UNIVERSITÉ DE MONTRÉAL

EEG BIOMETRICS DURING SLEEP AND WAKEFULNESS: PERFORMANCE OPTIMIZATION AND SECURITY IMPLICATIONS

ROCIO BEATRIZ AYALA MEZA

DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION

DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES

(GÉNIE INFORMATIQUE)

NOVEMBRE 2017

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

EEG BIOMETRICS DURING SLEEP AND WAKEFULNESS: PERFORMANCE OPTIMIZATION AND SECURITY IMPLICATIONS

présenté par : AYALA MEZA Rocio Beatriz

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. PAL Christopher J., Ph. D., président

M. FERNANDEZ José M., Ph. D., membre et directeur de recherche

M. JERBI Karim, Ph. D., membre et codirecteur de recherche

M. ALOISE Daniel, Doctorat, membre

DEDICATION

Far better it is to dare mighty things, to win glorious triumphs, even though checkered by failure, than to rank with those poor spirits who neither enjoy much nor suffer much, because they live in that grey twilight that knows neither victory nor defeat.

-Theodore Roosevelt

RÉSUMÉ

L'internet des objets et les mégadonnées ont un grand choix de domaines d'application. Dans les soins de santé ils ont le potentiel de déclencher les diagnostics à distance et le suivi en temps réel.

Les capteurs pour la santé et la télémédecine promettent de fournir un moyen économique et efficace pour décentraliser des hôpitaux en soulageant leur charge. Dans ce type de système, la présence physique n'est pas contrôlée et peut engendrer des fraudes d'identité. Par conséquent, l'identité du patient doit être confirmée avant que n'importe quelle décision médicale ou financière soit prise basée sur les données surveillées. Des méthodes d'identification/authentification traditionnelles, telles que des mots de passe, peuvent être données à quelqu'un d'autre. Et la biométrie basée sur trait, telle que des empreintes digitales, peut ne pas couvrir le traitement entier et mènera à l'utilisation non autorisée post identification/authentification.

Un corps naissant de recherche propose l'utilisation d'EEG puisqu'il présente des modèles uniques difficiles à émuler et utiles pour distinguer des sujets. Néanmoins, certains inconvénients doivent être surmontés pour rendre possible son adoption dans la vraie vie : 1) nombre d'électrodes, 2) identification/authentification continue pendant les différentes tâches cognitives et 3) la durée d'entraînement et de test. Pour adresser ces points faibles et leurs solutions possibles ; une perspective d'apprentissage machine a été employée.

Premièrement, une base de données brute de 38 sujets aux étapes d'éveil (AWA) et de sommeil (Rem, S1, S2, SWS) a été employée. En effet, l'enregistrement se fait sur chaque sujet à l'aide de 19 électrodes EEG du cuir chevelu et ensuite des techniques de traitement de signal ont été appliquées pour enlever le bruit et faire l'extraction de 20 attribut dans le domaine fréquentiel. Deux ensembles de données supplémentaires ont été créés : SX (tous les stades de sommeil) et ALL (vigilance + tous les stades de sommeil), faisant 7 le nombre d'ensembles de données qui ont été analysés dans cette thèse. En outre, afin de tester les capacités d'identification et d'authentification tous ces ensembles de données ont été divises en les ensembles des Légitimes et des Intrus. Pour déterminer quels sujets devaient appartenir à l'ensemble des Légitimes, un ratio de validation croisée de 90-10% a été évalué avec différentes combinaisons en nombre de sujets. A la fin, un équilibre entre le nombre de sujets et la performance des algorithmes a été trouvé avec 21 sujets avec plus de 44 epochs dans chaque étape. Le reste (16 sujets) appartient à l'ensemble des Intrus.

De plus, un ensemble Hold-out (4 epochs enlevées au hasard de chaque sujet dans l'ensemble des Légitimes) a été créé pour évaluer des résultats dans les données qui n'ont été jamais employées pendant l'entraînement.

Deuxièmement, pour obtenir une évaluation générale, une classification préliminaire (utilisant toutes les epochs[1]) a été faite pour chaque électrode en utilisant 4 algorithmes (KNN, SVM, RF XG) avec des hyper paramètres par défaut, un ratio de 90-10% et une validation croisée de 10x10.

Troisièmement, un enlèvement des epochs a été fait afin d'équilibrer la taille des ensembles de données et aussi réduire la durée d'inscription (le moment où une personne est ajoutée à l'ensemble des Légitimes). Ensuite, les 4 algorithmes avec des hyper paramètres par défaut, un ratio de 90-10% et une validation croisée 10x10 ont été une fois de plus utilisés pour évaluer les changements de performance.

Quatrièmement, tous les algorithmes ont été optimisés en utilisant la recherche aléatoire. L'algorithme le plus performant a été testé sur les ensembles des Intrus et Hold-out. La matrice de confusion et diverses mesures d'évaluation ont été appliquées pour évaluer les résultats.

Finalement, en utilisant l'algorithme le plus performant, une dernière classification a été faite en utilisant chaque attribut. C'était pour déterminer l'importance de chacun dans la classification.

En général, les résultats sont bons et sont au-dessus du niveau de chance.

---

[1] Epoch: Fenêtre de temps extraite du signal EEG continu. C'est ainsi que le signal a été "coupé" en segments.

ABSTRACT

Internet of Things and Big Data have a variety of application domains. In healthcare they have the potential to give rise to remote health diagnostics and real-time monitoring. Health sensors and telemedicine applications promise to provide and economic and efficient way to ease patients load in hospitals.

The lack of physical presence introduces security risks of identity fraud in this type of system. Therefore, patient's identity needs to be confirmed before any medical or financial decision is made based on the monitored data.

Traditional identification/authentication methods, such as passwords, can be given to someone else. And trait-based biometrics, such as fingerprints, may not cover the entire treatment and will lead to unauthorized post-identification/authentication use.

An emerging body of research proposes the use of EEG as it exhibits unique patterns difficult to emulate and useful to distinguish subjects. However certain drawbacks need to be overcome to make possible the adoption of EEG biometrics in real-life scenarios: 1) number of electrodes, 2) continuous identification/authentication during different brain stimulus and 3) enrollment and identification/authentication duration.

To address these shortcomings and their possible solutions; a machine learning perspective has been applied.

Firstly, a full night raw database of 38 subjects in wakefulness (AWA) and sleep stages (Rem, S1, S2, SWS) was used. The recording consists of 19 scalp EEG electrodes. Signal pre-processing techniques were applied to remove noise and extract 20 features in the frequency domain. Two additional datasets were created: SX (all sleep stages) and ALL (wakefulness + all sleep stages), making 7 the number of datasets that were analysed in this thesis. Furthermore, in order to test identification/authentication capabilities all these datasets were split in Legitimates and Intruders sets. To determine which subjects were going to belong to the Legitimates set, a 90-10% cross validation ratio was evaluated with different combinations in number of subjects. At the end, a balance between the number of subjects and algorithm performance was found with 21 subjects with over 44 epochs in each stage. The rest (16 subjects) belongs to the Intruders set. Also, a Hold-

out set (4 randomly removed epochs from each subject in the Legitimate set) was produced to evaluate results in data that has never been used during training.

Secondly, to get performance estimation, a preliminary classification (using all epochs[2]) was done for each electrode using 4 algorithms (KNN, SVM, RF, XG) with default hyper parameters, 90-10% ratio and 10x10 stratified shuffled split cross validation.

Thirdly, a removal of epochs was done to balance the size of the datasets but also to reduce enrollment time (i.e. the moment where an individual is added to the set of legitimate users). Then the 4 algorithms with default hyper parameters, 90-10% ratio and 10x10 stratified shuffled split cross validation were once more employed to evaluate changes in performance.

Fourthly, all algorithms were optimized using 2 rounds of random search. The best performing one was tested on the Intruders and Hold-out sets. A confusion matrix and various evaluation metrics were applied to evaluate results.

Finally, to determine features importance, an ultimate classification, using each feature was executed.

Overall, results are good and above chance level.

[2] Epoch: Time window extracted from the continuous EEG signal. It is how the signal has been "chopped" into segments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ADD	Attention Deficit disorder
AWG	Arbitrary Waveform Generator
AEP	Auditory evoked potential
AR	Autoregressive model
DFT	Discrete Fourier Transform
DTW	Dynamic Time Warping
ECG	Electrocardiography
ED	Euclidean distance
EEG	Electroencephalography
EER	Equal error rate
EMD	Empirical Mode Decomposition
EMG	Electromyography
EOG	Electrooculogram
ERP	Event related potential
ERR	Error rate
FAR	False accept rate
FDA	Fisher Discriminant Analysis
FFT	Fast Fourier Transform
FRR	False reject rate
HTER	Half Total Error Rate
HHT	Hilbert-Huang Transform
HSA	Hilbert Spectral Analysis
IMF	Intrinsic Mode Functions

IoT	Internet of things
IQR	Interquartile range
KNN	K-nearest neighbour
LDA	Linear Discriminant Analysis
LVQ	Learning Vector Quantizer
MFCC	Mel-frequency Cepstral Coefficients
NN	Neural Networks
OOB	Out-of-bag
PSD	Power spectral density
PSG	Polysomnography
REC	Resting eyes closed
REO	Resting eyes opened
RF	Random Forest
SE	Shannon entropy
SSEP	Somatosensory evoked potential
SSS	Stratified Shuffled Split
SVM	Support Vector Machine
VEP	Visual evoked potential
WPD	Wavelet packet decomposition
XG	Extreme Gradient Boosted trees - XGBoost

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

Hospitals are crowded as many people go for not emergency causes: chronic diseases, elderly care or small problems like cough, etc. This makes the service unavailable to people who are really in need. Thanks to Internet of Things (IoT) and big data, remote diagnostics and real-time monitoring will be possible.

However, spoofing techniques are becoming more sophisticated, for example: a) Fingerprints can be spoofed by reactivating the print left by the last user (by dusting or breathing on the collection plate), using a special tape with the fingerprint of somebody else, or using prosthetic fingers; b) Faces can be spoofed using 3D masks, high resolution photos or videos and c) Iris can be spoofed using high resolution images or contact lenses. Also, all of these traits are subject to physical damage.

Furthermore, the anti-spoofing measures have a limited validity period because they require careful feature engineering to detect real traits from fake ones. Special attention has been paid to the "liveness detection" measure which works by detecting physiological properties from a living body such as: electroencephalography (EEG), pupil dilatation and blinking. So, to prevent identity fraud by spoofing attacks, EEG has been proposed. As this physiological biometric is not static, it is possible to extract unique features to distinguish subjects. Other advantages over conventional biometrics are: it is "universal" which means that every person alive has them; it can monitor vital signs and emotional states (which could be useful for preventive medicine); and as it can be recorded for long periods in real time, the subject can be identified/authenticated continuously.

A lot of research has been done using different types of acquisition protocols: VEP, resting state, and cognitive activities. Many types of features have been used: Fast Fourier Transform (FFT), Wavelet, Correlation, entropy, etc. And also many algorithms: SVM, NN, KNN, LDA, etc. Performance is promising, but technology has not been adopted yet. Some possible drawbacks are: EEG is affected by arousal, attentional states, age and disease. All this variability makes it ideal as the perfect password, as it is unpredictable, but also difficult to manage. In addition, the enrollment (training) and identification/authentication (test) time must be fast for its deployment in real life. Finally, a small number of electrodes (ideally 1) must be capable to identify/authenticate.

1.1 Research Objectives

GENERAL OBJECTIVE

Identify and authenticate persons using their EEG during awake and sleep stages.

Many studies have been done with EEG using different protocols: VEP, resting state (eyes opened and closed) and cognitive activities; but never in subjects during wakefulness, sleep and a combination of them. So, the objective of this research is to examine the potential of EEG data as a reliable biometric tool. The main application could therefore be in the context of telemedicine, which is of increasing interest (easing the load of hospitals by decentralizing resources).

In order to answer to this question, this general objective has been divided in small specific research questions.

SPECIFIC OBJECTIVES

- *What stage is more efficient for identification/authentication?*
  From a neuroscience perspective, it is interesting to know in what stage the EEG power distribution is characteristic for an individual and therefore may reflect individual traits of brain morphology that may eventually be used to personalize clinical practices.
  Furthermore, as the resting state has been reported as a non-static phenomenon (due to fluctuations in attention), then maybe with a natural recurring state like sleep it could be answered if a brainprint exists.

- *Is it possible to distinguish individuals using 1 electrode?*
  This is important as EEG biometrics has to comply with the usability criteria. If too many electrodes are used, then the technology will not be adopted.

- *What is the minimal duration, in seconds, to identify/authenticate a person?*
  This objective is inside the usability criteria. Enrollment and identification/authentication time need to be small in order to make the technology fast and therefore easily adopted.

- *Is there entropy in brain signal?*

  Previous studies have proposed the use of EEG signals to generate cryptographic keys. In cryptanalysis, entropy is used as a way to measure the unpredictability of a cryptographic key. So, a comparison of the importance of the « spectral entropy » feature against EEG bands was done to answer this question.

- *What is the effect of hyper-parameter optimization in performance?*

  Most studies focus on feature engineering. But the performance with novel features has shown to be comparable or worse than the conventional used ones. So instead of looking for more effective features, this research uses a conventional set of features and focuses on hyper-parameter optimization to boost performance. In addition, the optimization ensures a fair comparison among algorithms when choosing the best performing one.

- *Is continuous identification/authentication possible?*

  This form of identification/authentication will be useful to recognize someone regardless of how the brain is engaged during the telemedicine session. In this research, 2 states have been studied: Wakefulness and Sleep.

1.2 Thesis structure

There are 3 chapters in this thesis excluding the Introduction and the Conclusion ones.

Chapter 2 presents advanced concepts related to machine learning, biometrics and neuroscience; and a survey of existing literature related to EEG biometrics. In Chapter 3, the applied methodology and how it was useful to answer the research questions, is described. And finally in Chapter 4, the results are explained.

CHAPTER 2 BIOMETRICS AND MACHINE LEARNING: CONCEPTS AND LITERATURE REVIEW

2.1 Biometrics

It is a form of identification/authentication and access control which uses human traits to describe individuals (A. K. Jain, Bolle, & Pankanti, 2006).

According to the trait, it can be of three types:

- **Physiological:** Related to the shape of the body. Ex: fingerprints, palm veins, face recognition, DNA, palm print, hand geometry, iris recognition, retina and odour.
- **Behavioural:** Related to behaviour patterns. Ex: typing rhythm, gait and voice.
- **Of intent:** by analyzing physiological features, behaviour can be predicted. Ex: EEG, ECG.

The compliance of a particular biometric to a specific application involves several factors:

- **Universality:** every person should possess the trait.
- **Uniqueness:** the trait should be enough different among population to distinguish each one from each other.
- **Permanence:** how the trait varies in time. A good trait should be reasonably invariant over time.
- **Measurability:** ease of trait measurement.
- **Performance:** related to the accuracy, speed and robustness of the technology used to process the trait.
- **Acceptability:** how easy the population accepts the technology. If they are willing to have their trait measured.
- **Circumvention:** how easy the trait can be imitated or copied.

2.2 Modules for an EEG biometric system

Depending on the application context, a biometrics system can work in two modes: authentication and identification (A. Jain, Flynn, & Ross, 2008).

- **Authentication:** Used to prevent different people from using the same identity.
  A user willing to be recognized claims an identity, usually using a user name. The system performs a one-to-one comparison (captured biometric with stored template) to determine whether the claim is true or not. (Is this the biometric of Alice?)
- **Identification:** It aims to prevent a single person from using multiple identities. The system performs a one-to-many comparison (searching through all stored templates for a match) to determine the identity of an unknown individual (whose biometric data is this one?)

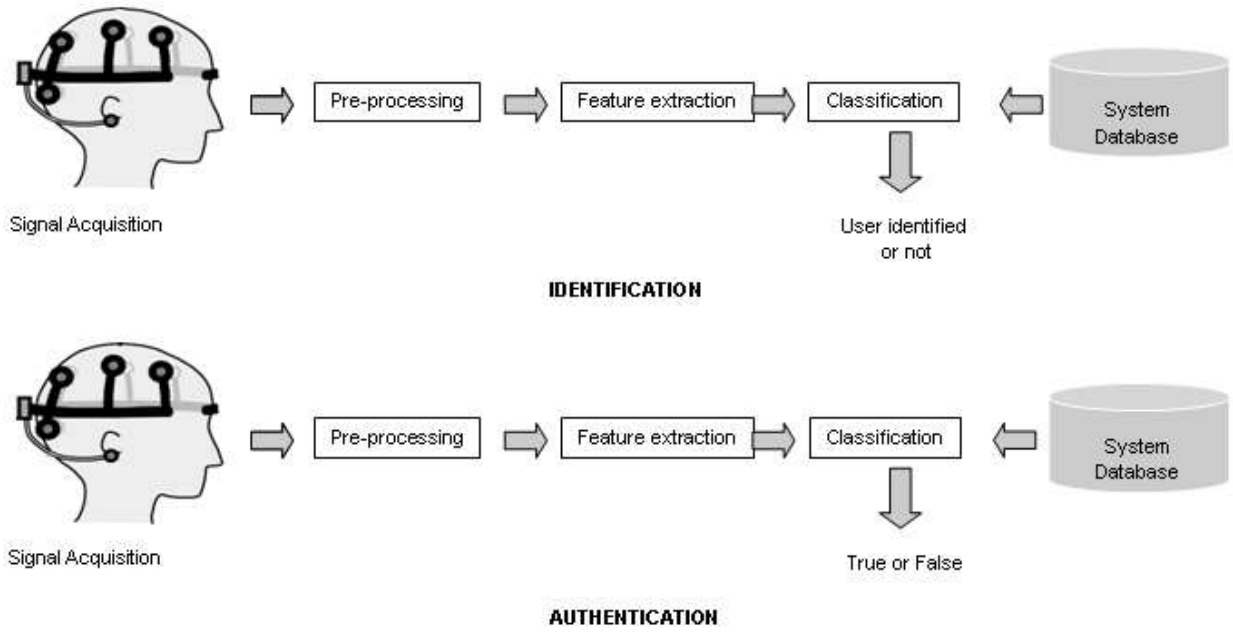


Figure 2-1 : Modules in an EEG biometric system

Figure 2-1 (A. Jain et al., 2008) (Abbas, Abo-Zahhad, & Ahmed, 2015) displays the modules of a biometric system: signal acquisition, preprocessing technique, features extraction and feature classification.

2.2.1 Signal acquisition

The trait used in this research is the Electroencephalography (EEG). The use of electrodes, a specialized headset and a brain stimulus are necessary.

a) **Electroencephalography (EEG)**

EEG is the recording of the electrical activity of the brain using electrodes placed along the scalp. It measures spontaneous electrical activity of neurons within a period of time (Niedermeyer & Silva, 2004). The amplitude of signals ranges between 10 and 200 uV with a frequency in the range of 0.5 – 100 Hz. The waveforms can be classified into different frequency bands:

Typical EEG bands: delta (< 4Hz), theta (4 Hz – 7 Hz), alpha (8 Hz – 15 Hz), beta (16 Hz – 31 Hz) and gamma (>32 Hz) (Tatum, 2014).

b) **Headsets**

In Figure 2-2 (Maskeliunas, Damasevicius, Martisius, & Vasiljevas, 2016) ("Neurosky," 2007), it can be seen the two types of headsets that have been used in biometrics research: a) medical-grade and b) low cost.

- **Medical-grade:** Contains a large number of electrodes usually of the wet type (they need to be moistened by electrolytes). Signals of better quality can be obtained but the setting up for signal acquisition is a shortcoming in biometrics applications.

  The distribution of electrodes over the scalp follow the 10-20 system or the 10-10 system (Homan, Herman, & Purdy, 1987). Both are internationally recognized methods to describe the location of electrodes for a test or experiment.

- **Low-cost:** Contains very small amount of electrodes of the dry type. Offers a low signal quality but the setting up for signal acquisition is not complex when compared to medical grade headsets. Table 2.1 shows some commercial low-cost headsets that have been used in biometrics research.

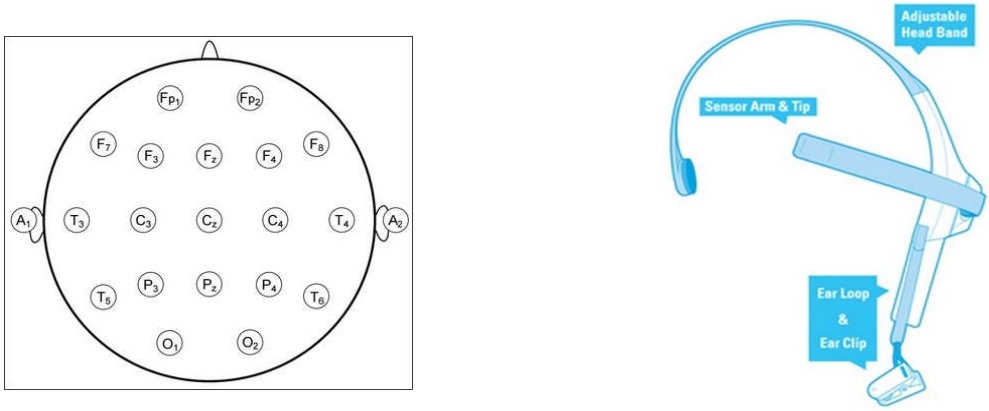


Figure 2-2 : a) A 19 electrodes positioning map based on the 10-20 system of a medical grade headset, b) A single Fp1 electrode low cost headset

Table 2-1 : Commercial low-cost headsets

Device	No. Electrodes	Released
Neurosky	1 – DRY	2007
Emotiv EPOC	14 – WET	2009
MindWave	1 – DRY	2011
iFocusBand	1 – DRY	2014
Muse	4 – DRY	2014
OpenBCI	8 or 16 – DRY/WET	2014
Aurora Dreamband	1 – DRY	2015
Melomind	4 – DRY	2015

c) **Brain stimulus**

It can be classified in 3 acquisition protocols: resting state, sensory (audio/visual) stimuli and cognitive tasks (verbal instructions). Additionally, sleep stages will be described; as they have been used in this study.

- **Resting state:** Subjects are sat on a chair in a quiet environment with either eyes open or closed. They provide EEG signals without any additional instruction. Articles that have used this protocol are presented in Appendix A.
- **Sensory stimuli:** An event related potential (ERP) is a brain response triggered by a ” specific sensory, cognitive or motor event” (Luck, 2005).

According to the source of stimuli it can be classified as Visual Evoked Potential (VEP), Auditory Evoked Potential (AEP) and Somatosensory Evoked Potential (SSEP). Until the date only VEP have been used in biometrics research (Del Pozo-Banos, Alonso, Ticay-Rivas, & Travieso, 2014).

The P300 wave belongs to the ERP category. "When recorded by EEG, it surfaces as a positive deflection in voltage with a delay between stimulus and response of 250 to 500 ms" (Polich, 2007).

To obtain a VEP, subjects are instructed to watch a series of pictures. After each picture, the EEG is recorded. P300 can be accurately measured during feature extraction. This makes them an asset for biometric systems. A drawback is the need for external devices to trigger the P300 signals. This results in a more complex system when compared to other protocols. Articles that have used this protocol are presented in Appendix B and (R. Palaniappan & Raveendran, 2002), (Ravi & Palaniappan, 2005), (R. Palaniappan, 2006), (R. Palaniappan & Mandic, 2007).

- **Cognitive activities:** It involves asking subjects to perform cognitive tasks during data collection including: mathematical calculation, geometric figure rotation, mental letter composition or visual counting. Also imaginary movements involving hands, legs, etc. Articles that have used this protocol are presented in Appendix C and (R Palaniappan, 2005).

- **Sleep stages:** Sleep scoring is a critical step used in clinical routines to diagnose pathologies such as: insomnia, hypersomnia, circadian rhythm disorders, epilepsy, apnea, etc. Sleep stages can be of 2 types: Rem and Non-Rem. Non-Rem is subdivided in: S1, S2 and SWS (sometimes known as S3). It happens in cycles which go from 90 to 120 minutes. Each cycle follows the following order: S1, S2, SWS; and after a period in SWS it goes back through S2 and S1; then it may enter Rem and a new cycle starts ("How sleep works," n.d.).

  In Figure 2-3 ("How sleep works," n.d.), the sleep cycle is shown:

  A. Rem: This is the rapid eye movement stage. It counts for up to 20 to 25 % of total sleep time in adults.

  B. S1: This is a stage between wakefulness and sleep. It usually lasts less than 10 minutes. So, it represents about 5% of the total sleep time.

C. **S2**: Particular waveforms happen here (Sleep spindles and K-complexes) to suppress any response to outside stimuli. It represents 45 to 50% of total sleep time.

