see that the auditory stimuli case outperformed the resting state by a 9.27% difference in accuracy. This can be explained that the brain response to the auditory stimuli generates different oscillation patterns in EEG signals. Therefore, the authentication performance in the presence of auditory stimuli is better. As seen in Fig , in terms of implementability, as seen in Fig 5.10 shows that subjects were 89.47% of the subjects have chosen auditory stimuli as their first choice for satisfaction and 89.47% as a second choice. This means the use of auditory stimuli is preferable from both performance and implementability perspectives.

2. **Dose the auditory conduction method affects the performance of EEG-biometric?**
   Answering the question based on Fig 5.7, Fig 5.8, and Fig 5.9 is not enough to tell if there is a significant difference. Therefore, To answer this question, the paired t.test was performed between In-Ear auditory stimuli and bone-conducting stimuli. The results shows $P=0.038<0.05$ for accuracy, $P=0.046<0.05$ for FAR, and $P=0.032<0.05$ for FRR. This means there is a significant difference between In-Ear auditory stimuli and bone-conducting auditory stimuli. The average accuracy shows that in-Ear ($69.33 \pm 8.92$) is slightly higher than Bone-conducting ($67.60 \pm 8.78$). Regarding user satisfaction and implementability, from Fig 5.10, In-Ear auditory stimuli were preferable by 36.84% of the subjects compared to 52.63% of subjects for bone-conducting auditory stimuli. These contradictory results between performance and implementability give the developer the option to choose their preferred traits.

3. **Does the EEG-biometric performance differs between native, non-native, and neutral music?**
   Answering this question can not be straight forward similarly to the previous one. To discover whether there is a significant difference between the three groups, the ANOVA test was applied. The test resulted in a p-Value>0.05, which is not significant. This means that EEG-biometric performance is independent of the language of the auditory stimuli. This comes in line with what [Jayarathne et al., 2016] they have found; they found that EEG-biometric authentication is independent of the genre of the music.

# Chapter 7

# Conclusion

This thesis has discussed the designing and implementability of authentication systems based on EEG Signals. The results of this work have answered several questions that were asked in the literature. The main contribution of this work could be summarized in five points:

1. Finding a feature set that reduces the computation time by 82.37%.
2. Finding a temporal threshold equals 4 seconds, which balances between performance and implementability.
3. Finding that using auditory stimuli could improve the authentication performance by 9.27%.
4. Finding that using In-Ear auditory stimuli is better than using bone-conducting auditory stimuli in terms of performance, while the contradictory situation in terms of implementability.
5. Finding that EEG-biometric performance is independent of the language of the auditory stimuli.

In the end, further research is still required to improve both the performance and implementability of EEG-based security systems. There are still several gaps to be bridged until the system can reach a real-life implementation.

# Bibliography

[Acharya et al., 2009] Acharya, U. R., Chua, C. K., Lim, T.-C., Dorithy, and Suri, J. S. (2009). Automatic identification of epileptic eeg signals using nonlinear parameters. *Journal of Mechanics in Medicine and Biology*, 9(04):539–553.

[Alchalabi and Faubert, 2019] Alchalabi, B. and Faubert, J. (2019). A comparison between bci simulation and neurofeedback for forward/backward navigation in virtual reality. *Computational intelligence and neuroscience*, 2019.

[Altahat, 2017] Altahat, S. H. Q. (2017). *Robust EEG Channel Set for Biometric Application*. PhD thesis, University of Canberra.

[Alzahab et al., 2019] Alzahab, N. A., Alimam, H., Alnahhas, M. S., Alarja, A., and Marmar, Z. (2019). Determining the optimal feature for two classes motor-imagery brain-computer interface (l/r-mi-bci) systems in different binary classifiers. *International Journal of Mechanical and Mechatronics Engineering*, 19(1):132–150.

[Alzahab et al., 2021] Alzahab, N. A., Apollonio, L., Di Iorio, A., Alshalak, M., Iarlori, S., Ferracuti, F., Monteriù, A., and Porcaro, C. (2021). Hybrid deep learning (hdl)-based brain-computer interface (bci) systems: A systematic review. *Brain Sciences*, 11(1):75.

[Bandt and Pompe, 2002] Bandt, C. and Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. *Physical review letters*, 88(17):174102.

[Bhattacharyya et al., 2017] Bhattacharyya, S., Khasnobish, A., Ghosh, P., Mazumder, A., and Tibarewala, D. (2017). A review on brain imaging techniques for bci applications. In *Medical Imaging: Concepts, Methodologies, Tools, and Applications*, pages 300–330. IGI Global.

[Bidgoly et al., 2020] Bidgoly, A. J., Bidgoly, H. J., and Arezoumand, Z. (2020). A survey on methods and challenges in eeg based authentication. *Computers & Security*, 93:101788.

[Cantor and Evans, 2013] Cantor, D. S. and Evans, J. R. (2013). *Clinical neurotherapy: application of techniques for treatment*. Academic Press.

[Carrión-Ojeda et al., 2019] 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)*, pages 1–6. IEEE.

[Castano-Candamil et al., 2015] Castano-Candamil, S., Höhne, J., Martínez-Vargas, J.-D., An, X.-W., Castellanos-Domínguez, G., and Haufe, S. (2015). Solving the eeg inverse problem based on space–time–frequency structured sparsity constraints. *NeuroImage*, 118:598–612.

[Castermans et al., 2014] Castermans, T., Duvinage, M., Cheron, G., and Dutoit, T. (2014). Towards effective non-invasive brain-computer interfaces dedicated to gait rehabilitation systems. *Brain sciences*, 4(1):1–48.

[Chen and Guestrin, 2016] Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

[Chen and Yin, 2020] Chen, Y. and Yin, J. (2020). Design of electroencephalogram authentication access control to smart car. *Healthcare Technology Letters*, 7(4):109–113.

[Da Silv, 2005] Da Silv, F. L. (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Lippencott Wlliams & Wilkins.

[Del Pozo-Banos et al., 2014] Del Pozo-Banos, M., Alonso, J. B., Ticay-Rivas, J. R., and Travieso, C. M. (2014). Electroencephalogram subject identification: A review. *Expert Systems with Applications*, 41(15):6537–6554.

[Delgado-Bonal and Marshak, 2019] Delgado-Bonal, A. and Marshak, A. (2019). Approximate entropy and sample entropy: A comprehensive tutorial. *Entropy*, 21(6):541.

[Elssied et al., 2014] Elssied, N. O. F., Ibrahim, O., and Osman, A. H. (2014). A novel feature selection based on one-way anova f-test for e-mail spam classification. *Research Journal of Applied Sciences, Engineering and Technology*, 7(3):625–638.

[Esteller et al., 1999] Esteller, R., Vachtsevanos, G., Echauz, J., and Lilt, B. (1999). A comparison of fractal dimension algorithms using synthetic and experimental data. In *1999 IEEE International Symposium on Circuits and Systems (ISCAS)*, volume 3, pages 199–202. IEEE.

[Fairhurst, 2018] Fairhurst, M. (2018). *Biometrics: A Very Short Introduction*. Oxford University Press, USA.

[Fulop and Fitz, 2006] Fulop, S. A. and Fitz, K. (2006). Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications. *The Journal of the Acoustical Society of America*, 119(1):360–371.

[Gamassi et al., 2005] Gamassi, M., Lazzaroni, M., Misino, M., Piuri, V., Sana, D., and Scotti, F. (2005). Quality assessment of biometric systems: a comprehensive perspective based on accuracy and performance measurement. *IEEE Transactions on Instrumentation and Measurement*, 54(4):1489–1496.

[Giardini et al., 2000] Giardini, M. E., Corti, M., Lago, P., and Gelmetti, A. (2000). Portable microcontroller-based instrument for near-infrared spectroscopy. In *Biomedical Diagnostic, Guidance, and Surgical-Assist Systems II*, volume 3911, pages 250–255. International Society for Optics and Photonics.