D. **SWS (Slow Wave Sleep)**: This is deep sleep. It represents around 15% to 20% of total sleep time.

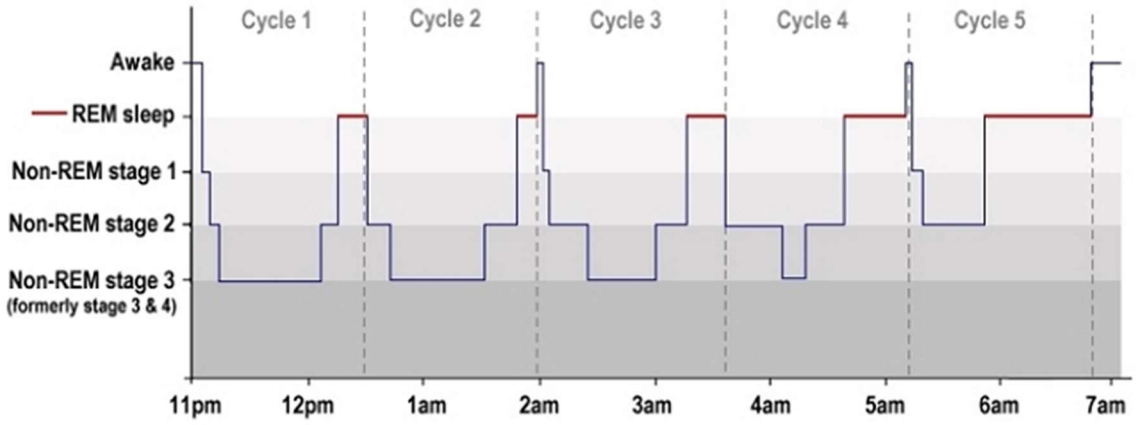


Figure 2-3 : Sleep cycles

2.2.2 Feature extraction

The selection of discriminative features is important in any type of classification problem. In EEG biometrics, the extraction of features has been conducted in the time and frequency domain.

A review of the most employed methods is provided in this section:

a) Power Spectral density (PSD)

It indicates the spectral density distribution of a signal in the frequency domain. It is computed from the Fourier Transform (FT). But, as EEG signals are non-stationary series; the truncated Fourier Transform $\hat{x}_T(\omega)$ over a finite interval [0, T] is computed. The signal is assumed to be stationary within that interval and the PSD $S_x(\omega)$ of the signal *x(t)* may be computed. Where $\hat{x}_T(\omega)$ is the FT of *x(t)* (Oppenheim, Schafer, & Buck, 1999).

$$S_x(w) = \lim_{T\to\infty} E[|\hat{x}_T(w)|^2]$$

Articles that have used this method are presented in Appendix A, B, C and (R. Palaniappan & Mandic, 2007), (R. Palaniappan & Raveendran, 2002), (Safont, Salazar, Soriano, & Vergara, 2012), (Nakanishi, Baba, & Miyamoto, 2009).

b) Autoregressive Model (AR)

It is a time domain representation of a random process. It is used to describe time-varying processes such as EEG. As its name suggests, it specifies that the output variable depends linearly on its own previous values and on a stochastic term (Pardey, Roberts, & Tarassenko, 1996).

It is defined as an AR model of order *p*:

$$X_t = c + \sum_{i=1}^{p} \varphi_i B^i X_t + \varepsilon_t$$

where the signal X_t is represented by a series of AR coefficients φ_i, white noise ε_t, a constant (c) and a lag operator (B). Articles that have used this method are presented in Appendix A, B, C and (Paranjape, Mahovsky, Benedicenti, & Koles, 2001), (Brigham & Kumar, 2010).

c) Wavelet Packet decomposition (WPD)

It decomposes the signal into both time and frequency representations which provide more information into the feature space (Daubechies, 1992).

It is defined as follows:

$$WT_\psi\{x\}(a,b) = \prec x, \psi_{a,b} \succ = \int_R x(t)\psi_{a,b}(t)dt$$

Where: $WT_\psi\{x\}(a,b)$ are the wavelet coefficients, *x(t)* is the time domain signal and $\psi_{a.b}(t)$ is the wavelet function.

It is applied by multiplying the EEG signal with different wavelet functions, in which each one can have different scale *(a)* and shift *(b)* according to specific applications.

Articles that have used this method are presented in Appendix A, B, C and (Abdullah, Subari, Leo, Loong, & Ahmad, 2010).

d) Other methods

- **Hilbert-Huang Transform (HHT):** It is a fairly new algorithm, designed to specifically work with non-stationary and nonlinear signals. As it preserves the characteristics of the varying frequency (N. Huang et al., 1998).

  It has two main steps:

  1) Uses the Empirical Mode Decomposition (EMD) method to generate the Intrinsic Mode Functions (IMF).

  2) Performs the Hilbert Spectral Analysis (HSA) on each IMF to obtain instantaneous frequency data.

  Articles that have used this method are presented in Appendix A, B, C and (Kumari, Kumar, & Vaish, 2014).

- **Euclidean distance (ED) and Dynamic Time Warping (DTW):** Employed by (Gui, Jin, Ruiz Blondet, Laszlo, & Xu, 2015), to measure the similarity between two EEG signals.
- **Shannon entropy (SE):** (Phung, Tran, Ma, Nguyen, & Pham, 2012) claimed that similar performance was obtained when compared against AR but much faster speed in identification.
- **Time domain peak matching algorithm:** Novel technique proposed by (Singhal & Ramkumar, 2007).
- **Equivalent root mean square (rms) values for each electrode signal over a 1 second period:** Time domain feature with low computational cost but high performance, proposed by (Altahat, Tran, & Sharma, 2012).
- **Conventional Mel-frequency Cepstral Coefficients (MFCC):** Suggested by (Nguyen, Tran, Huang, & Sharma, 2012). It is usually employed for voice recognition.

2.2.3 Feature classification

A classification scheme is necessary to make predictions. There are many types of algorithms. A brief description of the most popular ones used in EEG biometric is provided as well as the ones used in this research.

a) Algorithms

- **K-Nearest Neighbour (KNN):** It is a lazy learning algorithm because it constructs hypotheses directly from the training set stored in memory rather than generalizing. This makes it sensitive to the class distribution: the more frequent class will dominate the prediction of the tested epochs[3]. To overcome this, it is a good idea to assign distance weights so that the closer neighbors are the ones that contribute the most to the final prediction.

  It is also sensitive to noisy data, as it uses the distance among epochs to define the belonging to a class. To solve this, it is important to standardize the feature vectors.

  An advantage of this algorithm is its ability of adaption to unseen data and its fast classification process: as it only needs the training epochs to be stored together with their respective class labels and be compared with the query epochs directly (Kotsiantis, Zaharakis, & Pintelas, 2006).

  Articles that have used this algorithm are presented in Appendix A, B, C and (Ramaswamy Palaniappan & Ravi, 2006), (Yazdani, Roodaki, Rezatofighi, Misaghian, & Setarehdan, 2008), (Su, Xia, Cai, & Ma, 2010).

- **Kernel methods:** Support Vector Machine is one of its best known members who has been found to be competitive with neural networks on tasks such as handwriting recognition (Cortes & Vapnik, 1995).

  SVM searches for a hyperplane to separate different classes. This hyperplane has to maximize the margin between different classes to ensure generalization capabilities.

  Depending on the training set distribution, a non-linear hyperplane might be needed. For this effect a kernel function is required. Some popular examples: polynomial, radial basis function, sigmoid (C.-W. Hsu & Lin, 2002).

---

[3] Epoch: Time window extracted from the continuous EEG signal. It is how the signal has been "chopped" into segments.

By design it is a binary algorithm but several methods have been proposed to extend it to multiclass problems:

A. One-versus-all: It distinguishes between one of the classes and the rest. There is one classifier per class. The one with the highest performance assigns the class. It is computationally efficient (only n_classes classifiers are needed) and interpretable.

B. One-versus-one: It distinguishes between every pair of classes. At prediction time, the class which received the most votes is selected. It is slower than One-versus-all but scales well when using kernel methods (n_classes * (n_classes - 1) / 2).

Articles that have used this algorithm are presented in Appendix A, B, C and (Hu, 2010), (Ashby, Bhatia, Tenore, & Vogelstein, 2011).

- **Linear Discriminant Analysis (LDA):** It can be used as a classifier or as a dimensionality reduction technique. It searches for a linear combination of variables that best separates classes but assumes that all these share the same covariance matrix (McLachlan, 2004). Articles that have used this algorithm are presented in Appendix A, B, C (Lee, Kim, & Park, 2013).

- **Neural Networks (NN):** It is a collection of learning algorithms inspired by the brain's neural networks. It consists of three main layers of nodes (neurons): input, hidden and output; being the hidden layer the only one that varies in number. All these layers are interconnected and the signal propagates from the input to the output. In each node of the hidden layer there is an activation function which only responds to the output of its previous node. The results of these operations are fused and a prediction is made (Duda, Hart, & Stork, 2001). Articles that have used this algorithm are presented in Appendix A, B, C and (R Palaniappan & Mandic, 2005), (Chunying, Haifeng, Lin, & Bing, 2014), (Gui, Jin, & Xu, 2015).

- **Random Forest (RF):** It is a form of ensemble learning. It produces many trees based on a random selection of epochs and random selection of features. Then it predicts classes based on the majority vote of all trees (Breiman, 2001). It has not been used on EEG biometrics.

- **Extreme Gradient Boosted trees (XG):** It is an advanced Gradient Boosted classifier proposed by (Chen & He, 2015). It is very popular among winning solutions in Kaggle's machine learning competitions. It focuses on computational speed and model performance. Models are built sequentially and each one can correct the errors of prior models. Then all models are combined together into a strong one to make a final prediction. It has not been used on EEG biometrics.

2.3 Hyperparameter optimization

Each algorithm, according to its functioning, may or may not possess hyperparameters. The hyperparameters are the parameters that cannot be directly learnt from the dataset and therefore require fixing before model training to control model complexity. There are 2 popular brute-force techniques:

a) Grid search: It is an exhaustive search (and therefore expensive) through a finite set of "reasonable" values for each hyperparameter. It suffers from the curse of dimensionality but its workload can be separated in parallel tasks (Bergstra & Yoshua, 2012).

b) Random search: It works by sampling from a distribution of possible hyperparameter values, where the number of sampled candidates has to be specified. The higher this number is, the finer the search becomes. Frequently, some hyperparameters do not change significantly the algorithm's performance; making random search more effective than grid search in high dimensional spaces (Bergstra & Yoshua, 2012).

2.4 Performance evaluation

In biometrics there are many metrics used to evaluate a system. In this research the following ones have been used (Sokolova & Lapalme, 2009), (West Point Academy, 2012).

a) Accuracy: It is the number of all correct predictions divided by the total number of samples of the dataset. It can also be calculated by 1 – ERR. Its paradox is that for highly unbalanced datasets it will reflect the underlying class distribution. *TP* is True Positive, *TN* is True Negative, *FP* is False Positive and *FN* is False Negative.

$$ACC = \frac{TP + TN}{TP + TN + FN + FP}$$

b) Precision: Also called Positive Predictive Value. It is the proportion of predicted positives which are actual positive. *TP* is True Positive and *FP* is False Positive.

$$PRE = \frac{TP}{TP + FP}$$

c) Recall: Also called Sensitivity or True Positive Rate. It is the proportion of actual positives which are predicted positive. *TP* is True Positive and *FN* is False Negative.

$$SN = \frac{TP}{TP + FN}$$

d) F1-score: Also known as balanced F-score or F-measure. It is the harmonic mean between precision and recall.

$$F1 = \frac{2\ (precision * recall)}{precision + recall}$$

e) Type I Error: Also called False Negative Rate or False Reject Rate. It is when a system rejects access to a legitimate identity. *TP* is True Positive and *FN* is False Negative.

$$ErrorI = \frac{FN}{TP + FN}$$

f) Type II error: Also called False Positive Rate or False Accept Rate. It is the proportion of actual negatives which are predicted as positive. It is when a system grants access to an illegitimate identity. *TN* is True Negative and *FP* is False Positive.

$$ErrorII = \frac{FP}{TN + FP}$$

g) HTER: Half Total Error Rate. It is an aggregate of FAR and FRR. Most authentication systems are measured and compared using HTERs or variations of it.

$$HTER = \frac{ErrorI + ErrorII}{2}$$

2.5 Previous work and current limitations

In this section previous works are compared and analysed. This serves to give a comprehensive review of what has been done and also to justify the chosen methodology for this research.

For EEG to be considered a successful biometric technology, it has to satisfy two conditions: 1) Reliable recognition performance and 2) Usability: number of electrodes, location of electrodes, duration of training set (enrollment) and test set (recognition).

As the number of electrodes is an important condition for usability and the studied database consists of wakefulness and sleep stages; only works that have used 1 electrode (wet or dry) in resting state will be discussed. For a review of other works using different number of electrodes or other types of brain stimulus check Appendix A, B and C.

The first series of related works started in 1998 by (Poulos, 1999). Their work is based on the findings by (Vogel, 1970) about the inheritance of EEG traits. A dataset of 4 genuine users and 75 intruders in resting state with eyes closed (REC) was used. The proposed methodology fitted a linear model of the AR type on the alpha rhythm recorded by the O2 electrode and classified it by a Computational Geometry algorithm (convex polygon intersections). It reached a 95% success rate on classification. However, the fact that the dataset contained only 4 subjects prevents to make a final conclusion.

In a future work, (Poulos, Rangoussi, Alexandris, & Evangelou, 2002) used all the main brain bands (alpha, beta, delta, theta), on that same dataset. But this time using a bilinear model to extract features; prompted by an already investigated conjecture of the existence of non-linear components in EEG; and the use of a Learning Vector Quantizer (LVQ) neural network for classification. An accuracy of 99.5% was obtained.

(Paranjape et al., 2001) tested a dataset of 40 subjects in REO and REC states. The methodology also used an AR type model on the alpha rhythm for feature extraction, but on the P4 electrode. As

a classifier, a Linear Discriminant was employed reaching an impressive 99% in accuracy when all dataset was used and 85% when partitioning the dataset. Even though the 99% performance might be due to overfitting the data; the obtained accuracy of 85% when partitioning the dataset is still impressive and proved that the identification of subjects by means of EEG is possible.

In 2010, (Su, Xia, Cai, Wu, & Ma, 2010) published a study using a low cost headset with an Fp1 electrode, in which it was addressed the effects of diet and/or circadian rhythms. The dataset consisted of 40 subjects in REC where AR coefficients and PSD from 5 to 32 Hz features were extracted. Three classifiers were tested, but it was the combination of Fisher Discriminant Analysis (FDA) and KNN that reached a 97.5% when all dataset was used (without constraints of diet or circadian rhythm).

(Zhao et al., 2010) also employed a low cost headset but with an electrode in Cz; to record a dataset of 10 subjects in REC. A 97.63% was reached using KNN and linear features. These results are a contradiction because (Poulos et al., 2002) demonstrated that non-linear features were better suited for EEG biometrics. It is possible that the used algorithm (Neural Networks) has influenced the difference in results.

Finally, (Dan, Xifeng, & Qiangang, 2013) compared 3 algorithms: SVM, LDA and NN also using a low cost headset with an electrode in Fp1. The best performance was of 87% with SVM and AR features.

What all these works have in common is a high performance (over 80%), the use of a single electrode (wet or dry) in different brain locations, the continuous search for better features to discriminate among subjects and the many types of algorithms to perform classification. But there are no details whether hyperparameter optimization has been done. And this is an important step to control model complexity and to do a fair comparison among algorithms. According to (Yang, 2015) and by inspecting the Appendix A, B and C; the performance with novel features is comparable with the conventional ones (AR coefficients and PSD) or worse. So instead of looking for more effective features, this research focuses on hyperparameter optimization to boost performance, and to study the effect of reducing the number of epochs to make the biometrics system faster.

Determined by neurophysiology, brain rhythms occur in different brain regions according to the stimulus. Under the REC condition the occipital, temporal and parietal areas are the ones providing

the most discriminative information (Campisi et al., 2011), (Rocca, Campisi, & Scarano, 2012), (Parisa, Mo, & Mohammad, 2006). Meanwhile, under REO anterior brain regions are implicated (Bear, M; Connors, B; Paradiso, 2014), (Abdullah, Subari, Leo, et al., 2010), (Abdullah, Subari, Loong, & Ahmad, 2010), (P Tangkraingkij, Lursinsap, Sanguansintukul, & Desudchit, 2009), (Preecha Tangkraingkij, Lursinsap, Sanguansintukul, & Desudchit, 2010). But this represents a problem for the deployment of EEG biometrics as it does not comply with the usability criteria. Users have to see EEG biometrics as an unobtrusive option to monitor and improve their health. So, instead of using a brain region that is best for the stimulus, one that is best for the user was chosen: the frontal lobe. The advantage of this region is that the difficulty of applying electrodes correctly to the scalp and the reduced performance for individuals with long or coarse hair is almost nonexistent (Mihajlovic, Grundlehner, Vullers, & Penders, 2015), (David Hairston et al., 2014), (Ekandem, Davis, Alvarez, James, & Gilbert, 2012).

All EEG recordings are affected by muscle (frequencies > 20 Hz) (Muthukumaraswamy, 2013) and EOG artifacts (Urigüen & Begoña, 2015), (Schlögl et al., 2007). So, the innovative of this approach is that EEG electrodes in the frontal lobe are usually not taken seriously for the reasons mentioned before, even after removing the artifacts by software. But by focusing on the algorithm rather than on the features, I want to prove that classification performance can be increased without losing in usability. Therefore, the electrodes in this region are going to be tuned by performing hyperparameter optimization, and compared against the ones in the best region for the stimulus.

Many studies have focused on EEG as a potential biometric technology. Many brain stimuluses have been tested, each one with its advantages and disadvantages. VEP needs an external device to elicit ERP. Cognitive activities are not feasible for everyone; ex: Attention Deficit disorder (ADD) or handicapped patients. Meanwhile, a resting state is quite easy for everybody. Nevertheless, there are some studies which consider it a dynamic state, due to the fluctuations in attention (Gonçalves et al., 2006), (Laufs et al., 2003), (Stark & Squire, 2001) as humans cannot be just forced to think about nothing. And it is this dynamic that possibly influences the variability across subjects and therefore explains the high rates in performance. In contrast, “Sleep is a naturally recurring state of mind and body, characterized by altered consciousness, relatively inhibited sensory activity, inhibition of nearly all voluntary muscles, and reduced interactions with surroundings” (American Academy of Neurology, 2012). Then maybe with sleep states it could be answered if a real brainprint exists.

CHAPTER 3 METHODOLOGY AND MATERIALS

A full night raw database of wakefulness and sleep EEG recordings was used. Signal pre-processing techniques were applied to remove noise and create five datasets: AWA, Rem, S1, S2, SWS. The same features used in (Lajnef et al., 2015) were extracted for each of the 19 electrodes, and two additional datasets were created (SX, ALL) which are a combination of the other stages. So, 19x7 datasets were studied in this experiment. In order to test identification/authentication capabilities, all these datasets were split in Legitimates, Intruders and Hold-out sets. Assignment of subjects to these sets was based on the availability of sufficient number of epochs across the various stages. More details are given in Step 2 of Figure 3.1.

Firstly, four algorithms were used to get a preliminary classification performance: Support Vector Machine (rbf), K-nearest neighbour, Random Forest and XGBoost only on the Legitimates datasets. Secondly, a removal of epochs[4] was done to reduce training and testing duration. And, the previously mentioned algorithms were once more used to evaluate changes in performance. Finally, all algorithms were optimized and the most performing one was chosen to be tested against the Intruders and Hold-out sets.

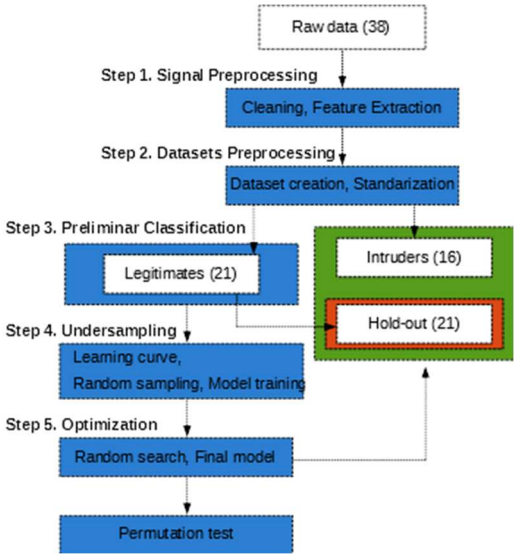


Figure 3-1 : Flowchart of biometric model

[4] Epoch: Time window extracted from the continuous EEG signal. It is how the signal has been "chopped" into segments.

Data analysis and visualization: feature extraction until permutation test were done using Python 3 and Scikit libraries. Data format conversions were performed with in-house software package for electrophysiological signal analysis (ELAN) developed at INSERM U1028, Lyon, France (Aguera, Jerbi, Caclin, & Bertrand, 2011).

3.1 Raw data

One full night of polysomnographic (PSG) recordings in 38 healthy subjects aged 29.2 +/- 8 years was collected at the DyCog Lab of the Lyon Neuroscience Research Center (Lyon, France). Each recording consists of EOG, EMG and 19 scalp EEG channels; positioned according to the international 10-20 system (See Figure 3-2 (Maskeliunas et al., 2016)). A sampling frequency of 1000Hz was used (Eichenlaub, Ruby, & Morlet, 2012). For this research, only the EEG channels were used.

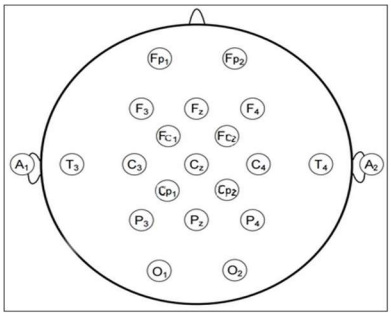


Figure 3-2 : Placement of electrodes

3.2 Biometric model steps

Step 1. Signal preprocessing
Cleaning and feature extraction (Lajnef et al., 2015)

The acquired polysomnographic signals were processed as follows:

- Filtering with cut-off frequencies at 0.2 and at 40Hz to minimize the effect of artefacts
- Segmentation into 30s epochs, according to sleep scoring standards
- Complete removal of remaining artifact contaminated segments

In total 20 features were computed from each electrode giving a matrix of Nx20, where N is the number of epochs in each class.

Relative spectral power (5), Power ratios combinations (14) and Spectral entropy were calculated from the power spectral density estimation (PSD).

Table 3-1 : Extracted features

Features used in the study		
0. relPowDelta	7. powDelta/powAlpha	14. powAlpha/powTheta
1. relPowTheta	8. powDelta/powBeta	15. powAlpha/powBeta
2. relPowAlpha	9. powDelta/powSigma	16. powAlpha/powSigma
3. relPowBeta	10. powTheta/powDelta	17. powBeta/powTheta
4. relPowSigma	11. powTheta/powAlpha	18. powBeta/powSigma
5. SpectralEntropy	12. powTheta/powBeta	19. powSigma/powTheta
6. powDelta/powTheta	13. powTheta/powSigma	

The Welch's averaged periodogram method was applied for this purpose (Oppenheim et al., 1999). So, each epoch was divided into 6 non-overlapping segments to which a Hamming window was applied. Then, by averaging these segments, the final power spectral density was obtained.

a) Relative spectral power: It represents what percentage of the signal is made up of oscillations. So, for a particular EEG band it is the absolute power divided by the sum of powers across the other frequencies bands. The absolute power is the sum of the power values within that particular band.

b) Power ratios combinations: These are the combinations of power in the EEG bands: delta (0.5 – 4.5 Hz), theta (4.5 – 8.5 Hz), alpha (8.5 – 11.5 Hz), sigma (11.5 – 15.5 Hz) and beta (15.5 – 32.5 Hz).

c) Spectral entropy: Measures data complexity or how predictable it can be in the frequency domain. It is based on the Fourier transform and calculated according to the Shannon entropy. It was computed from the relative power (Nunes, Almeida, & Sleigh, 2004), (Tsinghua University Press & Sun, 2016).

Step2. Datasets Preprocessing
Creation and standardization

The five datasets (AWA, Rem, S1, S2, SWS) were inspected and two additional datasets were created: SX, ALL. The former is the combination of all sleep stages (Rem, S1, S2, SWS); meanwhile the latter is the combination of all sleep stages with wakefulness (AWA, Rem, S1, S2, SWS).

According to the number of epochs some subjects were completely removed meanwhile others were chosen to become part of the Legitimates and the Intruders sets.

Subjects in the Legitimates and Intruders sets, were not chosen randomly. Assignment was based on the availability of sufficient number of epochs across the various stages. To determine who was going to belong to the Legitimates set: a 90-10% sss cross validation ratio was evaluated with different combinations in number of subjects. At the end, a balance between the number of subjects and algorithm performance was found with 21 subjects with over 44 epochs in each stage. The rest (16 subjects) belongs to the Intruders set. "More subjects mean lower performance (more classes makes the classification problem more difficult). Less subjects mean higher performance (the classification problem is easier but the biometric system is not "real")."