[Grégoire Cattan, 2018] Grégoire Cattan, Pedro L. C. Rodrigues, . M. C. (2018). Eeg alpha waves dataset.

[Gui et al., 2019] Gui, Q., Ruiz-Blondet, M. V., Laszlo, S., and Jin, Z. (2019). A survey on brain biometrics. *ACM Computing Surveys (CSUR)*, 51(6):1–38.

[Hardstone et al., 2012] Hardstone, R., Poil, S.-S., Schiavone, G., Jansen, R., Nikulin, V. V., Mansvelder, H. D., and Linkenkaer-Hansen, K. (2012). Detrended fluctuation analysis: a scale-free view on neuronal oscillations. *Frontiers in physiology*, 3:450.

[Higuchi, 1988] Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, 31(2):277–283.

[Homan et al., 1987] Homan, R. W., Herman, J., and Purdy, P. (1987). Cerebral location of international 10–20 system electrode placement. *Electroencephalography and clinical neurophysiology*, 66(4):376–382.

[Jain et al., 2011] Jain, A. K., Ross, A. A., and Nandakumar, K. (2011). *Introduction to biometrics*. Springer Science & Business Media.

[Jasper, 1958] Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.*, 10:370–375.

[Jayarathne et al., 2016] Jayarathne, I., Cohen, M., and Amarakeerthi, S. (2016). Brainid: Development of an eeg-based biometric authentication system. In *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 1–6. IEEE.

[Jelinek et al., 2019] Jelinek, H. F., Donnan, L., and Khandoker, A. H. (2019). Singular value decomposition entropy as a measure of ankle dynamics efficacy in a y-balance test following supportive lower limb taping. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2439–2442. IEEE.

[Jurcak et al., 2007] Jurcak, V., Tsuzuki, D., and Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage*, 34(4):1600–1611.

[Kannathal et al., 2005] Kannathal, N., Choo, M. L., Acharya, U. R., and Sadasivan, P. (2005). Entropies for detection of epilepsy in eeg. *Computer methods and programs in biomedicine*, 80(3):187–194.

[Katz, 1988] Katz, M. J. (1988). Fractals and the analysis of waveforms. *Computers in biology and medicine*, 18(3):145–156.

[Krishnan et al., 2020] Krishnan, P. T., Raj, A. N. J., Balasubramanian, P., and Chen, Y. (2020). Schizophrenia detection using multivariateempirical mode decomposition and entropy measures from multichannel eeg signal. *Biocybernetics and Biomedical Engineering*, 40(3):1124–1139.

[Lim et al., 2018] Lim, W., Sourina, O., and Wang, L. (2018). Stew: simultaneous task eeg workload data set. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(11):2106–2114.

[Lindauer et al., 2010] Lindauer, U., Dirnagl, U., Füchtemeier, M., Böttiger, C., Offenhauser, N., Leithner, C., and Royl, G. (2010). Pathophysiological interference with neurovascular coupling–when imaging based on hemoglobin might go blind. *Frontiers in neuroenergetics*, 2:25.

[Lotte, 2008] Lotte, F. (2008). *Study of electroencephalographic signal processing and classification techniques towards the use of brain-computer interfaces in virtual reality applications*. PhD thesis, INSA de Rennes.

[Lotte et al., 2007] Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of neural engineering*, 4(2):R1.

[Maity et al., 2015] Maity, A. K., Pratihar, R., Mitra, A., Dey, S., Agrawal, V., Sanyal, S., Banerjee, A., Sengupta, R., and Ghosh, D. (2015). Multifractal detrended fluctuation analysis of alpha and theta eeg rhythms with musical stimuli. *Chaos, Solitons & Fractals*, 81:52–67.

[Marcel and Millán, 2007] Marcel, S. and Millán, J. d. R. (2007). Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 29(4):743–752.

[Marino et al., 2020] Marino, L., Alluri, V., Kumar, A., Li, S., Leider, A. M., and Tappert, C. C. (2020). Cognitive biometrics.

[Michael-Titus et al., 2010] Michael-Titus, A., Revest, P., and Shortland, P. (2010). *The Nervous System: Basic Science and Clinical Conditions*. Systems of the body. Churchill Livingstone.