To see the total number of available epochs of these 2 sets, see Table 3.2.

- **Intruders (16 subjects):** The Intruders were never used during training, as they do not have permission to access the system.
- **Legitimates (21 subjects):** The Legitimates have access to the system. So they were used to train the model. From this set, 1 additional subset was created: Hold-out.
    A. **Hold-out:** To create it, 4 epochs were randomly[5] removed from each subject. This subset was used to test the generalization capabilities of the final model; and therefore it had never been used during the model training.

[5] Random removal: The ideal way is to use data from the same subject on a different day and training session to take into account the inter-session variability. As the closer the epochs in time, the more similar they are. However, this does not apply to the Intruders as these subjects do not have access to the system.

The Hold-out and Intruders sets were never used during training and therefore they were used for benchmarking.

Distance and margin based classifiers, such as KNN and SVM respectively, are dependent on feature standardization. Meanwhile tree-based ones, such as RF and XG aren't. But for model comparisons a robust scaler was applied on all datasets.

The robust scaler ensures that outliers do not affect the data standardization. So instead of using the mean and variance; it uses the median and the interquartile range (IQR) on each feature.

The IQR is a measure of variability based on dividing a dataset into four equal parts ("RobustScaler," n.d.).

Step 3. Preliminary Classification

This step is applied only to the Legitimates set.

Even though the datasets of the legitimate users had different number of epochs per subject and stage; 4 algorithms (SVM-rbf, KNN, RF and XG) were tested to get a preliminary performance estimation. As Accuracy can be misleading in datasets with large class imbalance (Machine Learning Mastery, n.d.), additional metrics were used: Precision, Recall and F1-score.

A 10 repetitions of 10 (10x10) stratified shuffled split (sss) cross validation was used to partition each dataset into a 90%-10% ratio for training and testing. This type of cross validation guarantees that classes are uniformly distributed in each partition and that the epochs from each one are not sorted.

Table 3-2 : Number of epochs in each dataset

LEGITIMATES

	AWA	Rem	S1	S2	SWS	SX	ALL
L1=s1	139	70	167	260	153	650	789
L2=s2	199	134	84	218	139	575	774
L3=s3	80	200	63	326	210	799	879
L4=s5	88	197	60	346	219	822	910
L5=s6	53	124	49	293	247	713	766
L6=s7	183	147	64	230	221	662	845
L7=s8	186	129	44	168	305	646	832
L8=s9	60	214	119	346	128	807	867
L9=s10	52	177	60	320	190	747	799
L10=s12	96	96	166	340	185	787	883
L11=s13	73	177	93	140	334	744	817
L12=s15	67	197	54	210	293	754	821
L13=s16	50	154	56	345	184	739	789
L14=s17	117	156	102	308	179	745	862
L15=s20	48	189	47	299	267	802	850
L16=s22	90	222	62	180	243	707	797
L17=s24	124	150	349	277	194	970	1094
L18=s27	139	61	84	112	296	553	692
L19=s31	134	152	61	353	315	881	1015
L20=s35	53	159	108	348	164	779	832
L21=s37	76	73	57	302	202	634	710
xStage	2107	3178	1949	5721	4668	15516	17623

INTRUDERS

	AWA	REM	S1	S2	SWS	SX	ALL
I1=s4	40	144	37	232	207	620	660
I2=s11	20	125	114	342	306	887	907
I3=s14	24	182	46	270	209	707	731
I4=s18	16	181	31	262	291	765	781
I5=s19	37	176	19	191	277	663	700
I6=s21	29	250	35	335	218	838	867
I7=s23	20	214	16	201	335	766	786
I8=s25	35	238	52	352	243	885	920
I9=s26	39	170	60	354	230	814	853
I10=s28	13	156	11	316	327	810	823
I11=s29	10	243	13	327	335	918	928
I12=s30	24	210	39	285	343	877	901
I13=s32	21	159	91	353	198	801	822
I14=s33	109	206	21	330	202	759	868
I15=s34	16	187	54	354	256	851	867
I16=s38	63	129	*27*	348	231	735	798
xStage	516	2970	666	4852	4208	12696	13212

Step 4. Undersampling

Learning curve[6], Random sampling and Model training

A stratified random sampling was used. It consists on randomly removing the same number of epochs from each subject. This is important not only to balance each dataset and make fair comparisons among subjects and stages, but also to reduce training duration (enrollment).

To determine the minimum number of epochs for training, a learning curve has been plotted. This type of plot is an asymptotic analysis to describe limiting behaviour. It shows how better the model becomes as the number of epochs increases in both training and testing sets while keeping the 90-10% ratio and the 10x10 sss cross validation.

A limitation in the datasets was the different number of epochs in each sleep stage and subject. So the learning curve was also limited to both of them. AWA and S1 datasets are the ones with the fewest number of epochs.

Many partitions' ratios for training and testing were evaluated: 90%-10%, 80%-20% and 70%-30%; even a One-sample-out cross validation. In order to simplify the biometric system and give every subject and stage the same chance while being classified, the total number of epochs for a ratio of 90%-10% must be a multiple of 10 to ensure that each subject will have the same number of epochs during training and testing for each stage. For example, in Table 3.3 when 38 epochs per subject are used this makes: 718 epochs for training and 80 epochs for testing. So, 17 subjects will be tested with 4 epochs meanwhile 4 subjects with only 3 (17x4 + 4x3 = 80). This does not happen if 40 epochs per subject are used (21x4=84).

So, a good balance was found in 40 epochs: 36 for training and 4 for testing.

- Legitimates: 40 random epochs x 21 subjects = 840 for each stage
- Intruders: 4 random epochs x 16 subjects = 64 for each stage
- Hold-out: 4 random epochs x 21 subjects = 84 for each stage (same size as the test set during cross validation)

---

[6] Learning curve: In machine learning it is useful for many purposes. For example: adjusting optimization to improve convergence or comparing different algorithms. In this research, it has been used to determine the amount of data used for training and testing.

And once more, the previous algorithms were tested to check how performance was affected with a reduced number of epochs.

Table 3-3 : Number of epochs in learning curve – ratio: 90% - 10%

LEARNING CURVE: SUBJECTS=21, RATIO=90%-10%			
# epochs x subject	Total	Train	Test
10*	210 (10)	189 (9)	21 (1)
12	252	226	26
14	294	264	30
16	336	302	34
18	378	340	38
20*	420 (20)	378 (18)	42 (2)
22	462	415	47
24	504	453	51
26	546	491	55
28	588	529	59
30*	630 (30)	567 (27)	63 (3)
32	672	604	68
34	714	642	72
36	756	680	76
38	798	718	80
40*	840 (40)	756 (36)	84 (4)

Table 3-4 : Number of epochs after undersampling

	LEGITIMATES	INTRUDERS	HOLD-OUT
AWA	840	64	84
Rem	840	64	84
S1	840	64	84
S2	840	64	84
SWS	840	64	84
SX	3360	256	336
ALL	4200	320	420

Step 5. Optimization

As performance was reduced after undersampling, within all algorithms, a hyperparameter optimization was used to increase it. The hyperparameters are the parameters that cannot be directly learned from the dataset and therefore require fixing before model training to control model complexity.

A random search (RS) strategy was used, as according to (Bergstra & Yoshua, 2012), it is more efficient at finding better models within a small fraction of the computation time, than trials on a grid search. It works by sampling from a distribution of possible parameter values, where the number of sampled candidates is specified. The higher this number is, the finer the search becomes.

The identification of the best hyperparameters was performed at each fold using a nested cross validation technique, to avoid the risk of leaking information to the algorithm and therefore overfitting. So, in the inner loop the best hyperparameters combination was chosen, and in the outer loop its performance was evaluated. To partition the datasets, a stratified shuffled split cross-validation method was used.

19x7 final models (one for each electrode and from each dataset) were created. They were used to determine the best sleep stage, best electrode for each stage, the features importance, to make a topographic map of the brain and to test performance against the benchmark sets.

a) Support Vector Machine (SVM): SVM searches for a hyperplane to separate different classes. This hyperplane has to maximize the margin between different classes to ensure generalization capabilities. Depending on the training set distribution, a non-linear hyperplane might be needed. For this effect a kernel function is required. Some popular examples: polynomial, radial basis function, sigmoid (C. W. Hsu, Chang, & Lin, 2016).

As an rbf kernel has been used, the hyperparameters tuned were:

- C: Penalty parameter. It determines the size of the margin.
- gamma: Kernel coefficient. It determines how wiggly the decision boundary becomes. By taking into account the influence of a single sample (far or close).

b) K-nearest neighbor (KNN): KNN defines the class of an epoch by the majority vote of its closest neighbors (Kotsiantis et al., 2006).

The tuned hyperparameters were:

- metric: it quantifies the similarity (distance) among epochs to define the belonging to a neighbourhood. Some examples: Euclidean distance, Minkowski distance, Manhattan distance, Cosine similarity and Correlation distance.
- n_neighbours: number of neighbour epochs required to vote the class membership of another epoch.
- Weight: assigns a weight to the epochs in a neighbourhood. The closer the epochs, the higher their weights.

A coarse random search of 25 parameters combinations was done, as a preliminary optimization step, on the ALL datasets using a 5x5 nested stratified shuffle split. Then a finer search of 20 parameters combinations on all the datasets was applied.

c) Random Forest (RF): It is a form of ensemble learning. It produces many trees based on a random selection of epochs and random selection of features. Then it predicts classes based on the majority vote of all trees.

Usually it is thought that RF does not need hyperparameter optimization, but it does according to (B. F. F. Huang & Boutros, 2016).

The tuned hyperparameters were:

- n_estimators: This is the number of trees to be used for the prediction of classes. The bigger the number the more stable the prediction becomes. But the computation time increases.
- max_features: This is the number of random features used to split each tree.

  According to (Breiman, 2001), (Breiman & Cutler, n.d.), it is recommended to use the Out-of-bag score (OOB) to find an optimal range for both hyperparameters. The OOB score is a fast and accurate estimation of the classifier performance without the need of an independent test set.

  So an OOB score calculation with the ALL datasets was done to estimate the range of these two hyperparameters. Then a finer random search of 90 parameters combinations, using these ranges, was done on all the datasets.

d) XGBoost (XG): It is an advanced Gradient Boosting classifier proposed by (Chen & He, 2015). It is very popular among winning solutions in Kaggle's machine learning competitions. As it focuses on computational speed and model performance. It builds models sequentially. The new created models can correct the errors of prior models and then all of them are combined together into a strong one to make a final prediction.

The tuned hyperparameters were:

- n_estimators: it is the number of trees.
- max_depth: it is the pruning of each tree. Unlike RF, in XG large trees easily overfit.
- learning_rate: Makes the model more robust when generalizing. It reduces the weights at each iteration.
- colsample_bytree: This is similar to max_features in RF. It chooses randomly a number of features.

As the trees are correlated, the OOB error is not a reliable estimation to tune hyperparameters in XG ("XGBoost - Add Out-of-Bag performance validation #1070," n.d.).

So, a coarse random search of 20 parameters combinations was done, as a preliminary optimization step, on the ALL datasets using a 5x5 nested stratified shuffle split. Then a finer search of 90 parameters combinations on all the datasets was applied.

Table 3-5 : Searching space of hyperparameters optimization

Model	Parameter	Preliminar RS	Final RS
SVM (rbf)	C (Penalty)	e-10 -> e10	e2 -> e6
	Gamma	e-10 -> e3	e-3 -> 1

Model	Parameter	Preliminar RS	Final RS
Knn	n_neighbors	2, 3, 5, 7, 10	2 -> 6
	weights	distance	distance
	metric	minkowski, euclidean, manhattan, cosine, correlation	euclidean, manhattan, cosine

Model	Parameter	Preliminar RS	Final RS
Random Forest	n_estimators	10 -> 400	170 -> 300
	max_features	2, 5, 10, 15, 20	5 -> 20

Model	Parameter	Preliminar RS	Final RS
XGBoost	n_estimators	100 -> 1000	300 -> 700
	max_depth	3 -> 10	6 -> 10
	learning_rate	e-3 -> e-1	e-2 -> e-1
	Colsample_bytree	0.5 ->1	0.5 ->1

Step 6. Permutation

In order to know if the best performing algorithm was appropriate for the classification task (if the classification score was significant) a standard Permutation test was used (Good, 1994).

In this test, the null hypothesis assumes that the features and the labels are independent $p(X,y)=p(X)p(y)$. To evaluate this hypothesis, the labels were permuted 1000 times; and for each permutation the classification pipeline was executed again.

The obtained *p-value* is the percentage of runs for which the permutation score was greater than the classification score obtained in the first place. As it was lower than 1/1000, the null hypothesis was rejected. This means that features and classes are dependant and that the algorithm was capable of finding a predictive structure between them.

3.3 Benchmark sets: Hold-out and Intruders

The final model obtained from the previous section was evaluated on dedicated evaluation sets that were not used during training: Hold-out and Intruders, to check generalization capabilities. For this purpose, the confusion matrix was used to estimate different metrics: Accuracy, precision, recall, f1-scores, error type I, error type II and HTER (Half Total Error Rate).

In biometrics, specific sleep stages are not relevant. So, for this reason the AWA, SX and ALL datasets were the only ones used to evaluate performance.

The Hold-out set has the same size as the test set during cross validation: 4 epochs per subject.

Other projects have not implemented the continuous identification/authentication; which is necessary to avoid unauthorized post-identification/authentication use of a telemedicine system. So, subjects must be able to be recognized no matter their current state. For this research: awoke (AWA) or aslept (SX).

For this purpose, the following combinations were done:

- Train: All Fp1, Test: AWA FP1
- Train: SX Fp1, Test: AWA Fp1

CHAPTER 4 RESULTS AND DISCUSSION

4.1 Preliminary classification

This is step 3 in Figure 3.1.

The 4 algorithms: SVM (rbf), KNN, RF and XG were used to get a preliminary performance on the Legitimates set for each electrode. A 10x10 sss cross-validation was used to partition each dataset into a 90% - 10% ratio for training and testing. As each dataset has different number of epochs[7] per each subject (class imbalance), 4 different metrics were used: accuracy, precision, recall and f1-score to ensure performance results.

Tables 4.1, 4.2, 4.3 and 4.4 show the highest accuracy value reached for each electrode (Max_acc); and the average across all the 19 electrodes for each stage (Mean_Acc).

[7] Epoch: Time window extracted from the continuous EEG signal. It is how the signal has been "chopped" into segments.

4.1.1 SVM (rbf)

With this algorithm it can be seen that S1 is the most discriminative stage and that Fp1, Pz, C4, Fz, F4, P3 and Cz are the most performing electrodes. In Figure 4.1, it can be seen that the class imbalance is not a problem for the classification.

Table 4-1 : Preliminary performance: Best electrode and Best stage – SVM (rbf)

Stage	Elec	Max_Acc
S1	Fp1	93.1
Rem	Pz	92.62
AWA	C4	92.36
S2	Fz	83.9
SWS	F4	83.42
SX	P3	64.05
ALL	Cz	61.3

Stage	Mean_Acc
Rem	88.42
S1	87.36
AWA	85.76
SWS	80.56
S2	80.32
SX	61.96
ALL	58.91

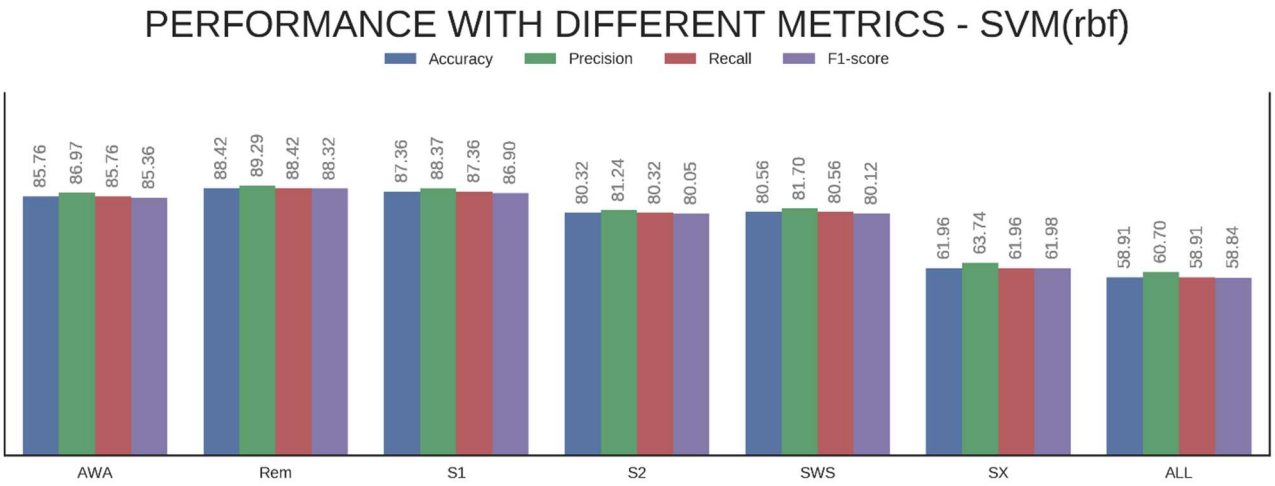


Figure 4-1 : Performance with different metrics (using Mean_Acc) – SVM (rbf)

4.1.2 KNN

With this algorithm it can be seen that Rem is the most discriminative stage and that Pz, C4, Fp1, P3 and Cz are the most performing electrodes.

In Figure 4.2, it can be seen that the class imbalance is not a problem for the classification.

Table 4-2 : Preliminary performance: Best electrode and Best stage - KNN

Stage	Elec	Max_Acc
Rem	Pz	93,68
AWA	C4	92,02
S1	Fp1	91,86
SWS	P3	87,7
S2	P3	86,86
SX	P3	77,31
ALL	Cz	74,53

Stage	Mean_Acc
Rem	90,38
AWA	86,25
S1	86,22
SWS	86,14
S2	84,5
SX	74,82
ALL	72,41

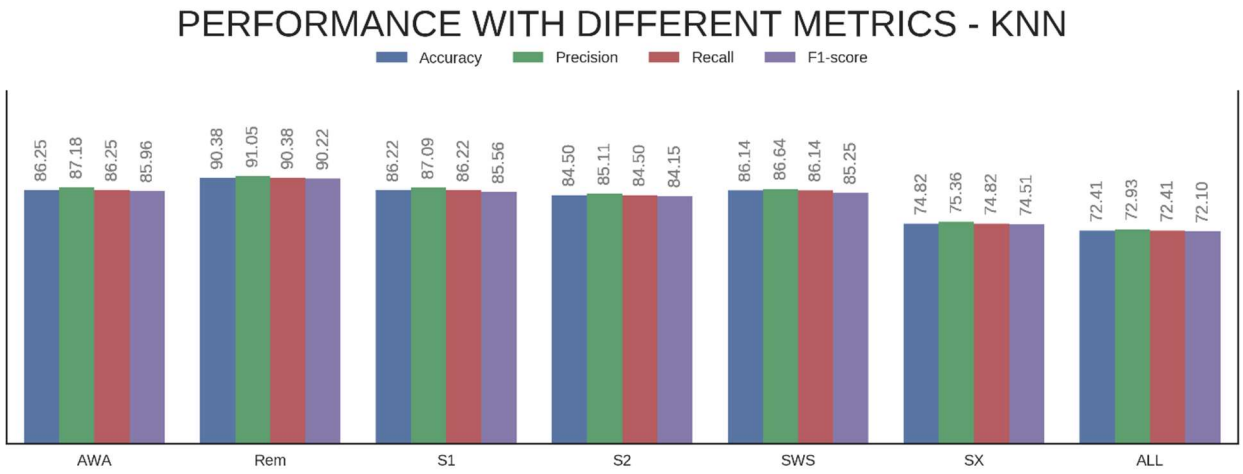


Figure 4-2 : Performance with different metrics (using Mean_Acc) – KNN

4.1.3 RF

With this algorithm it can be seen that Rem is the most discriminative stage and that T3, C4, Fz, P4 and P3 are the most performing electrodes.

In Figure 4.3, it can be seen that the class imbalance is not a problem for the classification.

Table 4-3 : Preliminary performance: Best electrode and Best stage - RF

Stage	Elec	Max_Acc
Rem	T3	95,55
AWA	C4	93,64
S1	Fz	93,3
SWS	P4	90,21
S2	P3	89,03
SX	P3	79,6
ALL	P3	76,89

Stage	Mean_Acc
Rem	91,84
S1	90,28
AWA	89,38
SWS	88,6
S2	86,21
SX	77,69
ALL	75,55

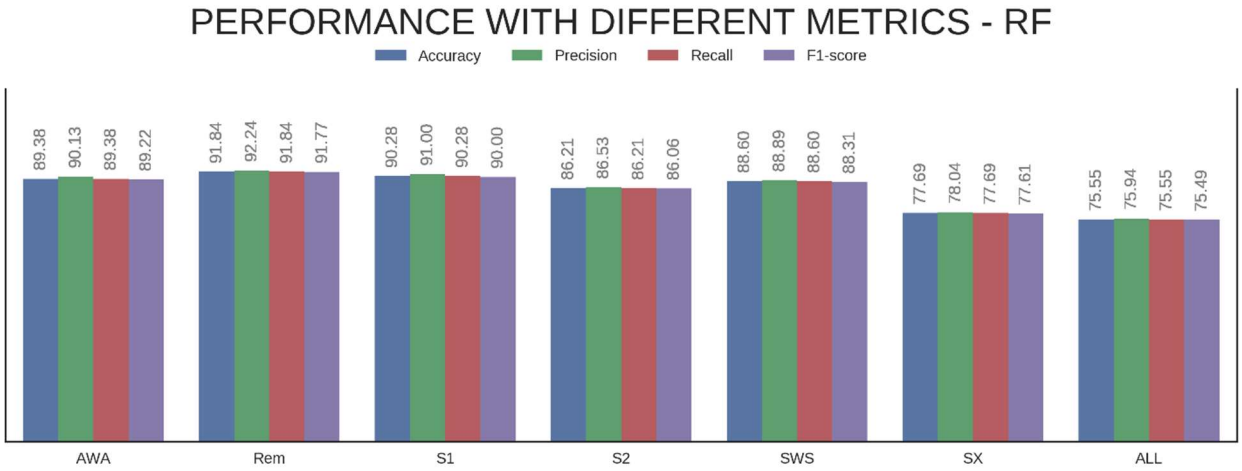


Figure 4-3 : Performance with different metrics (using Mean_Acc) – RF

4.1.4 XG

With this algorithm it can be seen that Rem is the most discriminative stage and that Pz, Fz, C4, P3, and T3 are the most performing electrodes.

In Figure 4.4, it can be seen that the class imbalance is not a problem for the classification.

Table 4-4 : Preliminary performance: Best electrode and Best stage - XG

Stage	Elec	Max_Acc
Rem	Pz	95,7
S1	Fz	95,13
AWA	C4	94,21
S2	Fz	87,58
SWS	P3	87,35
SX	P3	65,23
ALL	T3	61,46

Stage	Mean_Acc
Rem	92,73
S1	92,22
AWA	91,07
SWS	85,53
S2	84,01
SX	62,44
ALL	59,07

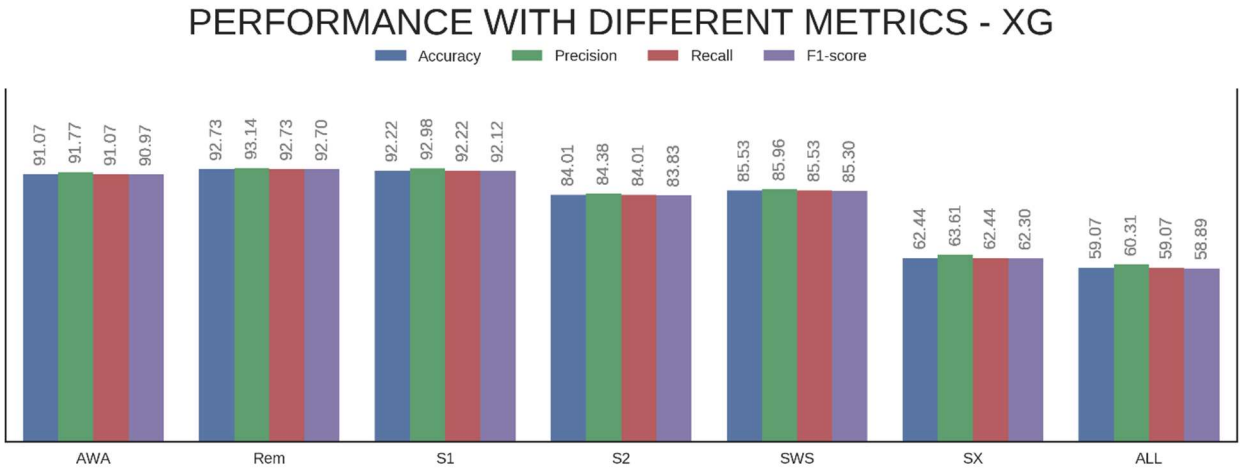


Figure 4-4 : Performance with different metrics (using Mean_Acc) - XG

4.1.5 Best algorithm and Best Electrode

The most performing algorithms is RF in almost all stages. Also, RF and KNN are the most stable across all stages; as SVM (rbf) and XG decrease abruptly in stages SX and ALL. It can be concluded from Figure 4.5 that RF is the most performing and the most stable of all algorithms. And that the best electrode is Pz with XG algorithm in Rem stage.

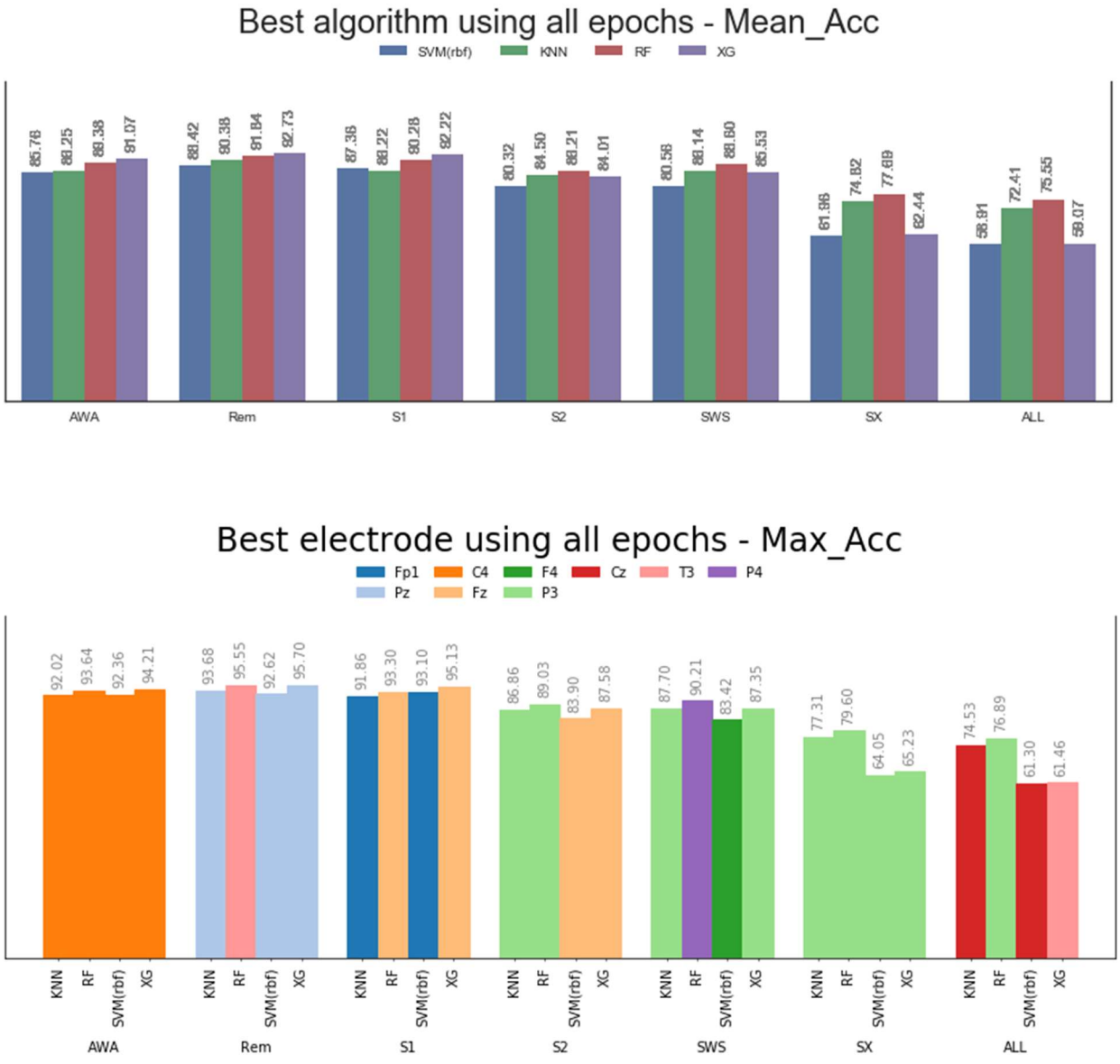


Figure 4-5 : a) Best algorithm – Mean_Acc and b) Best electrode – Max_Acc; preliminary performance (all epochs)

4.2 Undersampling

This is step 4 in Figure 3.1.

As RF was the most performing and more stable algorithm in the previous step; it was used to plot a learning curve with electrodes Fp1 and Fp2[8]. This type of plot shows how better the model becomes as the number of epochs increases in both training and testing sets while keeping the 90-10% ratio and the 10x10 sss cross validation. Accuracy was used as the performance metric as the number of epochs per subject in each dataset from now on will always be balanced.

Tables 4.5, 4.6, 4.7 and 4.8 show the highest accuracy value reached for each electrode (Max_acc); and the average across all the 19 electrodes for each stage (Mean_Acc).

4.2.1 Learning curve

AWA and S1 datasets are the ones with the fewest number of epochs. So the maximum number of epochs to keep in each dataset was limited to these two. In Figures 4.6 and 4.7, it can be seen that accuracy increases as the number of epochs does. Also, that the most performing stages are S1, AWA and Rem. Meanwhile the worst ones are SX and ALL. The best performance is reached with 38 and 40 epochs per subject. But, in order to give every subject and stage the same chance while being classified, a good balance was found with 40 epochs: 36 for training (1080 s) and 4 for testing (120 s).

[8]The choice of Fp1 and Fp2 has already been discussed in Chapter 2.

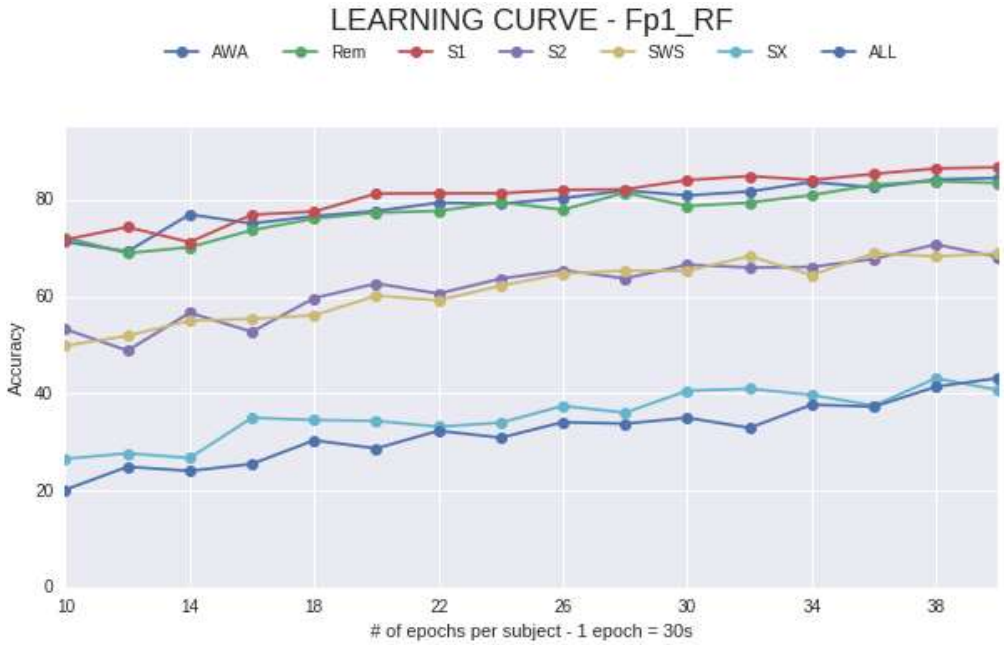


Figure 4-6: Learning curve – Fp1_RF

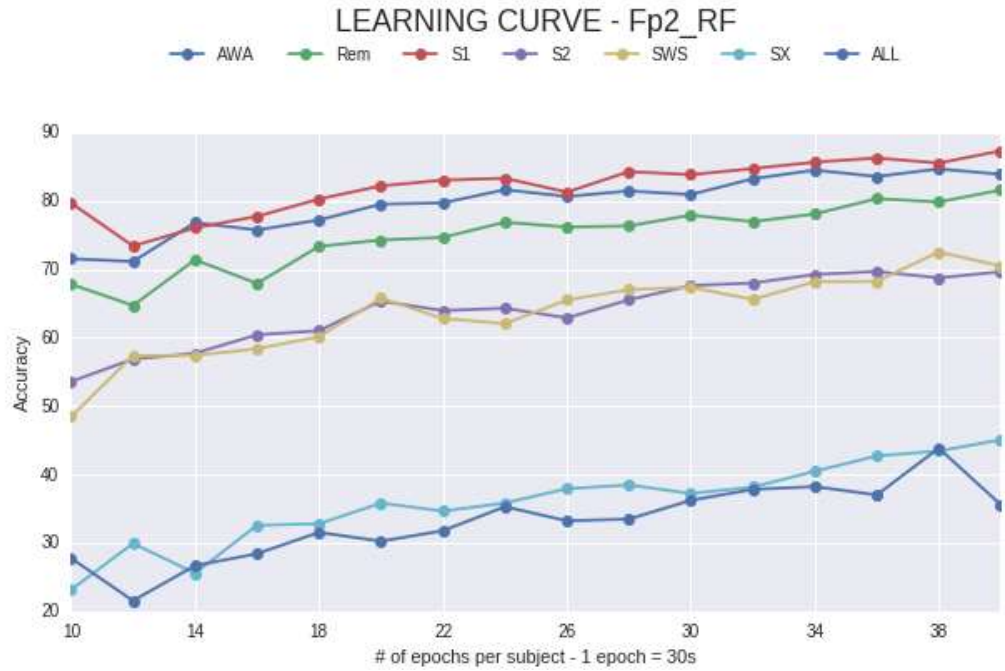


Figure 4-7: Learning curve – Fp2_RF

4.2.2 SVM (rbf)

With this algorithm it can be seen that Rem is now the most discriminative stage (in the preliminary classification it was S1) and that Pz, Fz, C4, P4, Cz and C3 are the most performing electrodes.

Table 4-5: Performance after undersampling: Best electrode and Best stage – SVM (rbf)

Stage	Elec	Max_Acc
Rem	Pz	87,48
S1	Fz	87,13
AWA	C4	86,89
S2	P4	68,67
SWS	Cz	67,21
SX	C3	52,55
ALL	Cz	49,86

Stage	Mean_Acc
Rem	80,95
S1	80,63
AWA	80,31
S2	63,92
SWS	62,48
SX	49,07
ALL	46,98

4.2.3 KNN

With this algorithm it can be seen that AWA is the most discriminative stage and that C4, Pz, Fz, O2, and Cz are the most performing electrodes.

Table 4-6: Performance after undersampling: Best electrode and Best stage – KNN

Stage	Elec	Max_Acc
AWA	C4	85.07
Rem	Pz	84.56
S1	Fz	82.96
SWS	O2	65.27
S2	Pz	64.23
SX	Fz	57.95
ALL	Cz	56.9

Stage	Mean_Acc
AWA	80.07
Rem	80.02
S1	78.19
SWS	61.12
S2	60.65
SX	54.57
ALL	53.62

4.2.4 RF

With this algorithm it can be seen that now AWA is the most discriminative stage (in preliminary classification it was Rem) and that T3, C4, Fz, P3 and Cz are the most performing electrodes.

Table 4-7 : Performance after undersampling: Best electrode and Best stage – RF

Stage	Elec	Max_Acc
Rem	T3	90,17
AWA	C4	88,49
S1	Fz	87,64
S2	Fz	73,05
SWS	P3	71,62
SX	Fz	63,37
ALL	Cz	62,06

Stage	Mean_Acc
AWA	84,7
Rem	84,68
S1	83,3
S2	69,14
SWS	69,12
SX	60,97
ALL	59,51

4.2.5 XG

With this algorithm it can be seen that Rem is still the most discriminative stage and that T3, C4, Fz, P4, P3 and Fz are the most performing electrodes.

Table 4-8 : Performance after undersampling: Best electrode and Best stage – XG

Stage	Elec	Max_Acc
Rem	T3	91,46
AWA	C4	90,08
S1	Fz	89,26
S2	P4	79,13
SWS	P3	73,8
SX	T3	60,88
ALL	Fz	57,78

Stage	Mean_Acc
Rem	86,82
AWA	86,7
S1	86,15
S2	73,07
SWS	71,43
SX	58,46
ALL	55,42

4.2.6 Best algorithm and Best electrode

Accuracy has decreased in all stages, and the algorithms leading to the highest accuracies were RF and XG. All algorithms show a drop in accuracy in stages SX and ALL, especially SVM (rbf) and KNN. It can be concluded from Figure 4.8 that RF is still the most performing and the most stable of all algorithms. And that the best electrode is T3 with XG algorithm in Rem stage.

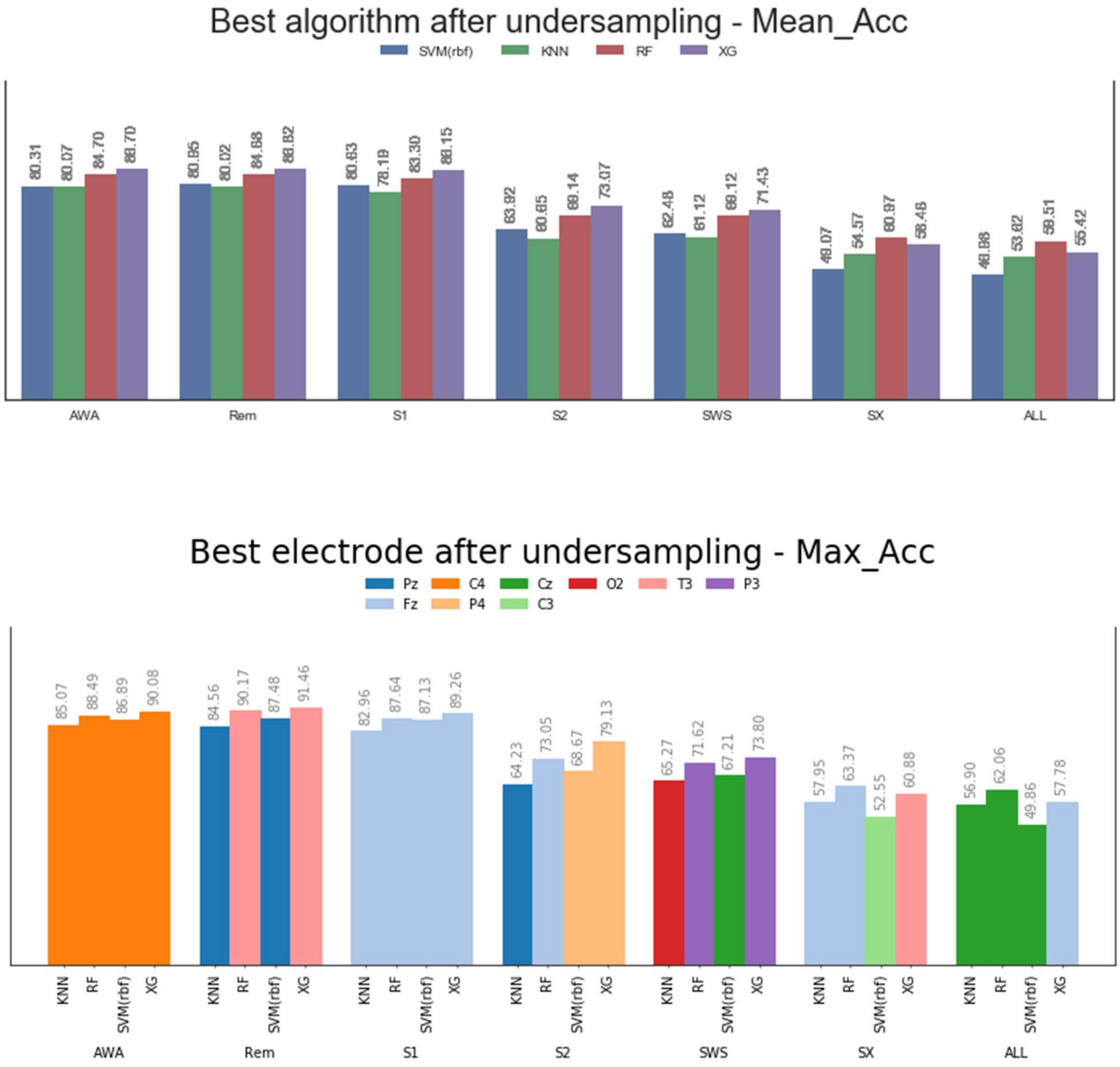


Figure 4-8 : a) Best algorithm - Mean_Acc and b) Best electrode – Max_Acc; after undersamplig (840 epochs)

4.3 Optimization

This is step 5 in Figure 3.1.

As performance decreased after undersampling, a hyperparameter optimization was used to increase it. 19x7 final models (one for each electrode and stage) were created. And the 4 algorithms: SVM (rbf), KNN, RF and XG were optimized to make a fair comparison and make a final decision to determine the best stage, the best electrode for each stage, the features importance, to make a topographic map of the brain and to test performance against the benchmark sets. A 10x10 sss cross-validation was used to partition each dataset into a 90% - 10% ratio for training and testing.

Tables 4.9, 4.10, 4.11 and 4.12 show the highest accuracy values reached for each electrode (Max_acc); and the average across all the 19 electrodes for each stage (Mean_Acc).

4.3.1 SVM (rbf)

When optimizing this algorithm, a heatmap was used to easily find a range of values to fine tune the search for better hyperparameters. The ALL datasets were used for each of the 19 electrodes. With this algorithm it can be seen that Rem is the most discriminative stage and that Pz, Fz, Fp1, F4, P4, C3 and Fz are the most performing electrodes.

Table 4-9 : Performance optimization: Best electrode and Best stage – SVM (rbf)

Stage	Elec	Max_Acc
Rem	Pz	93.57
S1	Fz	92.14
AWA	Fp1	90.6
SWS	F4	78.69
S2	P4	77.98
SX	C3	66.61
ALL	Fz	65.12

Stage	Mean_Acc
Rem	88.54
AWA	87.82
S1	87.56
SWS	74.01
S2	73.56
SX	64.08
ALL	61.43

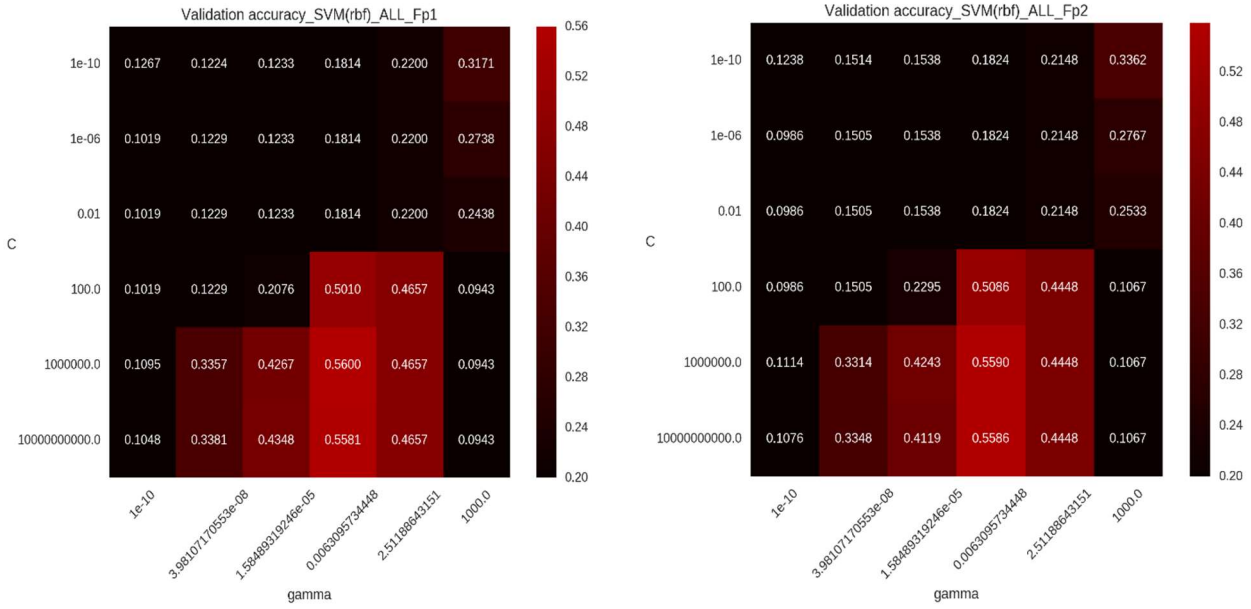


Figure 4-9 : Heatmap : Preliminary optimization on Fp1 and Fp2 – SVM (rbf)

4.3.2 KNN

When optimizing this algorithm, a heatmap was used to easily find a range of values to fine tune the search for better hyperparameters. The ALL datasets were used for each of the 19 electrodes. With this algorithm it can be seen that Rem is the most discriminative stage and that C4, Pz, Fz and O2 are the most performing electrodes.

Table 4-10 : Performance optimization: Best electrode and Best stage – KNN

Stage	Elec	Max_Acc
AWA	C4	88.81
Rem	Pz	88.57
S1	Fz	87.98
SWS	O2	74.4
S2	Pz	72.5
SX	Fz	64.97
ALL	Fz	62.31

Stage	Mean_Acc
Rem	84.59
AWA	83.34
S1	81.77
SWS	69.74
S2	66.91
SX	60.72
ALL	59.51

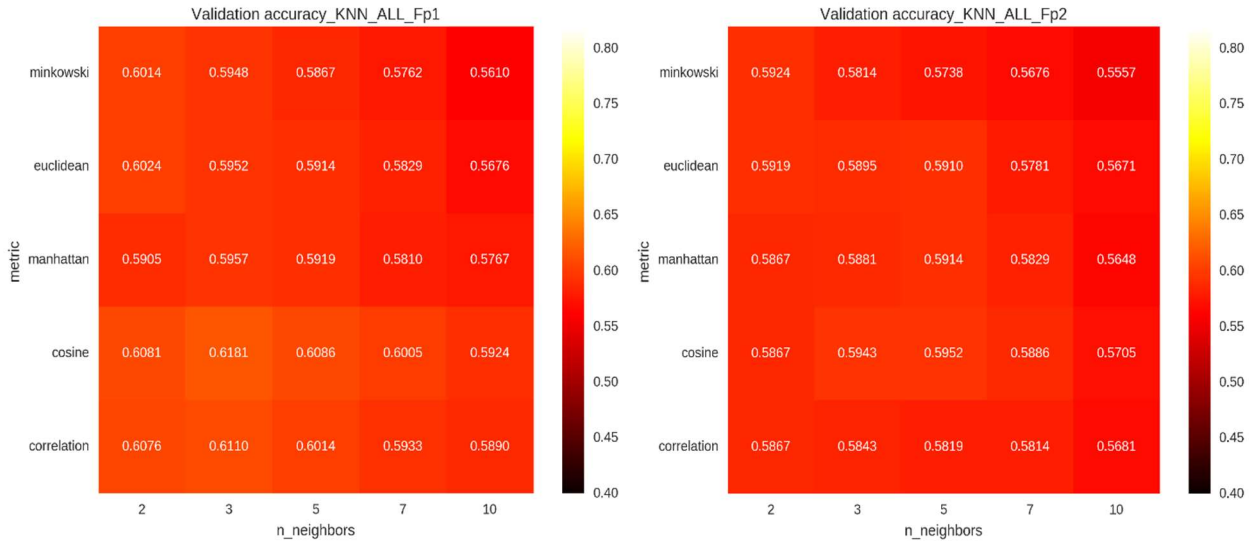


Figure 4-10 : Heatmap: Preliminary optimization on Fp1 and Fp2 - KNN

4.3.3 RF

To optimize this algorithm, the OOB error was used to easily find a range of values to fine tune the search for better hyperparameters. The ALL datasets were used for each of the 19 electrodes. With this algorithm it can be seen that Rem is once again the most discriminative stage and that C4, Pz, Fz and O2 are the most performing electrodes.