[Morabito et al., 2012] Morabito, F. C., Labate, D., La Foresta, F., Bramanti, A., Morabito, G., and Palamara, I. (2012). Multivariate multi-scale permutation

entropy for complexity analysis of alzheimer's disease eeg. *Entropy*, 14(7):1186–1202.

[Mu et al., 2017] Mu, Z., Hu, J., Min, J., and Yin, J. (2017). Comparison of different entropies as features for person authentication based on eeg signals. *IET Biometrics*, 6(6):409–417.

[Nam et al., 2018] Nam, C. S., Nijholt, A., and Lotte, F. (2018). *Brain–computer interfaces handbook: technological and theoretical advances*. CRC Press.

[Petrosian, 1995] Petrosian, A. (1995). Kolmogorov complexity of finite sequences and recognition of different preictal eeg patterns. In *Proceedings eighth IEEE symposium on computer-based medical systems*, pages 212–217. IEEE.

[Phung et al., 2014a] Phung, D., Tran, D., Ma, W., Nguyen, P., and Pham, T. (2014a). Investigating the impacts of epilepsy on eeg-based person identification systems. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 3644–3648. IEEE.

[Phung et al., 2014b] Phung, D. Q., Tran, D., Ma, W., Nguyen, P., and Pham, T. (2014b). Using shannon entropy as eeg signal feature for fast person identification. In *ESANN*, volume 4, pages 413–418.

[Pincus, 1991] Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301.

[Plastino and Rosso, 2005] Plastino, A. and Rosso, O. (2005). Entropy and statistical complexity in brain activity. *europhysics news*, 36(6):224–228.

[Podder et al., 2014] Podder, P., Khan, T. Z., Khan, M. H., and Rahman, M. M. (2014). Comparative performance analysis of hamming, hanning and blackman window. *International Journal of Computer Applications*, 96(18).

[Rahi and Mehra, 2014] Rahi, P. K. and Mehra, R. (2014). Analysis of power spectrum estimation using welch method for various window techniques. *International Journal of Emerging Technologies and Engineering*, 2(6):106–109.

[Ravi and Palaniappan, 2007] Ravi, K. and Palaniappan, R. (2007). A minimal channel set for individual identification with eeg biometric using genetic algorithm. In *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, volume 2, pages 328–332. IEEE.

[Richman and Moorman, 2000] Richman, J. S. and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*.

[Rumelhart et al., 1995] Rumelhart, D. E., Durbin, R., Golden, R., and Chauvin, Y. (1995). Backpropagation: The basic theory. *Backpropagation: Theory, architectures and applications*, pages 1–34.

[Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533–536.

[Sanei and Chambers, 2007] Sanei, S. and Chambers, J. (2007). *Introduction to EEG*. Wiley Online Library.

[Sanyal et al., 2015] Sanyal, S., Banerjee, A., Pratihar, R., Maity, A. K., Dey, S., Agrawal, V., Sengupta, R., and Ghosh, D. (2015). Detrended fluctuation and power spectral analysis of alpha and delta eeg brain rhythms to study music elicited emotion. In *2015 International Conference on Signal Processing, Computing and Control (ISPCC)*, pages 205–210. IEEE.

[Sebastien et al., 2014] Sebastien, M., Nixon, M., and Li, S. (2014). Handbook of biometric anti-spoofing: trusted biometrics under spoofing attacks.

[Sharanreddy and Kulkarni, 2013] Sharanreddy, P. and Kulkarni, P. (2013). Eeg signal classification for epilepsy seizure detection using improved approximate entropy. *Int J Public Health Sci*, 2(1):23–32.

[Stollfuss et al., 2015] Stollfuss, J., Landvogt, N., Abenstein, M., Ziegler, S., Schwaiger, M., Senekowitsch-Schmidtke, R., and Wieder, H. (2015). Non-invasive imaging of implanted peritoneal carcinomatosis in mice using pet and bioluminescence imaging. *EJNMMI research*, 5(1):1–8.