Table 4-11 : Performance optimization: Best electrode and Best stage - RF

Stage	Elec	Max_Acc
Rem	T3	94.17
AWA	C4	92.62
S1	Fp1	92.62
S2	P4	83.81
SWS	Fp2	79.76
SX	Fz	74.4
ALL	Fz	72.38

Stage	Mean_Acc
Rem	89.5
AWA	88.61
S1	88.52
S2	77.09
SWS	76.19
SX	72.03
ALL	70.34

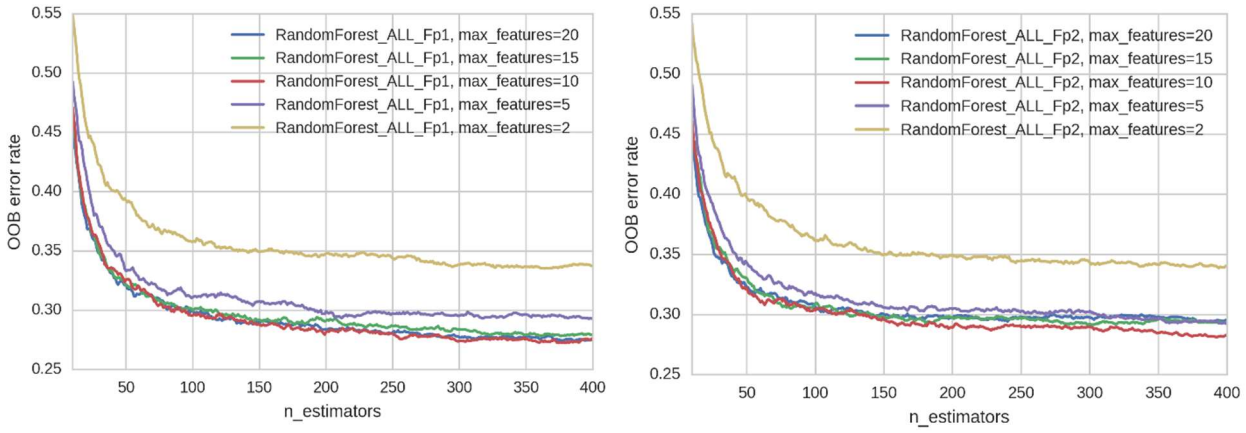


Figure 4-11 : OOB error : Preliminary optimization on Fp1 and Fp2 - RF

4.3.4 XG

When optimizing this algorithm, a report with the best parameters was used to easily find a range of values to fine tune the search for better hyperparameters. The ALL datasets were used for each of the 19 electrodes. With this algorithm it can be seen that Rem is the most discriminative stage and that Fz, P4, P3 and T3 are the most performing electrodes.

Table 4-12 : Performance optimization: Best electrode and Best stage - XG

Stage	Elec	Max_Acc
Rem	Fz	91.19
AWA	Fz	90.6
S1	Fz	90.36
S2	P4	81.67
SWS	P3	79.52
SX	T3	72.26
ALL	T3	70.67

Stage	Mean_Acc
Rem	87.71
AWA	87.41
S1	86.6
S2	74.66
SWS	73.88
SX	69.27
ALL	67.24

4.3.5 Best algorithm, Best electrode and Best stage

A standard permutation test was used to statistically assess the classification performance (Good, 1994). Labels were permuted 1000 times, and for each permutation the classification pipeline was executed. As a $p<0.001$ was obtained, this means that features and classes are dependant and that each algorithm was capable of finding a predictive structure between them for a probabilistic chance level of 1/21 x 100 = 4.76%.

From Figure 4.12, it can be concluded that Rem is the most performing stage. Also that the best algorithm in all stages is RF followed by XG, SVM (rbf) and KNN. All algorithms show a drop in accuracy in stages SX and ALL, especially SVM (rbf) and KNN.

As RF is the most performing algorithm, it will be used to evaluate the Benchmark sets.

The best electrode is T3 with RF algorithm in Rem stage.

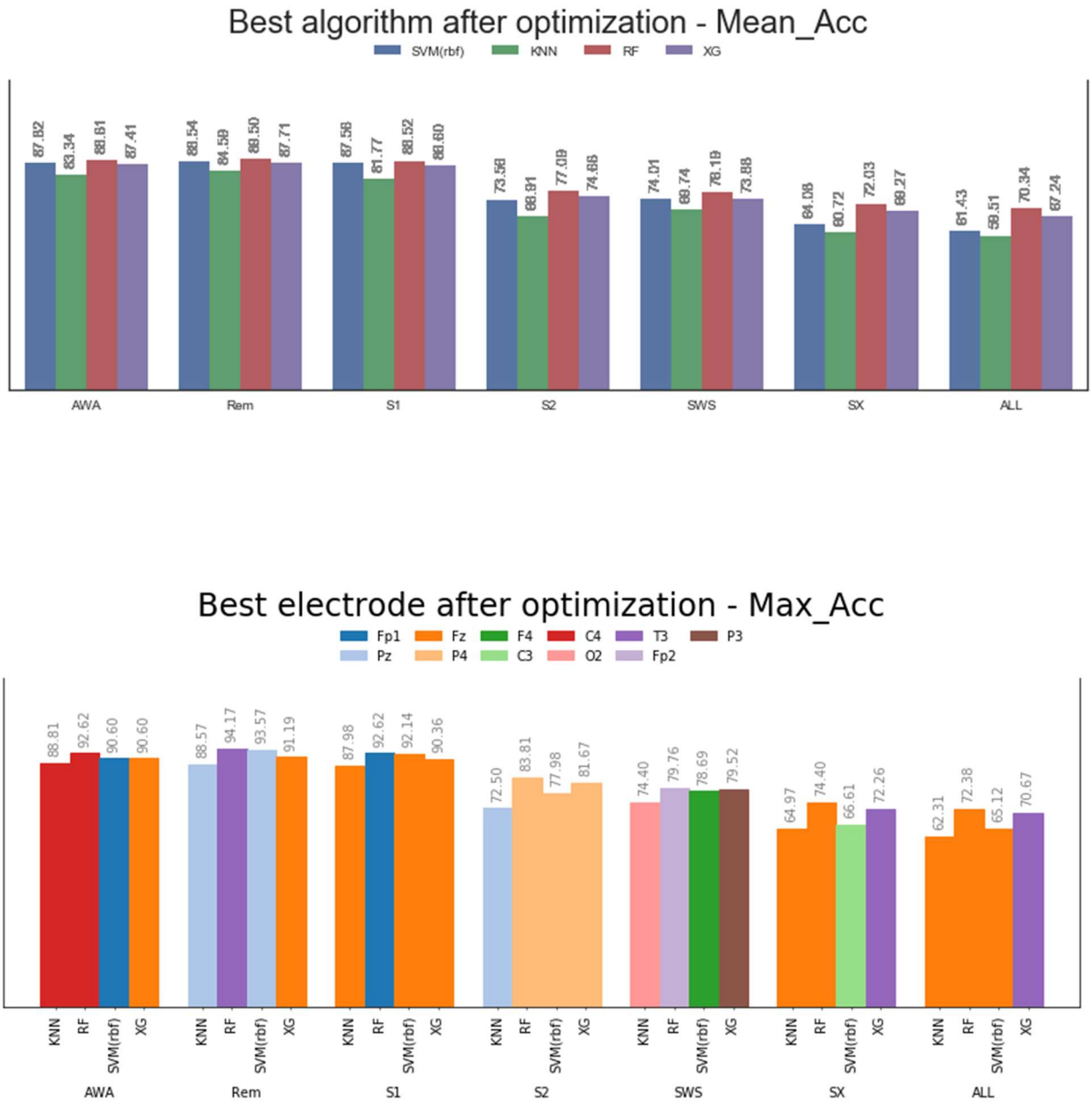


Figure 4-12 a) Best algorithm – Mean_Acc, b) Best electrode – Max_Acc – Mean_Acc; after optimization

4.3.6 Best electrode after optimization – Fp1 vs Fp2

One of the specific objectives was to know the possibility of discriminating subjects with only 1 electrode. Also it is important to make the system "usable"; so electrodes Fp1 and Fp2 were chosen. In Figure 4.13 it can be seen that Fp1 systematically provided higher decoding accuracies than Fp2 across all stages.

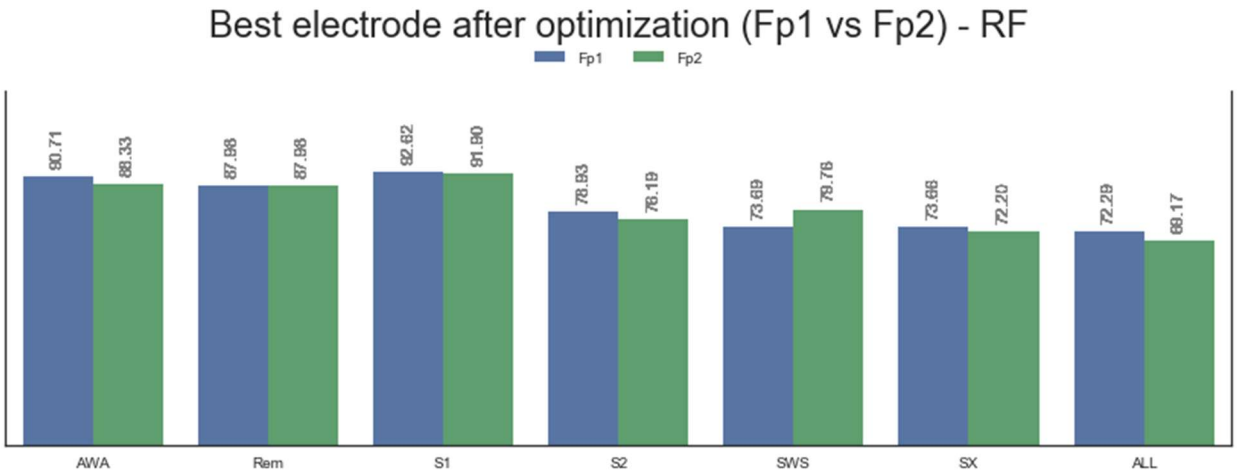


Figure 4-13 : Best electrode after optimization (Fp1 vs Fp2) – RF

4.3.7 Performance of Fp1 vs best electrode per stage

In Figures 4.14, 4.15 and 4.16, the undersampled and optimized Fp1 is compared with the best electrode for each stage in 3 situations: preliminary (All epochs), undersampling (840 epochs) and undersampling (840) + optimization; respectively.

In Figure 4.14, Fp1 performance is not the best. But it must be taken into account that there is a difference of more than the double in the number of epochs for each stage. So, exploring better techniques for undersampling would be interesting to do in the future for maximizing epoch's efficiency.

In Figure 4.15, Fp1 is better than almost all electrodes in all stages except in Rem, where T3 is more performing by 2%. This proves correct our initial assumption that an optimized frontal lobe electrode can be better than the ones in the best region for the stimulus.

In Figure 4.16, Fp1 is the worst in all stages.

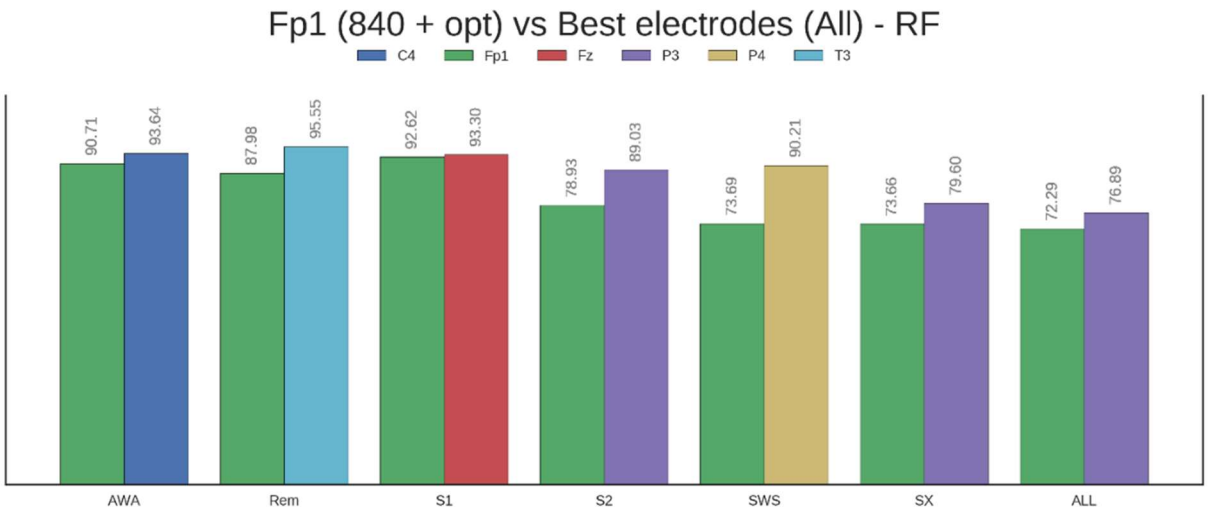


Figure 4-14 : Fp1 (840 + opt) vs Best electrodes per stage (All) - RF

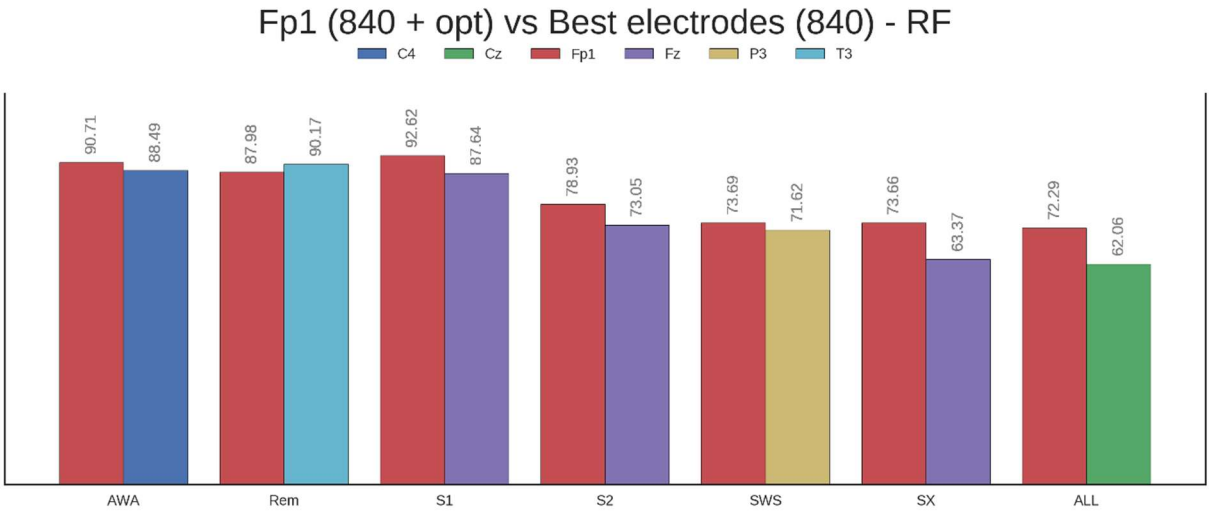


Figure 4-15 : Fp1 (840 + opt) vs Best electrodes per stage (840) – RF

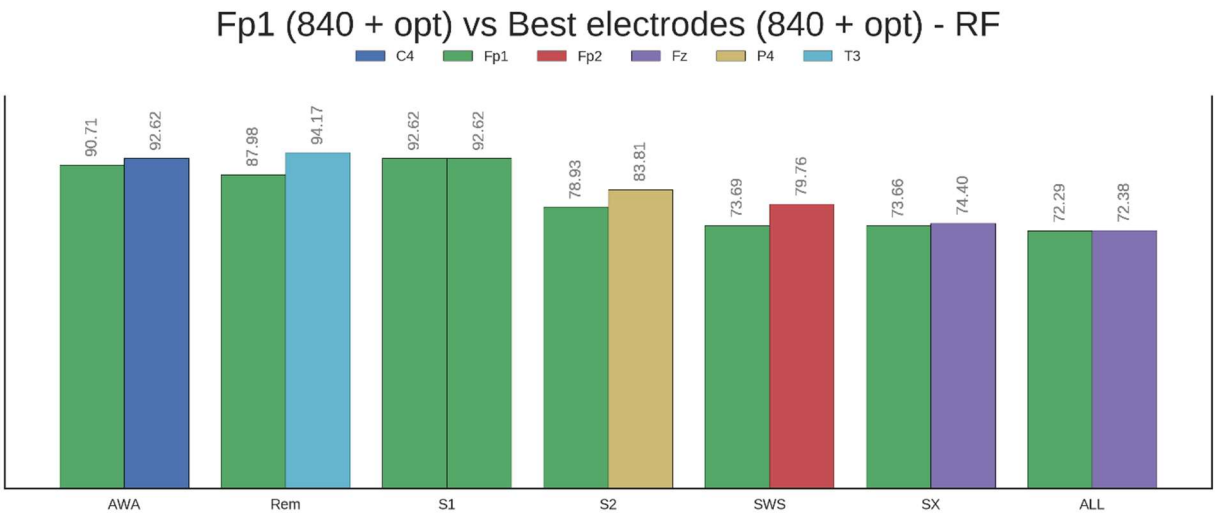


Figure 4-16 : Fp1 (840 + opt) vs Best electrodes per stage (840 + opt) – RF

4.3.8 Effect of optimization in Fp1

Figures 4.17, 4.18, 4.19 and 4.20 show the effect of steps 3, 4 and 5 of Figure 3.1 (Preliminar classification (All epochs), undersampling (840 epochs), undersamplig + optimisation (opt)) in each dataset. It can be seen that in some cases, few epochs and an optimized algorithm can be as good as or better than a lot of epochs and a non-optimized algorithm. Nevertheless, better techniques for undersampling, rather than stratified random sampling, must be explored.

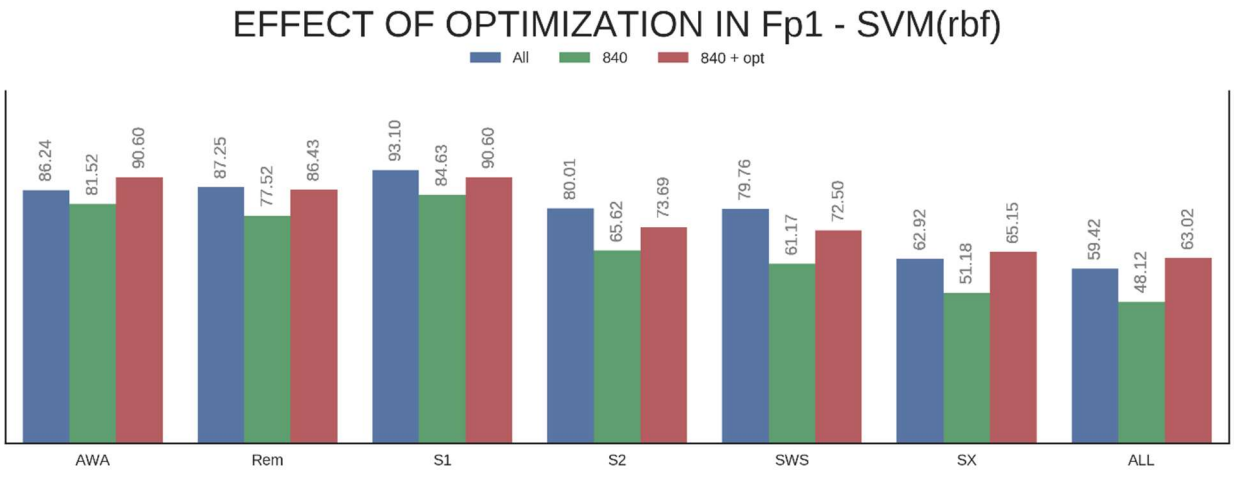


Figure 4-17 : Effect of optimization in Fp1 - SVM (rbf)

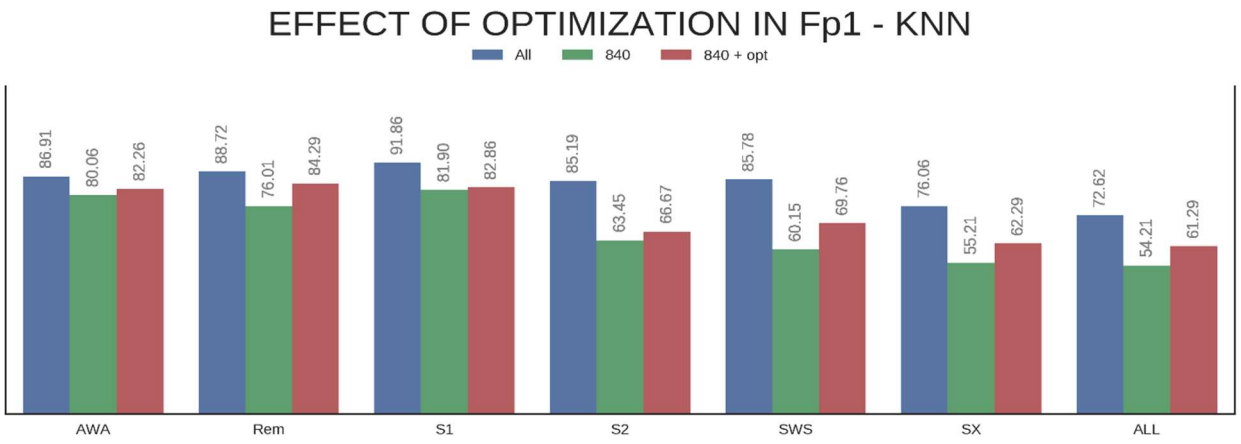


Figure 4-18 : Effect of optimization in Fp1 – KNN

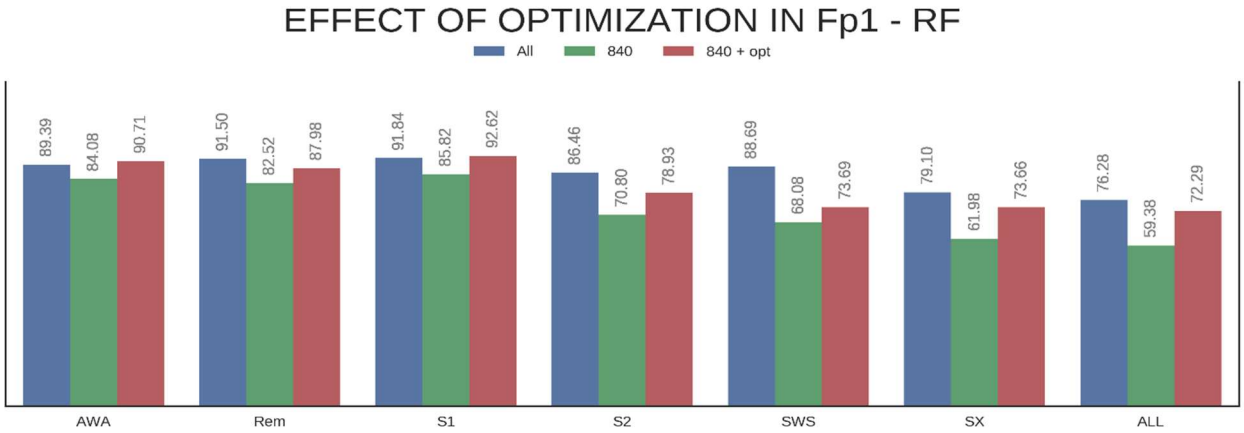


Figure 4-19 : Effect of optimization in Fp1 – RF

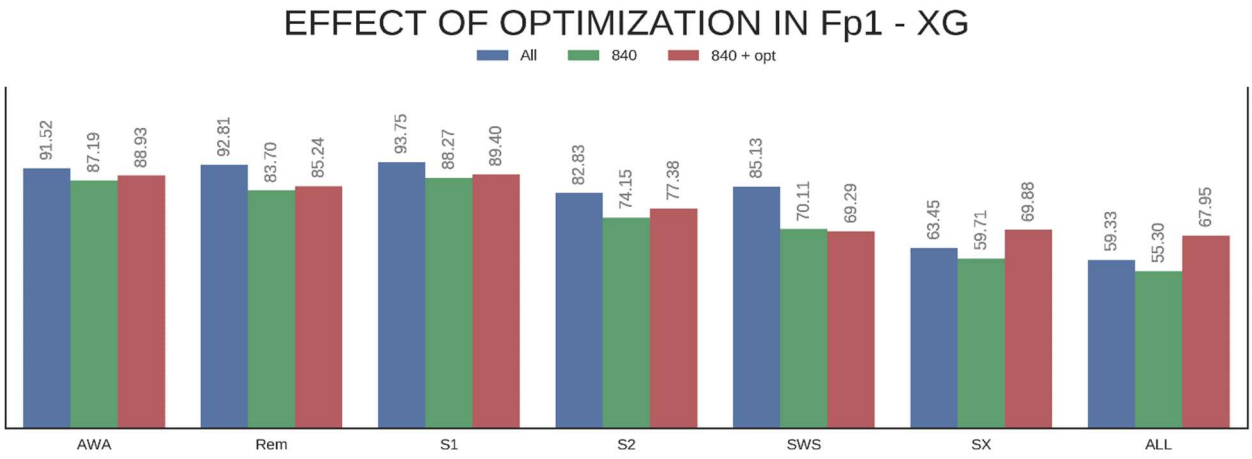


Figure 4-20 : Effect of optimization in Fp1 – XG

4.3.9 Topoplot with RF

Table 4.13 shows the performance of each electrode of each stage and Figure 4.21 displays it in a topographic map. A multifeature classification (all features together) and an optimised Random forest with a 10x10 sss cross-validation (to partition each dataset into a 90% - 10% ratio for training and testing) were used. The brighter the color the higher the accuracy for a specific electrode in a brain area.

Table 4-13 : Performance after optimization – RF

	C3	C4	CP1	CP2	Cz	F3	F4	FC1	FC2	Fp1	Fp2	Fz	O1	O2	P3	P4	Pz	T3	T4
AWA	90.36	92.62	88.45	88.69	86.79	85.71	86	86.6	89.4	90.71	88.3	90.83	89.52	89.4	87.62	86.43	89.52	87.86	88.81
Rem	90	86.79	91.43	90.48	91.9	86.9	82.4	92	89.64	87.98	88	91.79	88.93	92.4	89.52	85.95	92.14	94.17	88.21
S1	88.69	87.86	86.79	88.45	90.71	89.17	85.2	89.8	85.36	92.62	91.9	91.19	87.02	85.4	87.14	89.52	89.29	88.1	87.62
S2	73.93	77.5	77.74	69.29	75	76.55	75.1	76.7	75.71	78.93	76.2	79.17	76.31	77.9	79.52	83.81	82.38	80.71	72.38
SWS	76.07	75.83	71.07	77.74	74.88	78.1	78	77.7	74.4	73.69	79.8	72.02	77.74	78.9	78.21	78.81	71.19	73.81	79.64
SX	71.93	69.26	71.85	69.94	73.51	72.53	69.5	73.9	71.28	73.66	72.2	74.4	73.04	72.2	72.74	71.4	71.43	73.75	70.09
ALL	71.95	69.71	71.62	69.98	71.43	70.88	69	69.8	68.4	72.29	69.2	72.38	69.74	70.6	71.38	68.21	70.33	71.93	67.69

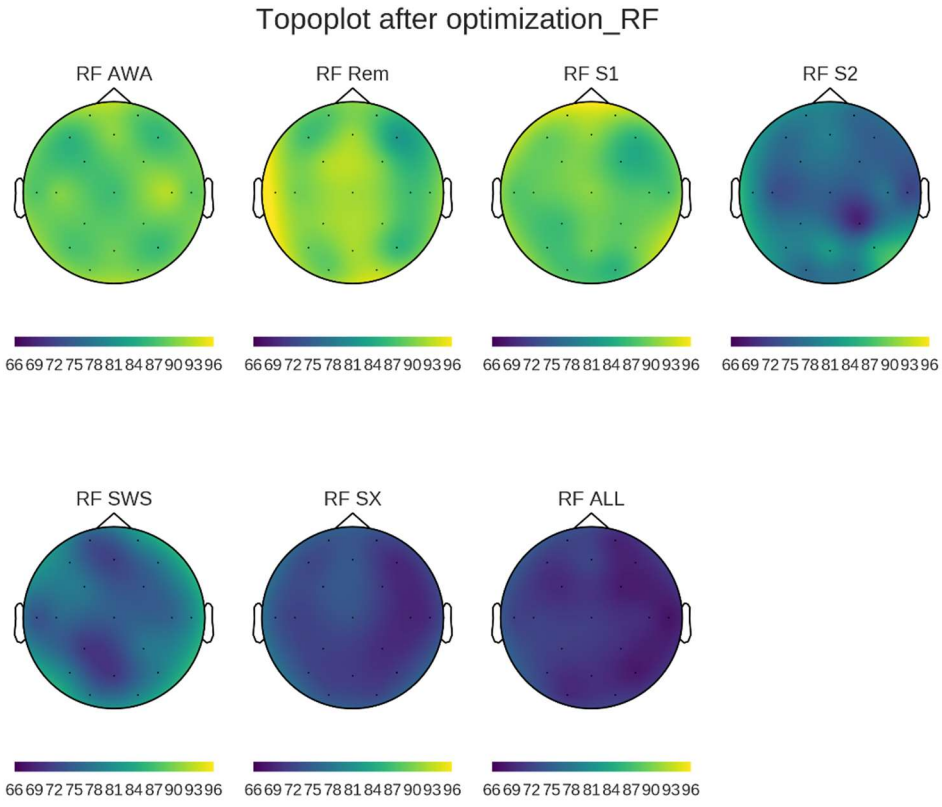


Figure 4-21 : Topoplot after optimization - RF

4.3.10 Features importance with RF

Appendix D, E and F show the importance of each feature (single feature classification) for each electrode of datasets AWA, SX and ALL; using the best algorithm (Random Forest) after optimization. The feature with the highest performance is the most relevant and therefore the one that is the most capable to discriminate among subjects. It can be seen that feature 5: Spectral entropy is the most important one. Electrodes: O1, C3 and CP2 have the highest values of it for each stage respectively.

4.4 Benchmark sets: Hold-out, Intruders

These sets were never used during training, which make them ideal to check the generalization capabilities of our model (is there overfitting or underfitting?). All their epochs were chosen randomly. For each of these ones each subject has 4 trials (4 epochs of 30s each) to try to access the system.

In biometrics, specific sleep stages are not relevant. And as continuous identification/authentication happens in periods of time, it is important to recognize subjects in different states: either aslept or awoke for this research. Therefore, only AWA, SX and ALL datasets were evaluated. The confusion matrix has been used to display the acceptance or rejection of subjects.

Two types of techniques were used: a Multiclass classification and a One versus all classification. In the multiclass classification technique, epochs are classified into one of the 21 subjects. Meanwhile in the One versus all classification technique, there is a single classifier for each of the 21 subjects.

4.4.1 Hold-out set (AWA)

The results of this section were done using a multiclass classification technique.

The subjects in this set have permission to access the system: 21 subjects, 84 epochs in total.

From Figure 4.22, the diagonal represents the correct classification for each subject and its difference in color is the quantity of epochs. For example:

- Yellow: during the 4 trials, 15 subjects were correctly recognized.
- Green: during 3 out of 4 trials, 4 subjects were correctly recognized.
- Cyan: during 2 out of 4 trials, 1 subject was correctly recognized.
- Blue: during 1 out of 4 trials, 1 subject was correctly recognized.

The off-diagonals show how often a subject (in this case: S1, S5, S9, S11 and S18) would have obtained permission to enter the system as an impostor. Subject 18 was the most difficult to

recognize. It was confused with subjects: s1, s2 and s8. And subject 1 was half of the time misclassified as subject 18. Results might be different if using other electrodes different from Fp1.

Table 4.14 shows the metrics used to evaluate the performance of the model. It can be seen that the accuracy is slightly lower when compared with the obtained one while doing cross validation with an optimized algorithm (89.29% vs 90.71%).

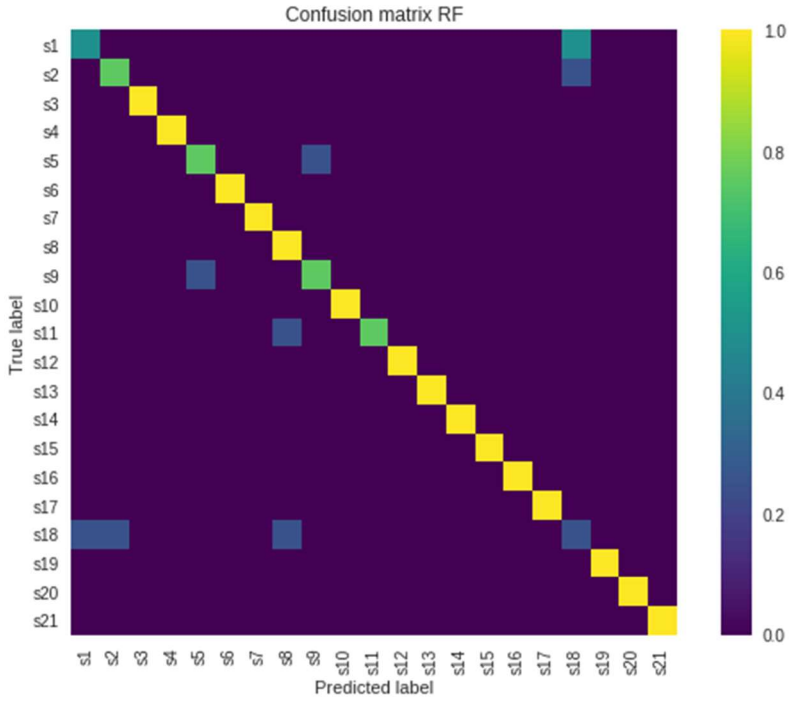


Figure 4-22 : Hold-out set (4 epochs per subject) – Fp1_RF

Table 4-14 : Evaluation metrics – Hold-out set_RF

Metrics: Hold-out – Random Forest	
Accuracy (%)	89.29
Recall (%)	89.29
Precision (%)	89.68
f1-score (%)	89.18
FAR (%)	0.54
FRR (%)	10.71
HTER (%)	5.62

4.4.2 Hold-out and Intruders set (AWA)

The results of this section were done using a one versus all classification. Therefore 21 binary classifiers were produced. One for each legitimate subject for the positive training data and all the data of the others legitimate subjects for the negative training data. Then for the test, a combination of Legitimates and Intruders was ran through all the 21 classifiers. If none of them is able to detect a legitimate subject, then the epoch can be deemed as an intruder.

This is an example of the number of epochs used for each binary classifier:

Classifier 1

Training: Legitimate s1 (40 epochs) vs (Legitimate s2 + … + Legitimate s21 (800 epochs))

Testing: Hold-out of Legitimate s1 (4 epochs) + all 16 intruders (64 epochs)

…

Classifier 21

Training: Legitimate s21 (40 epochs) vs (Legitimate s1 + … + Legitimate s20 (800 epochs))

Testing: Hold-out of Legitimate s21 (4 epochs) + all 16 intruders (64 epochs)

There are 3 types of subjects in this set: Those who have permission to access the system (Legitimates): 21 subjects, 84 epochs in total; those who do not have permission to access the system (Intruders): 16 intruders, 64 epochs in total and those who have permission to access the system but will try to impersonate other Legitimate subject (Impostors).

The off-diagonals show how often a subject (in this case: s1, s3, s5, s9, s11 and s15) would have obtained permission to enter the system as an impostor. Subject s1 was the most difficult to recognize. It was confused with subject s18 3 times out of 4.

From Figure 4.23, the diagonal represents the correct classification for each subject and its difference in color is the quantity of epochs. For example:

- Yellow: during the 4 trials, 15 subjects were correctly recognized.

- Green: during 3 out of 4 trials, 3 subjects were correctly recognized.
- Cyan: during 2 out of 4 trials, 2 subjects were correctly recognized.
- Blue: during 1 out of 4 trials, 1 subject was correctly recognized.

In the case of intruders, there is not a diagonal of correct recognitions. Epochs are everywhere. Intruders i3 and i4 have been granted access as Legitimates s20 and s5. It is interesting to see that most intruders are recognized as legitimate s5. Meanwhile legitimates 10, 13, 14, 15 and 17 are difficult to spoof.

Table 4.15 shows the metrics used to evaluate the performance of the model. The overall accuracy of the system is of 88.10%. FAR is very low, as only 2 intruders and 6 impostors were accepted as Legitimate subjects. FRR is also low, as 15 Legitimate subjects were correctly classified 4 times out of 4.

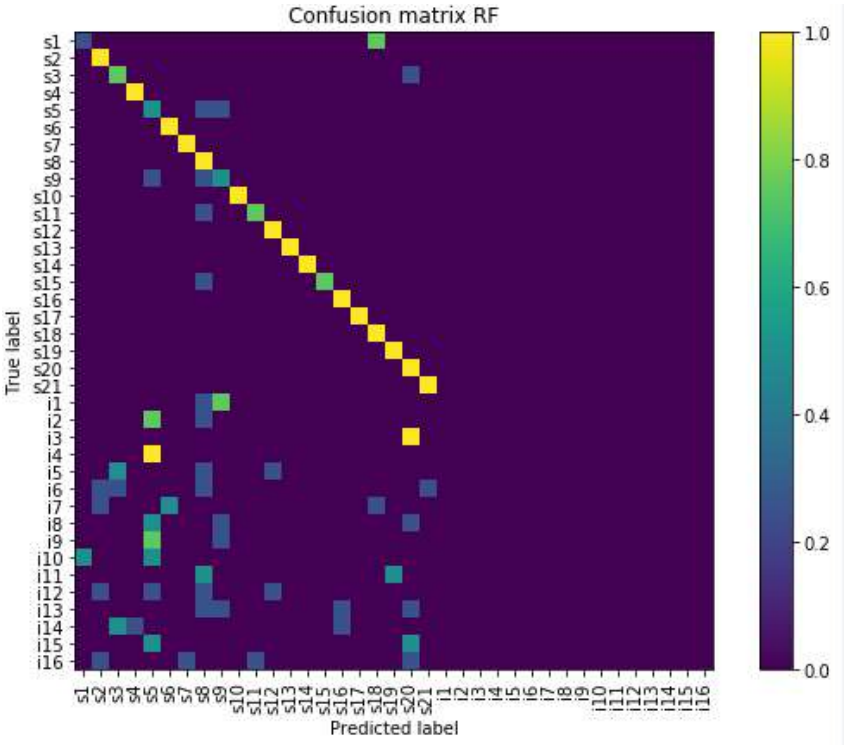


Figure 4-23 : Holdout_Intruders set – Fp1_RF

Table 4-15 : Evaluation metrics – Holdout_Intruders set_RF

Metrics: Hold-out and Intruders - RandomForest				
	Overall	Legitimates		Intruders
		Train	Test	
Accuracy (%)	88.10	97.3	88.10	96.88
Recall (%)	88.10			
Precision (%)	91.45			
f1-score (%)	87.61			
FAR (%)	0.60			
FRR (%)	11.90			
HTER (%)	6.25			

4.4.3 Continuous identification/authentication (SX-AWA)

The epochs of legitimate users in awake (AWA) and sleep states (SX) were used. The model created with SX epochs was tested with AWA epochs. It is interesting that most subjects while sleeping were predicted as s2 and s7 while awake (Figure 4-24).

Table 4.16 shows the metrics used to evaluate the performance of the model: FRR is very high. In conclusion, the intra-variance within each subject in SX and AWA is very high. Further exploration of better features should be done for future work.

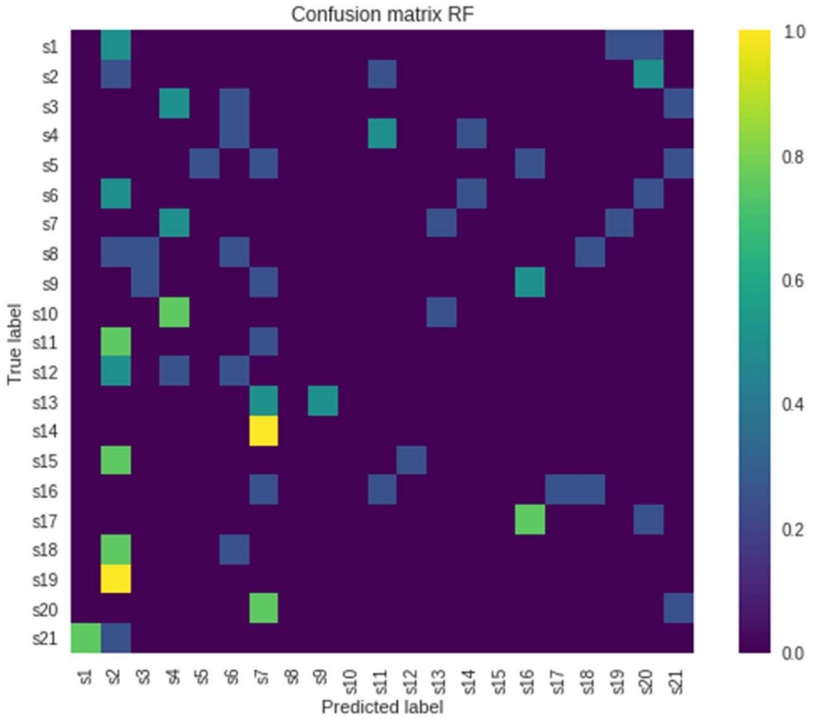


Figure 4-24: Continuous authentication - Fp1_RF (SX-AWA)

Table 4-16 : Evaluation metrics – Continuous_RF (SX-AWA)

Metrics: Continuous – Random Forest (SX-AWA)	
Accuracy (%)	2.38
Recall (%)	2.38
Precision (%)	4.98
f1-score (%)	2.27
FAR (%)	4.88
FRR (%)	97.62
HTER (%)	51.25

4.4.4 Continuous identification/authentication (ALL – AWA)

The epochs of legitimate users in awake (AWA) and all states (ALL) were used. The model created with ALL epochs was tested with AWA epochs.

Figure 4.25 shows that 12 subjects were correctly classified within the 4 trials. And s18 was the only one who was not recognized at all. These results are better than SX-AWA which means that in order to reduce the intra-variance of each subject, the state used for testing must be employed when creating the model. This is a disadvantage as subjects might be engaged in many different activities in a period of time. And it is not possible to train a model with all the features for a specific activity. Table 4.17 shows the metrics used to evaluate the performance of the model. Accuracy is worse when compared to 89.29% when using the same state for training and testing.

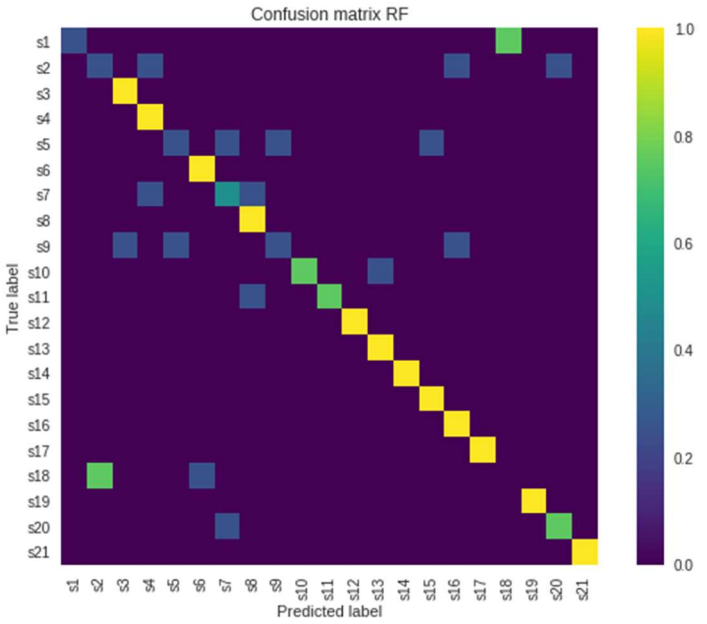


Figure 4-25 : Continuous authentication - Fp1_RF (ALL-AWA)

Table 4-17 : Evaluation metrics – Continuous_RF (ALL-AWA)

Metrics: Continuous – Random Forest (ALL-AWA)	
Accuracy (%)	75
Recall (%)	75
Precision (%)	74.76
f1-score (%)	72.55
FAR (%)	1.25
FRR (%)	25
HTER (%)	13.12

4.5 Discussion and Limitations

In this chapter, all the procedure to choose an algorithm and creating a model for a biometric system has been explained and evaluated against benchmark sets (Hold-out, Intruders and Continuous authentication). Also, the best stage, the best electrode for each stage and the features importance have been shown.

The choice of subjects to create the sets: Legitimates, Intruders was not random. It was done according to the number of available epochs. So, how much would results have changed if it wold have been possible to explore different combinations of subjects? This is an open question that will need to be explored for future work.

The unbalance in the number of epochs per subject, found in all datasets, was not terrible. Therefore, the performance results using metrics like: Precision, Recall and F1-score (which are not affected by the quantity of epochs) were not completely different from Accuracy within all algorithms.

When defining the best stage, it was only after optimization that all algorithms "agreed" in choosing Rem. Nevertheless, when choosing the best electrode for each stage, each algorithm gave different results.

Random Forest has shown to be superior over the other algorithms in all stages, when defining the best algorithm after optimization. It has never been used before in EEG biometrics.

When choosing between Fp1 and Fp2 optimized electrodes, the former was more performing than the latter one within all stages except by SWS. Also, when compared against the best electrode for each stage (840 epochs without optimization), Fp1 has been better in almost all stages. This is important for usability purposes as it proves that an optimized electrode is better in most cases than the best one (non-optimized) for a brain area. Nevertheless, when compared to the best electrode for each stage (all epochs without optimization), its performance was not better. It is worth exploring better undersampling techniques rather than using stratified random sampling; as this is good to achieve high performance while keeping low computational resources.

The good performance of Fp1 might have been influenced by its sensitivity to eye movements despite of the applied artifacts removal process (Muthukumaraswamy, 2013), (Urigüen & Begoña, 2015), (Schlögl et al., 2007). See Appendix G for decoding accuracies using EOG and EMG.

When analyzing the effect of optimization, accuracy increased in all stages with the exception of SWS using XG algorithm; where accuracy decreased in less than 1%. In AWA stage, all algorithms (XG is the exception), performed better than when using all epochs. And SX and ALL stages were better with XG and SVM (rbf) algorithms.

Results with the benchmark sets were very good. In the Hold-out set, 15 out of 21 legitimate users were granted access during all the 4 trials; for both classification techniques: multiclass and one vs all. In the case of Intruders, only 2 were falsely granted access during all the 4 trials.

In the continuous authentication (SX_AWA), results were not good. The intra-variance within each subject in SX and AWA was very high. Meanwhile in the continuous authentication (ALL_AWA); results were quite good. When comparing this scenario with the previous one, it could be concluded that for a good performance, stages used during training must be used during testing. But this is not realistic as any person can be engaged in many different situations which might affect their EEG. And this will make the training set very difficult to implement. A solution could be, if supervised learning is still used, "on-line machine learning". Here the model continuously updates and no epochs are stored. Other alternative could be unsupervised learning. Here the algorithm will find by itself significant patterns and trends in the ever changing brain signal.

The fact that the Hold-out set has been created using stratified random sampling makes the performance estimation optimistic (despite not being used during training and using a permutation test). Because EEG (especially in low frequency) is very similar between 2 consecutive time windows[9]. Also, when using the "stratified shuffled split" instead of a "time series split"; any time series nature of the data has been compromised. Since the training set can include data from both before and after the test set. And the principle of Train/Test is that Training data represents data known to the present, and Test data represents unseen data (perhaps literally from the future). For EEG biometrics, it is better to have different sessions and days in training and test data to have into account intersession variability.

[9] In this research a time window is an epoch of 30 seconds.

Another limitation was the number of subjects used during this research (21 Legitimates and 16 Intruders). Unfortunately, databases specifically for EEG biometrics research are scarce. And the collection of a diverse one with more than 50 subjects is difficult. The database is very important as Biometric systems do get progressively more challenged to keep FAR and FRR low, as subjects are added.

To sum up, EEG biometrics is often presented as an alternative to existing security technologies: fingerprint recognition, passwords, etc. But with this we are assuming that it is more secure for any given application. However, this is not true. EEG biometrics adds more value to security: cognitive states while authenticating/identifying (Morandi & Tzovaras, 2012).

Its recognition power[10] is still low and unstable when compared to physiological biometrics (ex. Fingerprint). But the risk of privacy loss is very high as it elicits personal information which can be misused to reveal other kind of information (intentions, emotions, medical conditions) and render the mean for discrimination (Morandi & Tzovaras, 2012).

Cognitive states, vital signals, fatigue and stress levels are useful in medicine: preventive and telemedicine. Or maybe in any other situation where it is important to fulfill a risky work: Truck driver, nuclear plant controller, air traffic controller, etc. So, EEG biometrics is better tailored for health environments where the additional value: **cognitive state** is useful for doctors to make superior diagnostics and where a series of ethical issues and major concerns in the privacy of health records are and will be covered by legislation.

Finally, it is important to comment in which ways this type of biometric system can be spoofed. The Nymi wristband is an ECG biometrics system, whose inventors claimed that there was no known means of falsifying this type of trait. But a published paper by Oxford University proved them wrong. The researchers succeeded in injecting ECG signals and also synthetic generated ones and therefore bypassing the security provided by the ECG biometrics. Something very interesting is that they succeeded without knowing the features used by the Nymi company. A conclusion

[10]EEG biometrics when combined with other types of biometrics (in a multimodal biometric system) improves the probability of matching with only one person and steps up demands for any fraudster.

could be that to make the spoofing more challenging, never to give access to raw data of any type of trait.

The techniques used by (Eberz et al., 2017) were:

- Hardware based AWG (Arbitrary Waveform Generator): software was written to load the ECG signals into the AWG and to set the necessary parameters to start signal generation.
- Software based AWG: A laptop's sound card was used as an ECG waveform generator.
- Playback of ECG signals encoded as audio files: ECG signals were filtered, scaled, sampled and stored as an audio file.

The generation of synthetic signals was important because the morphology of ECG signals depends on the device and electrode position that was used to measure them, much like with EEG. So a mapping function (training a model and optimising with genetic algorithms) was designed to eliminate statistical differences and make the signal match that of the target device (the Nymi wristband) (Eberz et al., 2017).