[Thomas and Vinod, 2016] Thomas, K. P. and Vinod, A. P. (2016). Biometric identification of persons using sample entropy features of eeg during rest state. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 003487–003492. IEEE.

[Thompson et al., 2020] Thompson, N. C., Greenewald, K., Lee, K., and Manso, G. F. (2020). The computational limits of deep learning. *arXiv preprint arXiv:2007.05558*.

[Tiwari and Chaturvedi, 2019] Tiwari, A. and Chaturvedi, A. (2019). A multiclass eeg signal classification model using spatial feature extraction and xgboost algorithm. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4169–4175. IEEE.

[Townsend, 2008] Townsend, D. W. (2008). Dual-modality imaging: combining anatomy and function. *Journal of Nuclear Medicine*, 49(6):938–955.

[Tsai et al., 2020] Tsai, M.-H., Hsia, C.-Y., Wu, S. K., and Chen, T.-L. (2020). An individual specific electroencephalography signal pattern verification model based

on machine learning and convolutional neural network. *J. Adv. Comput. Netw*, 8:1–9.

[Uhlhaas, 2015] Uhlhaas, P. (2015). Magnetoencephalography as a tool in cognitive neuroscience: A translational perspective. In *Schizophrenia Bulletin*, volume 41, pages S98–S99. OXFORD UNIV PRESS GREAT CLARENDON ST, OXFORD OX2 6DP, ENGLAND.

[Wilcox et al., 2008] Wilcox, T., Bortfeld, H., Woods, R., Wruck, E., and Boas, D. A. (2008). Hemodynamic response to featural changes in the occipital and inferior temporal cortex in infants: a preliminary methodological exploration. *Developmental science*, 11(3):361–370.

[Wilkinson and Friendly, 2009] Wilkinson, L. and Friendly, M. (2009). The history of the cluster heat map. *The American Statistician*, 63(2):179–184.

# Appendix 1

Metadata of the local dataset

| Subject ID | Initials | Age | Gender | Profession | BMI | Smoke | Alcoholic | Medication | Mother Language |
|---|---|---|---|---|---|---|---|---|---|
| S01 | AG | 27 | Male | Student | 25.93 | Yes | Occasionally | No | Italian |
| S02 | SC | 23 | Female | Student | 22.95 | Yes | Occasionally | No | Italian |
| S03 | SB | 26 | Male | Student | 20.45 | No | Occasionally | No | Arabic |
| S04 | JH | 29 | Male | Student | 27.46 | No | Occasionally | No | Arabic |
| S05 | MA | 28 | Male | Student | 27.77 | Yes | Never | No | Arabic |
| S06 | AA | 27 | Male | Student | 28.41 | Yes | Never | No | Arabic |
| S07 | DR | 25 | Male | Carpenter | 22.04 | No | Occasionally | No | Latvian |
| S08 | IA | 24 | Female | Student | 21.88 | No | Usually | Yes | Italian |
| S09 | LI | 30 | Female | Student | 23.18 | No | Usually | No | Arabic |
| S10 | LA | 26 | Male | student | 26.06 | No | Occasionally | Yes | Italian |
| S11 | BB | 25 | Male | student | 19.39 | No | Never | No | Arabic |
| S12 | DC | 27 | Male | student | 26.23 | No | Occasionally | No | Italian |
| S13 | AD | 27 | Male | student | 26.23 | No | Occasionally | No | Italian |
| S14 | NA | 26 | Male | student | 35.92 | No | Never | No | Arabic |
| S15 | SA | 25 | Male | student | 21.89 | No | Usually | No | Italian |
| S16 | FA | 26 | Male | student | 22.76 | No | Occasionally | No | English |
| S17 | YM | 21 | Male | student | 35.49 | No | Never | No | Arabic |
| S18 | MFA | 31 | Male | student | 20.43 | No | Never | Yes | Urdu |
| S19 | ST | 22 | Female | student | 21.01 | No | Never | No | Italian |
| S20 | RB | 25 | Male | Student | 22.49 | No | Occasionally | No | English |