ECG signals much like EEG ones, can be captured and digitally stored. For example, by using the printouts or screen shots of the signal found in medical records. And also by intercepting the transmitted signal of off the shelf health mobile devices (Eberz et al., 2017).

The methods used for spoofing ECG signals can be as well used with EEG. The level of difficulty is higher but not impossible.

CHAPTER 5 CONCLUSION AND RECOMMENDATIONS

In this research, all the procedure to choose an algorithm and create a model for a biometric system (using wakefulness and sleep stages) has been explained and evaluated against benchmark datasets (Hold-out, Intruders and Continuous authentication). Also; the best stage, the best electrode for each stage and the features importance have been studied.

Finally, to conclude this research, the specific objectives that had been formulated at the beginning will be answered.

- *What stage is more efficient for identification/authentication?*
  Rem (sleep state) is the most efficient stage for identification/authentication, according to 4 algorithms: SVM (rbf), KNN, RF and XG.

- *Is it possible to distinguish individuals using 1 electrode?*
  Yes, in single stages (AWA, Rem, S1, S2 and SWS) accuracy is higher than 70%. And in combination of stages (SX, ALL), the accuracy is over 60%. In both cases the classification was above chance level.

- *What is the minimal duration, in seconds, to identify/authenticate a person?*
  In this study all datasets were segmented in 30s epochs according to sleep standards, so the minimal duration achieved was of 30 seconds. For future work, smaller epochs will be explored.

- *Is there entropy in brain signal?*
  Yes, and it was the best feature among 19 others. For future work, the rate of entropy will be measured to know if it is unpredictable enough to generate a strong encryption key. Also, other types of entropy will also be evaluated.

- *What is the effect of hyperparameter optimization in performance?*
  With optimization it has been proved that a small dataset with an optimized algorithm can be as good as or better than a big dataset with a non-optimized algorithm. This is very important as usually many machine learning studies are lacking of enough samples and they

have to generate synthetic data. In a biometric system, small datasets are important to guarantee a fast response while using few computational resources.

For future work, better techniques like Bayesian optimization will be evaluated. Also better techniques to undersample datasets will be explored (instead of a random removal of samples).

- *Is continuous identification/authentication possible?*

  Yes, but only when the same stages used during training were used for testing.

  For example, in the combinations:

  Train: SX, Test: AWA; the results were very bad.

  Train: ALL, Test: AWA; the results were good (above 70%).

  Implementing a training set with samples of the many situations in which a subject might be engaged is very difficult. So, other machine learning techniques, like online adaptive learning, will be explored.

BIBLIOGRAPHY

Abbas, S. N., Abo-Zahhad, M., & Ahmed, S. M. (2015). State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals. *IET Biometrics*. https://doi.org/10.1049/iet-bmt.2014.0040

Abdullah, M. K., Subari, K. S., Leo, J., Loong, C., & Ahmad, N. N. (2010). Analysis of the EEG Signal for a Practical Biometric System. *World Academy of Science, Engineering and Technology*, *4*(2008), 1123–1127.

Abdullah, M. K., Subari, K. S., Loong, J. L. C., & Ahmad, N. N. (2010). Analysis of effective channel placement for an EEG-based biometric system. In *Proceedings of 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2010* (pp. 303–306). https://doi.org/10.1109/IECBES.2010.5742249

Aguera, P. E., Jerbi, K., Caclin, A., & Bertrand, O. (2011). ELAN: A software package for analysis and visualization of MEG, EEG, and LFP signals. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/158970

Altahat, S., Tran, D., & Sharma, D. (2012). Human identification with electroencephalogram (EEG) signal processing. *2012 International Symposium on Communications and Information Technologies (ISCIT)*, 1021–1026. https://doi.org/10.1109/ISCIT.2012.6380841

American Academy of Neurology. (2012). Understanding Sleep Disorder, *1*, 1–16.

Ashby, C., Bhatia, A., Tenore, F., & Vogelstein, J. (2011). Low-cost electroencephalogram (EEG) based authentication. In *2011 5th International IEEE/EMBS Conference on Neural Engineering, NER 2011* (pp. 442–445). https://doi.org/10.1109/NER.2011.5910581

Bear, M; Connors, B; Paradiso, M. (2014). NEUROSCIENCE Exploring the Brain. *Lippincott Williams & Wilkins*, (1), 1–5. https://doi.org/10.1007/s13398-014-0173-7.2

Bergstra, J., & Yoshua, B. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, *13*, 281–305. https://doi.org/10.1162/153244303322533223

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., & Cutler, A. (n.d.). Random Forests. Retrieved December 22, 2016, from https://www.stat.berkeley.edu/~breiman/RandomForests/cc_manual.htm

Brigham, K., & Kumar, B. V. K. V. (2010). Subject identification from Electroencephalogram (EEG) signals during imagined speech. In *IEEE 4th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2010*. https://doi.org/10.1109/BTAS.2010.5634515

Campisi, P., Scarano, G., Babiloni, F., DeVico Fallani, F., Colonnese, S., Maiorana, E., & Forastiere, L. (2011). Brain waves based user recognition using the "eyes closed resting conditions" protocol. In *2011 IEEE International Workshop on Information Forensics and Security, WIFS 2011*. https://doi.org/10.1109/WIFS.2011.6123138

Chen, T., & He, T. (2015). xgboost : eXtreme Gradient Boosting, 0–3.

Chunying, F., Haifeng, L., Lin, M., & Bing, J. (2014). Induced event-related coherence measures during auditory change detection. In *Proceedings - 2014 International Conference on Medical Biometrics, ICMB 2014* (pp. 118–124). https://doi.org/10.1109/ICMB.2014.28

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1023/A:1022627411411

Dan, Z., Xifeng, Z., & Qiangang, G. (2013). An identification system based on portable EEG acquisition equipment. *Proceedings of the 2013 3rd International Conference on Intelligent System Design and Engineering Applications, ISDEA 2013*, 281–284. https://doi.org/10.1109/ISDEA.2012.70

Daubechies, I. (1992). *Ten Lectures on Wavelets (CBMS-NSF Regional Conference Series in Applied Mathematics)*. *Society for Industrial and Applied Mathematics SIAM*. Retrieved from http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0898712742

David Hairston, W., Whitaker, K. W., Ries, A. J., Vettel, J. M., Cortney Bradford, J., Kerick, S. E., & McDowell, K. (2014). Usability of four commercially-oriented EEG systems. *Journal of Neural Engineering*, *11*(4), 46018. https://doi.org/10.1088/1741-2560/11/4/046018

Del Pozo-Banos, M., Alonso, J. B., Ticay-Rivas, J. R., & Travieso, C. M. (2014). Electroencephalogram subject identification: A review. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2014.05.013

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern Classification. *Neuropsychopharmacology*, *2*(6), 654. https://doi.org/10.1038/npp.2011.9

Eberz, S., Paoletti, N., Roeschlin, M., Patani, A., Kwiatkowska, M., & Martinovic, I. (2017). Broken Hearted: How To Attack ECG Biometrics. In *Proceedings 2017 Network and Distributed System Security Symposium*. https://doi.org/10.14722/ndss.2017.23408

Eichenlaub, J. B., Ruby, P., & Morlet, D. (2012). What is the specificity of the response to the own first-name when presented as a novel in a passive oddball paradigm? An ERP study. *Brain Research*, *1447*, 65–78. https://doi.org/10.1016/j.brainres.2012.01.072

Ekandem, J. I., Davis, T. a., Alvarez, I., James, M. T., & Gilbert, J. E. (2012). Evaluating the ergonomics of BCI devices for research and experimentation. *Ergonomics*, *55*(5), 592–598. https://doi.org/10.1080/00140139.2012.662527

Gonçalves, S. I., De Munck, J. C., Pouwels, P. J. W., Schoonhoven, R., Kuijer, J. P. A., Maurits, N. M., … Lopes Da Silva, F. H. (2006). Correlating the alpha rhythm to BOLD using simultaneous EEG/fMRI: Inter-subject variability. *NeuroImage*, *30*(1), 203–213. https://doi.org/10.1016/j.neuroimage.2005.09.062

Good, P. (1994). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. *Springer Series in Statistics*. Retrieved from http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/038798898X

Gui, Q., Jin, Z., Ruiz Blondet, M. V., Laszlo, S., & Xu, W. (2015). Towards EEG biometrics: Pattern matching approaches for user identification. In *2015 IEEE International Conference on Identity, Security and Behavior Analysis, ISBA 2015*. https://doi.org/10.1109/ISBA.2015.7126357

Gui, Q., Jin, Z., & Xu, W. (2015). Exploring EEG-based biometrics for user identification and authentication. In *2014 IEEE Signal Processing in Medicine and Biology Symposium, IEEE SPMB 2014 - Proceedings*. https://doi.org/10.1109/SPMB.2014.7002950

Homan, R. W., Herman, J., & Purdy, P. (1987). Cerebral location of international 10-20 system electrode placement. *Electroencephalography and Clinical Neurophysiology*, *66*(4), 376–382. https://doi.org/10.1016/0013-4694(87)90206-9

How sleep works. (n.d.). Retrieved December 13, 2015, from https://www.howsleepworks.com/types.html

Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, *13*(2), 415–425. https://doi.org/10.1109/72.991427

Hsu, C. W., Chang, C. C., & Lin, C. J. (2016). A practical guide to support vector classication. *National Taiwan University*. Retrieved from http://www.csie.ntu.edu.tw/%7B~%7Dcjlin/papers/guide/guide.pdf

Hu, J. F. (2010). Comparison of different classifiers for biometric system based on EEG signals. In *Proceedings - 2nd International Conference on Information Technology and Computer Science, ITCS 2010* (pp. 288–291). https://doi.org/10.1109/ITCS.2010.77

Huang, B. F. F., & Boutros, P. C. (2016). The parameter sensitivity of random forests. *BMC Bioinformatics*, *17*(1), 331. https://doi.org/10.1186/s12859-016-1228-x

Huang, N., Shen, Z., Long, S., Wu, M., SHIH, H., ZHENG, Q., … Liu, H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *454*(1971), 995, 903. https://doi.org/10.1098/rspa.1998.0193

Jain, A., Flynn, & Ross, A. (2008). *Handbook of Biometrics*.

Jain, A. K., Bolle, R., & Pankanti, S. (2006). *Biometrics: personal identification in networked society*. *Biometrics* (Vol. 479). https://doi.org/10.1007/978-0-387-32659

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, *26*(3), 159–190. https://doi.org/10.1007/s10462-007-9052-3

Kumari, P., Kumar, S., & Vaish, A. (2014). Feature extraction using emprical mode decomposition for biometric system. In *2014 International Conference on Signal Propagation and Computer Technology, ICSPCT 2014* (pp. 283–287). https://doi.org/10.1109/ICSPCT.2014.6885030

Lajnef, T., Chaibi, S., Ruby, P., Aguera, P. E., Eichenlaub, J. B., Samet, M., … Jerbi, K. (2015). Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines. *Journal of Neuroscience Methods*, *250*, 94–105.

https://doi.org/10.1016/j.jneumeth.2015.01.022

Laufs, H., Krakow, K., Sterzer, P., Eger, E., Beyerle, A., Salek-Haddadi, A., & Kleinschmidt, A. (2003). Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proceedings of the National Academy of Sciences*, *100*(19), 11053–11058. https://doi.org/10.1073/pnas.1831638100

Lee, H. J., Kim, H. S., & Park, K. S. (2013). A study on the reproducibility of biometric authentication based on electroencephalogram (EEG). *International IEEE/EMBS Conference on Neural Engineering, NER*, 13–16. https://doi.org/10.1109/NER.2013.6695859

Luck, S. J. (2005). An Introduction to the Event-Related Potential Technique. *Monographs of the Society for Research in Child Development*, *78*(3), 388. https://doi.org/10.1118/1.4736938

Machine Learning Mastery. (n.d.). Classification accuracy is not enough more performance measures you can use. Retrieved December 22, 2016, from http://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/

Maskeliunas, R., Damasevicius, R., Martisius, I., & Vasiljevas, M. (2016). Consumer grade EEG devices: are they usable for control tasks? *PeerJ*. https://doi.org/10.7717/peerj.1746

McLachlan, G. J. (2004). *Discriminant analysis and statistical pattern recognition*. *Wiley series in probability and statistics*. https://doi.org/10.1002/0471725293

Mihajlovic, V., Grundlehner, B., Vullers, R., & Penders, J. (2015). Wearable, wireless EEG solutions in daily life applications: What are we missing? *IEEE Journal of Biomedical and Health Informatics*, *19*(1), 6–21. https://doi.org/10.1109/JBHI.2014.2328317

Morandi, E., & Tzovaras, D. (2012). *Second Generation Biometrics: The ethical, Legal and Social Context*. *The International Library of Ethics, Law and Technology* (Vol. 11). https://doi.org/10.1007/s13398-014-0173-7.2

Muthukumaraswamy, S. D. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2013.00138

Nakanishi, I., Baba, S., & Miyamoto, C. (2009). EEG based biometric authentication using new

spectral features. In *2009 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)* (pp. 651–654). https://doi.org/10.1109/ISPACS.2009.5383756

Neurosky. (2007). Retrieved December 13, 2015, from https://store.neurosky.com/pages/mindwave

Nguyen, P., Tran, D., Huang, X., & Sharma, D. (2012). A Proposed Feature Extraction Method for EEG-based Person Identification. *The International Conference on Artificial Intelligence*.

Niedermeyer, E., & Silva, F. H. L. Da. (2004). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. *Lippincott Williams and Wilkins* (Vol. 1). Retrieved from http://books.google.com/books?id=tndqYGPHQdEC&pgis=1

Nunes, R. R., Almeida, M. P. de, & Sleigh, J. W. (2004). Entropia espectral: um novo método para adequação anestésica. *Revista Brasileira de Anestesiologia*, *54*(3), 404–422. https://doi.org/10.1590/S0034-70942004000300013

Oppenheim, A. V, Schafer, R. W., & Buck, J. R. (1999). *Discrete Time Signal Processing*. *Book* (Vol. 1999).

Palaniappan, R. (2005). Identifying Individuality Using Mental Task Based Brain Computer Interface. *Third International Conference on Intelligent Sensing and Information Processing, ICISIP*, 238–242. https://doi.org/10.1109/ICISIP.2005.1619442

Palaniappan, R. (2006). Vision related brain activity for biometric authentication. *IECON Proceedings (Industrial Electronics Conference)*, 3227–3231. https://doi.org/10.1109/IECON.2006.347485

Palaniappan, R., & Mandic, D. P. (2005). Energy of brain potentials evoked during visual stimulus: A new biometric? *Artificial Neural Networks: Formal Models and Their Applications - Icann 2005, Pt 2, Proceedings*, *3697*, 735–740.

Palaniappan, R., & Mandic, D. P. (2007). Biometrics from Brain Electrical Activity: A Machine Learning Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(4), 738–742. https://doi.org/10.1109/TPAMI.2007.1013

Palaniappan, R., & Raveendran, P. (2002). Individual identification technique using visual evoked

potential signals. In *Electronics Letters* (Vol. 38, p. 1634). https://doi.org/10.1049/el:20021104

Palaniappan, R., & Ravi, K. V. R. (2006). Improving visual evoked potential feature classification for person recognition using PCA and normalization. *Pattern Recognition Letters*, *27*(7), 726–733. https://doi.org/10.1016/j.patrec.2005.10.020

Paranjape, R. B., Mahovsky, J., Benedicenti, L., & Koles, Z. (2001). The electroencephalogram as a biometric. *Canadian Conference on Electrical and Computer Engineering*, *2*, 1363–1366. https://doi.org/10.1109/CCECE.2001.933649

Pardey, J., Roberts, S., & Tarassenko, L. (1996). A review of parametric modelling techniques for EEG analysis. *Medical Engineering and Physics*. https://doi.org/10.1016/1350-4533(95)00024-0

Parisa, S., Mo, B., & Mohammad, B. S. (2006). Person Identification by Using AR Model for EEG Signals. In *Proceeding of World Academy of Science, Engineering and Technology* (Vol. 11, pp. 281–285).

Phung, D., Tran, D., Ma, W., Nguyen, P., & Pham, T. (2012). Using Shannon Entropy as EEG Signal Feature for Fast Person Identification. In *The 22nd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2014)* (pp. 413–418). Retrieved from https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2014-121.pdf

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2007.04.019

Poulos, M. (1999). Parametric person identification from the EEG using computational geometry. *Electronics,Circuits and Systems*, *2*(2), 1005–1008. https://doi.org/10.1109/ICECS.1999.813403

Poulos, M., Rangoussi, M., Alexandris, N., & Evangelou, a. (2002). Person identification from the EEG using nonlinear signal classification. *Methods of Information in Medicine*, *41*(1), 64–75.

Ravi, K. V. R., & Palaniappan, R. (2005). Recognising individuals using their brain patterns. In *Proceedings - 3rd International Conference on Information Technology and Applications, ICITA 2005* (Vol. II, pp. 520–523). https://doi.org/10.1109/ICITA.2005.234

RobustScaler. (n.d.). Retrieved September 23, 2016, from http://scikit-learn.org/0.17/modules/generated/sklearn.preprocessing.RobustScaler.html#sklearn.preprocessing.RobustScaler%0A

Rocca, D. La, Campisi, P., & Scarano, G. (2012). EEG Biometrics for Individual Recognition in Resting State with Closed Eyes. *International Conference of the Biometrics Special Interest Group (BIOSIG)*, (Figure 1), 1–12. https://doi.org/10.1007/978-3-662-44485-6_22

Safont, G., Salazar, A., Soriano, A., & Vergara, L. (2012). Combination of multiple detectors for EEG based biometric identification/authentication. In *Proceedings - International Carnahan Conference on Security Technology* (pp. 230–236). https://doi.org/10.1109/CCST.2012.6393564

Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., & Pfurtscheller, G. (2007). A fully automated correction method of EOG artifacts in EEG recordings. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2006.09.003

Singhal, G. K., & Ramkumar, P. (2007). Person identification using evoked potentials and peak matching. In *2007 Biometrics Symposium, BSYM*. https://doi.org/10.1109/BCC.2007.4430555

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Stark, C. E. L., & Squire, L. R. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.221462998

Su, F., Xia, L., Cai, A., & Ma, J. (2010). Evaluation of recording factors in EEG-based personal identification: A vital step in real implementations. In *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics* (pp. 3861–3866). https://doi.org/10.1109/ICSMC.2010.5641768

Su, F., Xia, L., Cai, A., Wu, Y., & Ma, J. (2010). EEG-based personal identification: From proof-of-concept to a practical system. In *Proceedings - International Conference on Pattern Recognition* (pp. 3728–3731). https://doi.org/10.1109/ICPR.2010.908

Tangkraingkij, P., Lursinsap, C., Sanguansintukul, S., & Desudchit, T. (2009). Selecting Relevant EEG Signal Locations for Personal Identification Problem Using ICA and Neural Network. *2009 Eighth IEEEACIS International Conference on Computer and Information Science*, 616–621. https://doi.org/10.1109/ICIS.2009.156

Tangkraingkij, P., Lursinsap, C., Sanguansintukul, S., & Desudchit, T. (2010). Personal Identification by EEG Using ICA and Neural Network. In D. Taniar, O. Gervasi, B. Murgante, E. Pardede, & B. O. Apduhan (Eds.), *Computational Science and Its Applications -- ICCSA 2010: International Conference, Fukuoka, Japan, March 23-26, 2010, Proceedings, Part III* (pp. 419–430). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-12179-1_35

Tatum, W. O. (2014). Ellen R. grass lecture: Extraordinary EEG. *Neurodiagnostic Journal*, *54*(1), 3–21. https://doi.org/10.1080/21646821.2014.11079932

Tsinghua University Press, & Sun, K. (2016). Chaotic Secure Communication Principles and Technologies, 37. Retrieved from https://www.degruyter.com/view/product/460632

Urigüen, J. A., & Begoña, G.-Z. (2015). EEG artifact removal - State-of-the-art and guidelines. *Journal of Neural Engineering*. https://doi.org/10.1088/1741-2560/12/3/031001

Vogel, F. (1970). The genetic basis of the normal human electroencephalogram (EEG). *Humangenetik*. https://doi.org/10.1007/BF00295509

West Point Academy. (2012). *Biometrics Metrics Report*.

XGBoost - Add Out-of-Bag performance validation #1070. (n.d.). Retrieved December 22, 2016, from https://github.com/dmlc/xgboost/issues/1070

Yang, S. (2015). The use of EEG signals for biometric person recognition.

Yazdani, a, Roodaki, a, Rezatofighi, S. H., Misaghian, K., & Setarehdan, S. K. (2008). Fisher linear discriminant based person identification using visual evoked potentials. *2008 9th International Conference on Signal Processing*, 1677–1680. https://doi.org/10.1109/ICOSP.2008.4697459

Zhao, Q., Peng, H., Hu, B., Li, L., Qi, Y., Liu, Q., & Liu, L. (2010). Towards an Efficient and Accurate {EEG} Data Analysis in {EEG}-Based Individual Identification. In *Ubiquitous*

Intelligence and Computing (Vol. 6406, pp. 534–547). https://doi.org/10.1007/978-3-642-16355-5_41

APPENDICES

APPENDIX A – DIFFERENT STUDIES FOR RESTING STATE

PAPER	# OF SUBJECTS	MODE	EYE STATUS	ELECTRODES	BANDS	PRE-PROCESSING	FEATURE EXTRACTION	CLASSIFICATION ALGORITHM	ACCURACY
Person identification based on parametric processing of the EEG (Poulos et al.,1999)	4 genuines, 75 impostors	I	REC	O2 - WET	Alpha	Fourier Transform	AR	Kohonen Linear Vector Quantizer (NN)	72% - 84%
Parametric person identification from the EEG using computational geometry (Poulos et al., 1999)	4 genuines, 75 impostors	I	REC	O2 - WET	Alpha	Fourier Transform	AR	Intersection of convex polygons	95%
Improving Individual Identification in Security Check with an EEG Based Biometric Solution (Zhao et al., 1999)	10	I	REC	Cz – DRY	Alpha, theta, SMR	40 Hz low pass filtering, ICA	AR	KNN	97.63%
Neural network based person identification using EEG features (Poulos et al., 1999)	4 genuines, 75 impostors	I	REC	O2 - WET	Alpha	1-30 Hz bandpass filter. 3 subbands: [7-10Hz], [8-11Hz], [9-12Hz]	FFT	Kohonen Linear Vector Quantizer (NN)	80 to 100%
The electroencephalogram as a biometric (Paranjape et al., 2001)	40	I	REC, REO	P4 – WET	Alpha	epochs of 8.533s	AR	Discriminant analysis	61 to 99% (order 3 and 18, 21)-All dataset. 49, 85 and 82% (order 3, 15 and 21) – 50/50 dataset
Person Identification from the EEG using Nonlinear Signal Classification (Poulos et al., 2002)	4 genuines, 75 impostors	I, A	REC	O2 - WET	Alpha, beta, delta, theta	1-30Hz low pass filtering	bilinear model	LVQ (NN)	99.50%

PAPER	# OF SUBJECTS	MODE	EYE STATUS	ELECTRODES	BANDS	PRE-PROCESSING	FEATURE EXTRACTION	CLASSIFICATION ALGORITHM	ACCURACY
Unobtrusive biometric system based on electroencephalogram analysis (Riera et al., 2008)	51 genuines, 36 impostors	I, A	REC	FP1, FP2 – WET	Delta, theta, alpha, beta and gamma	Second-order pass band filter. Notch filter at 50Hz.	1-electrode: AR, Fourier. 2-electrodes: Mutual information, Coherence, cross-correlation.	Fisher's discriminant analysis	A: 98.1%, I: 95.1%
The Electroencephalographic Fingerprint of Sleep Is Genetically Determined: A Twin Study (De Gennaro et al., 2008)	40	Similarity	Sleep, awake	Fz, Cz, Pz – WET	0.5 to 26.0Hz	High-pass and low-pass filter. Ocular and muscle artifacts excluded	FFT	Newton–Raphson and Fisher scoring method	NREM=95% AWAKE=81%
EEG-based personal identification: from proof-of-concept to a practical system (Su et al., 2010)	40	I	REC	FP1 – DRY	4 to 33Hz.	5 to 32Hz	AR and PSD	KNN, KNN+FDA, SVM(multiclass), LVQ	KNN+FDA = 97.5% KNN = 70,7%, SVM = 79,6%, LVQ = 81,9%, KNN+FDA=84,2%(for 30s)
Brain waves based user recognition using the "Eyes Closed Resting Conditions" protocol (Campisi et al., 2011)	48	I	REC	10 different triplets – WET	40Hz, 33.33Hz, 20Hz	low pass filtering and downsampling. Windows of 3s	AR(order 6) reflection coefficients	Polynomial regression	T7-Cz-T8 for 40Hz= 95.7%, for 33.33Hz= 96.08%. O1-POZ-O2 for 20Hz=88%
EEG biometrics for individual recognition in resting state with closed eyes (LaRocca et al., 2012)	45	I	REC	combinations of (2), (3), (5) – WET	alpha, beta, theta, delta	Band-pass filtering (0.5-30Hz) and segmentation	AR(order 12) reflection coefficients	Polynomial regression	Fcz-Poz(2) = 97,09%, O1-Poz-O2(3)=98.73%, Cz-Tpz-Cpz-TP8-Pz(5)=98.56%

PAPER	# OF SUBJECTS	MODE	EYE STATUS	ELECTRODES	BANDS	PRE-PROCESSING	FEATURE EXTRACTION	CLASSIFICATION ALGORITHM	ACCURACY
An identification system based on portable EEG acquisition equipment (Dan et al (2013)	13	I	REC	FP1 – DRY	alpha	Artifact removal, notch filtering	AR	NN, SVM (polynomial), LDA	75%, 87.18%, 87%
Using Shannon Entropy as EEG Signal Feature for Fast Person Identification (Phung et al., 2014)	40	I	REC, REO	(23) – WET	alpha, beta, gamma	none	Shannon entropy	SVM (linear)	97.10%
Human brain distinctiveness based on EEG spectral coherence connectivity (LaRocca et al., 2014)	108	I	REC, REO	(56) – WET	delta, theta, beta and gamma	Downsampled and low pass filtered.	PSD and Spectral Coherence	Mahalanobis distance-based	Single: PSD, REC=90,49%; Fusion: COH,REO=100% (9 electrodes), COH,REC=100% (9 elecrodes)
An EEG-Based Biometric System Using Eigenvector Centrality in Resting State Brain Networks (Fraschini et al., 2015)	109	A	REC, REO	(64) – WET	Delta, theta, alpha, beta and gamma	Bandpass filtering	Nodal Eigenvector Centrality	Euclidean distance	REO= 0.006 – 0.328 (AUC), REC= 0.018 -0.402 (AUC)
EEG Feature Selection and the Use of Lyapunov Exponents for EEG - based Biometrics (Hwan et al., 2016)	7	I	REC	(16) – DRY	Theta, alpha, beta	Bandpass	PSD, median frequency (2-50Hz), Shannon entropy and maximum of the positive Lyapunov exponents	SVM(linear)	Best 12 features, 9 electrodes=94,9%, Lyapunov, T4-F4=85,6%
On the Invariance of EEG-based Signatures of Individuality with Application in Biometric Identification (Wang et al., 2016)	4	I	REC	(128), (28) – WET	delta, theta, alpha, beta, gamma	ICA to remove eyes movement. Background noise and muscle artifacts removed.	Continuous wavelet transform using Morlett as the mother	KNN	Same session: beta=99,14%. Different session: beta2=86,50%

APPENDIX B – DIFFERENT STUDIES FOR SENSORY STIMULI

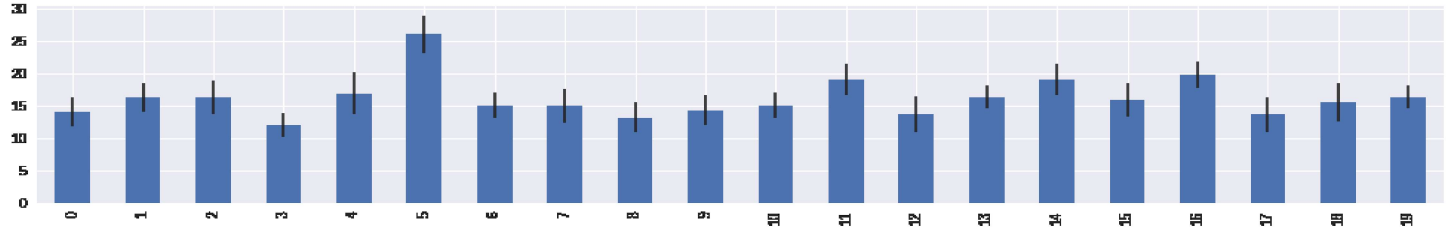
PAPER	# OF SUBJECTS	MODE	Visual stimuli	EEG CHANNELS	BANDS	PRE-PROCESSING	FEATURE EXTRACTION	CLASSIFICATION ALGORITHM	ACCURACY
A New Method to Identify Individuals Using Signals from the Brain (Palaniappan et al., 2003)	20	I	Snodgrass pictures	(61) – WET	gamma	eye blink removal, PCA	Power using Simplified Fuzzy ARTMAP	Neural Networks	94.18%
Method of identifying individuals using VEP signals and neural network (Palaniappan, 2004)	20	I	Snodgrass pictures	(61) – WET	gamma	eye blink removal, zero-phase Butterworth bandpass	Energy of each channel	Neural Networks	99.06%
Biometrics from Brain Electrical Activity: A Machine Learning Approach (Palaniappan et al., 2007)	40	I	Snodgrass pictures	(61) – WET	gamma	simultaneous diagonalization	MUSIC dominant power	Elman NN, Knn	98.12%, 96.13%
Biometric authentication using brain responses to visual stimuli (Zuquete et al., 2010)	70	A	Snodgrass pictures	(8)- WET	gamma	eye blink removal, Bandpass filter	Energy of each channel	Knn, Support Vector Data Description	AUC=0.981 – 0.996
Person authentication from neural activity of face-specific visual self-representation (Yeom et al., 2013)	10	I, A	Self and non-self face pictures	(18) – WET	all	Bandpass filtering (0.1 to 100Hz), notch filter (50Hz).	Temporal and dynamic features	SVM (gaussian)	86.10%
Novel HHT-Based Features for Biometric Identification Using EEG signals (Su et al., 2014)	118	I	Snodgrass pictures	Oz – WET	All (0 – 128 Hz)	Wavelet Packet Decomposition filtering	Hilbert Huang Transform	LDA	95.88%

APPENDIX C - DIFFERENT STUDIES FOR COGNITIVE ACTIVITIES

PAPER	# OF SUBJECTS	MODE	Task	ELECTRODES	BANDS	PRE-PROCESSING	FEATURE EXTRACTION	CLASSIFICATION ALGORITHM	ACCURACY
Electroencephalogram Signals from Imagined Activities: A Novel Biometric Identifier for a Small Population (Palaniappan, 2006)	5	I	baseline, imagined letter composing, figure rotation, counting and math	(6) – WET	alpha, beta, gamma	Band-pass filtering (0.1-100Hz, elliptic filter	AR, channel and inter-channel power spectrum + PCA	Linear Discriminat Classifier	32 – 100%
Person Authentication Using Brainwaves (EEG) and Maximum A Posteriori Model Adaptation (Marcel et al., 2007)	9	A	motor imagery (left and right hand), generation of words	(8) – WET	8-30 Hz	Surface Laplacian filtering	PSD	Gaussian Mixture model and Maximum A Posteriori	HTER: 7.1 – 42.6%
Multitask Learning for EEG-Based Biometrics (Su, 2008)	9	I	motor imagery (left and right index fingers)	(15) – WET	8-30 Hz	Bandpass filtering, normalization	CSP	NN	94.81 – 95.6%
Subject Identification from Electroencephalogram (EEG) Signals During Imagined Speech (Bringham et al., 2010)	6	I	imagined speech	(124) – WET	4-25 Hz	Artifacts removal, notch filtering	AR	SVM (linear), Knn	99.76%, 98.52 – 99.41%
Motor Imagery EEG-Based Person Verification (Nguyen et al., 2013)	9	A	motor imagery (left and right hands)	(3) – WET	8-30 Hz	none	PSD, AR	SVDD, GMM	EER: 0.0409, 0.0441
Novel HHT-Based Features for Biometric Identification Using EEG signals (Su et al., 2014)	105	I	motor imagery (left and right hands)	Cz – WET	All (0 – 80 Hz)	Wavelet Packet Decomposition filtering	Hilbert Huang Transform	KNN	99.00%
Towards EEG Biometrics: Pattern Matching Approaches for User Identification (Gui et al., 2015)	30	I, A	reading tasks	Oz – WET	none	normalization, standarization	none	Euclidean distance, Dynamic time warping	81.17%, 67.17%

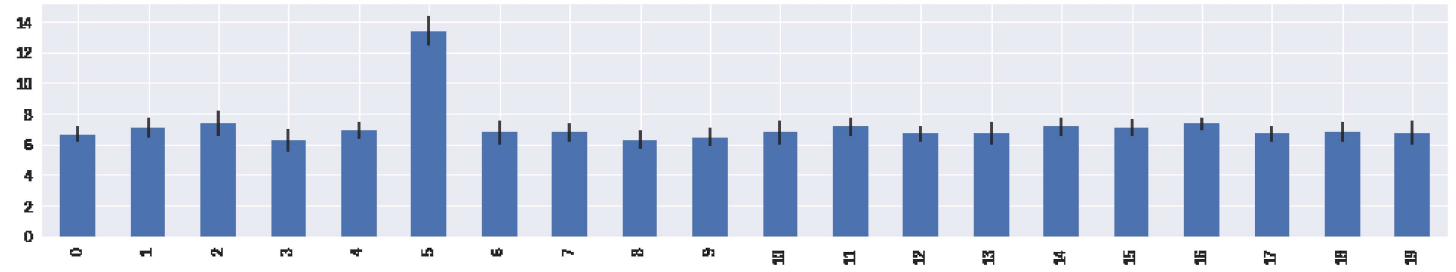
APPENDIX D – FEATURES IMPORTANCE PER ELECTRODE (AWA)

	0. relPowDelta	1. relPowTheta	2. relPowAlpha	3. relPowBeta	4. relPowSigma	5. SpectralEntropy	6. powDelta/powTheta	7. powDelta/powAlpha	8. powDelta/powBeta	9. powDelta/powSigma	10. powTheta/powDelta	11. powTheta/powAlpha	12. powTheta/powBeta	13. powTheta/powSigma	14. powAlpha/powTheta	15. powAlpha/powBeta	16. powAlpha/powSigma	17. powBeta/powTheta	18. powBeta/powSigma	19. powSigma/powTheta
Fp1	12.98	17.2	16.4	10.8	11.92	24.1	11.71	12.33	8.67	10.92	11.71	20.55	14.07	16.13	20.56	16.81	18.38	14	16.04	16.14
Fp2	10.25	16.67	14.04	10.62	16.08	25.81	15.51	12.23	11.1	15.62	15.3	21.36	15.29	13.17	21.36	10.8	17.46	15.17	14.01	13.17
C3	11.6	13.56	20.4	12	15.43	22.63	13.45	12.9	10.17	13.86	13.57	17.49	9.95	16.01	17.5	16.71	17.71	9.95	12.35	16.02
C4	14.3	14.5	16.95	14.32	14.15	25.14	15.01	13.73	16.96	16.06	14.99	16.29	11.3	14.02	16.37	17.81	21.14	11.31	14.76	14.07
Fz	12.85	17.55	18.2	14.89	22.13	27.56	14.8	11.8	13	13.01	14.89	18.08	17.55	18.62	18.1	13	22.04	17.36	18.25	18.68
Cz	16.38	17.29	9.32	15	21.32	27.32	13.87	14.67	15.6	14.62	13.8	17.69	16.36	19.99	17.68	12.02	23.1	16.36	19.8	19.99
Pz	14.35	16.56	16.44	9.95	20.39	25.05	14.45	13.6	13.82	15.99	14.57	16.24	13.65	17.79	16.21	13.58	18.49	13.77	17.79	18.07
T3	16.11	13.87	16.37	11.62	12.9	25.61	18.36	12.77	14.93	15.81	18.38	14.81	13.67	15.48	14.83	15.01	18.76	13.82	14.35	15.63
T4	16.71	11.65	17.65	10.12	20.04	22.62	18.08	13.21	15.85	13.86	18.32	17.37	13.54	18.61	17.49	19.24	20.14	13.64	20.67	18.62
O1	18.76	19.46	15.63	12.6	19.86	32.37	19.89	20.55	15.71	19.14	19.89	20.26	17.96	16.29	20.21	17.31	22	17.95	20.32	16.36
O2	12.69	19.67	20.04	14.62	15.74	24.77	13.93	15.89	11.75	16.87	13.86	21.27	15.55	17.82	21.14	15.52	20.73	15.55	15.95	18
F3	13.38	17.58	16.32	12.24	19.75	28.37	13.08	13.48	14.58	11.85	13.18	16.86	12.6	14.7	16.86	16.1	22.24	12.6	15.92	14.7
F4	13.44	15.83	14.65	11.8	18.29	26.37	15.57	15.02	10.87	17.2	15.52	19.32	14.25	15	19.04	15.2	17.77	14.25	17.2	15.15
P3	14.96	18.02	17.21	12.99	18.45	31.29	13.55	18.04	12.04	12.32	13.46	21.68	14.74	18.02	21.94	20.57	20.4	14.65	10.9	18.2
P4	13.14	13.24	16	8.26	10.74	23.13	13.99	13.5	10.6	11.35	14.07	21.69	9.38	15.63	21.69	14.65	17.13	9.38	9.57	15.63
FC1	10.2	16.32	13.05	12.85	14.29	22.62	14.65	14.67	13.18	12.23	14.64	17.2	10.83	14.74	17.25	14.15	15.74	10.7	14.35	14.8
FC2	17.19	19.9	17.9	11.54	15.39	23.85	15.81	17.4	12.96	11.38	15.82	20.85	17.6	16.31	20.82	17.85	21.07	17.45	13.65	16
CP1	12.92	14.38	14.01	12.76	15.36	26.4	13.68	20.23	12.87	13.04	13.57	23.76	8.95	16.04	23.88	16.55	21.77	8.96	13.6	16.14
CP2	15.11	16.36	19.13	9.44	17.86	29.37	16.14	17.37	16.08	15.54	16.11	19.29	12.17	15.24	19.25	19.89	20.73	12.13	15.06	15.24



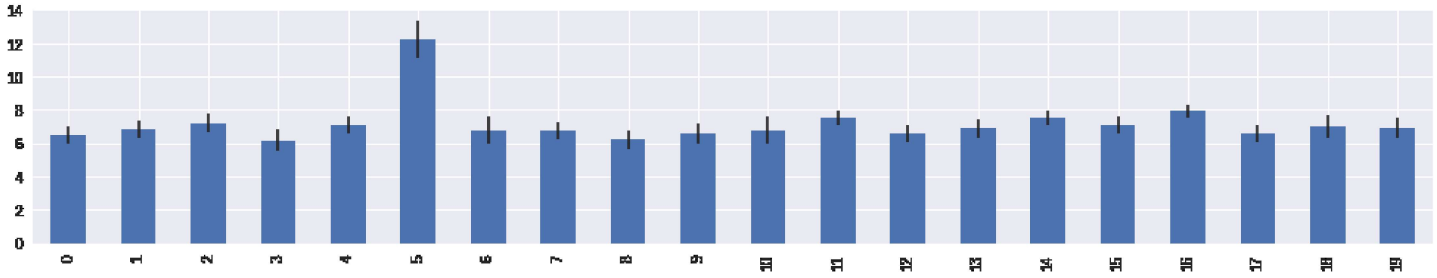
APPENDIX E - FEATURES IMPORTANCE PER ELECTRODE (SX)

	0. relPowDelta	1. relPowTheta	2. relPowAlpha	3. relPowBeta	4. relPowSigma	5. SpectralEntropy	6. powDelta/powTheta	7. powDelta/powAlpha	8. powDelta/powBeta	9. powDelta/powSigma	10. powTheta/powDelta	11. powTheta/powAlpha	12. powTheta/powBeta	13. powTheta/powSigma	14. powAlpha/powTheta	15. powAlpha/powBeta	16. powAlpha/powSigma	17. powBeta/powTheta	18. powBeta/powSigma	19. powSigma/powTheta
Fp1	6.96	7.61	6.31	7.52	7.47	12.99	6.37	6.73	6.99	6.83	6.37	6.93	6.38	6.8	6.92	6.73	7.43	6.37	7.27	6.8
Fp2	7.64	7.65	7.19	6.38	6.34	12.88	6.89	6.82	6.85	5.7	6.91	7.35	6.43	6.51	7.26	7.06	7.14	6.43	6.04	6.51
C3	6.54	6.92	8.42	6.83	6.91	14.93	7.56	7.78	6.34	7.08	7.55	8.11	6.85	7.21	8.11	7.51	6.97	6.83	6.93	7.21
C4	6.65	7.7	6.53	6.42	7.61	14.26	6.81	7.29	6.45	6.44	6.81	7.17	6.15	7.3	7.18	7.7	8.21	6.15	7.35	7.29
Fz	6.01	7.81	7.82	6.7	6.89	12.76	6.6	6.38	5.21	6.4	6.62	6.83	6.04	7.52	6.83	7.94	7.7	6.01	7.01	7.56
Cz	5.86	7.59	7.59	5.46	6.61	14.5	6.29	6.78	5.85	6.56	6.29	6.88	6.68	7.98	6.88	7.03	7.38	6.69	6.55	8
Pz	6.41	6.84	7.73	6.49	7.38	13.18	8.01	6.9	6.18	6.84	8.02	7.03	7.49	7.79	7.03	7.12	6.91	7.44	7.37	7.8
T3	7.22	6.72	7.07	5.24	6.9	12.83	8.35	7.24	6.49	6.97	8.35	8.49	6.77	7	8.52	6.97	7.86	6.75	6.74	7.05
T4	6.42	7.27	8.19	6.46	6.58	13.54	7.61	6.06	6.72	5.77	7.61	7.29	6.91	6.45	7.3	6.13	7.33	6.91	5.99	6.44
O1	6.26	5.96	7.55	5.94	6.32	14.26	7.47	6.78	7.22	7.38	7.45	6.96	6.3	6.18	6.96	6.98	6.84	6.29	5.63	6.19
O2	6.4	7.16	7.57	5.09	5.37	11.99	6.07	7.14	6.95	5.01	6.09	7.94	6.71	5.53	7.94	7.27	7.03	6.73	5.57	5.55
F3	6.96	6.53	8.23	6.47	6.7	14.02	6.27	5.61	4.95	6.19	6.28	7.52	6.02	6.65	7.52	7.42	7.41	6.02	7.29	6.65
F4	5.89	6.16	6.96	6.4	7.59	14.34	6	6.28	6.71	6.39	5.97	6.64	6.34	6.82	6.65	7.56	6.96	6.34	7.77	6.85
P3	7.37	6.57	5.9	6.35	6.3	13.75	7.13	6.52	5.78	5.63	7.14	6.96	6.9	4.91	6.96	6.33	7.23	6.88	6.21	4.91
P4	7.01	6.32	6.32	5.03	6.77	12.57	5.8	7.41	5.92	7.24	5.8	7.05	7.12	7.08	7.05	6.93	7.27	7.12	6.82	7.08
FC1	6.79	6.68	6.47	5.7	6.93	11.86	6.84	7.25	6.18	6.17	6.81	6.18	5.99	6.95	6.17	5.94	7.07	5.99	7.54	6.95
FC2	5.84	6.25	6.7	6.21	7.55	12.86	6	6.22	6.49	6.65	6.01	7.31	7.12	6.72	7.36	6.33	6.71	7.12	7.33	6.72
CP1	6.84	7.67	8.57	7.43	7.19	12.23	5.67	5.97	5.7	6.82	5.67	6.28	7.16	6.03	6.31	7.13	7.84	7.16	7.04	6.03
CP2	6.7	8.17	8.17	6.48	7.16	14.82	6.76	7.54	6.04	6.48	6.74	6.67	7.8	6.62	6.66	8.03	7.99	7.8	6.19	6.62



APPENDIX F: FEATURES IMPORTANCE PER ELECTRODE (ALL)

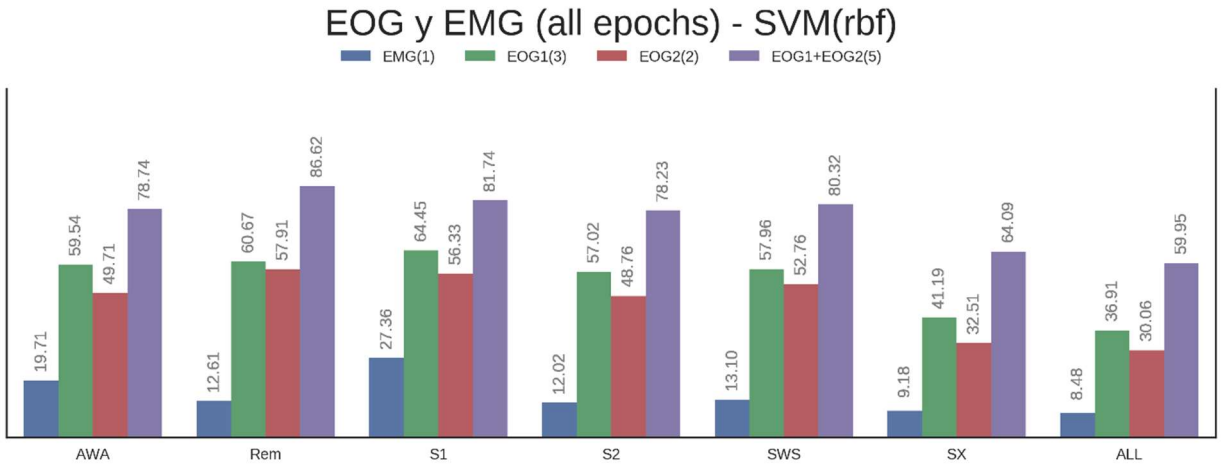
	0. relPowDelta	1. relPowTheta	2. relPowAlpha	3. relPowBeta	4. relPowSigma	5. SpectralEntropy	6. powDelta/powTheta	7. powDelta/powAlpha	8. powDelta/powBeta	9. powDelta/powSigma	10. powTheta/powDelta	11. powTheta/powAlpha	12. powTheta/powBeta	13. powTheta/powSigma	14. powAlpha/powTheta	15. powAlpha/powBeta	16. powAlpha/powSigma	17. powBeta/powTheta	18. powBeta/powSigma	19. powSigma/powTheta
Fp1	6.1	7	6.6	6.75	7.32	10.75	6.4	7.2	6.18	6.35	6.41	7.73	6.04	7.42	7.73	7.13	7.05	6.04	6.79	7.41
Fp2	7.06	7.5	7.07	5.99	6.79	11.74	6.61	6.3	6.1	5.93	6.59	8.02	6.65	6.47	8	6.61	7.46	6.65	5.4	6.46
C3	6.66	6.82	8.1	6.41	6.74	13.15	6.92	7.22	5.47	7.59	6.94	7.91	6.18	7.11	7.9	6.7	7.35	6.16	6.65	7.13
C4	6.81	7.79	7.12	6.39	7.53	13.08	7.38	6.73	7.07	6.65	7.38	7.07	6.37	6.53	7.07	8.13	8.48	6.36	7.81	6.54
Fz	6.23	6.78	7.32	5.83	7.72	12.18	6.39	6.11	5.66	6.12	6.42	7.39	5.78	7.17	7.4	7.35	8.11	5.73	7.25	7.22
Cz	5.56	7.66	7.07	5.55	7.73	12.53	6.17	7.2	6.45	6.58	6.17	7.22	6.41	7.98	7.22	6.82	8.09	6.41	7.95	8
Pz	6.3	5.85	8.16	6.09	7.42	10.96	7.58	6.85	6.77	6.45	7.58	7.08	6.73	7.62	7.08	6.56	7.84	6.72	7.25	7.62
T3	7.17	6.34	7.3	5.65	6.91	12.49	8.12	6.52	6.87	6.86	8.12	7.56	6.5	7.38	7.55	7.3	7.99	6.49	7.6	7.39
T4	6.71	6.48	7.85	6.73	7.04	12.56	8.1	5.77	6.25	6.38	8.12	7.16	6.9	7.45	7.17	6.15	8.31	6.9	7.71	7.45
O1	6.7	6.88	6.91	6.62	6.39	13.44	8.32	7.13	7.46	7.72	8.31	7.8	7.08	6.79	7.81	6.76	7.7	7.07	6.49	6.78
O2	6.12	7.26	6.77	5.5	6	11.28	5.98	7.42	6.14	4.98	6.02	8.13	6.51	6.46	8.13	7.08	7.77	6.52	6.05	6.47
F3	7.17	7.01	7.48	6.65	6.88	13.71	6.19	6.41	5.46	6.9	6.22	7.41	5.84	6.35	7.41	7.3	8.51	5.85	6.86	6.36
F4	5.87	6.62	6.76	5.21	7.71	13.09	6.04	6.12	5.81	6.91	6.06	7.65	6.5	6.92	7.65	7.06	7.74	6.49	7.64	6.94
P3	6.7	6.02	6.26	6.53	6.86	12.85	6.87	6.88	6.1	5.79	6.85	8.07	6.69	5.5	8.07	7.27	8.09	6.69	6.33	5.52
P4	6.4	6.43	6.87	5.29	6.64	10.96	6.17	7.36	6.26	6.63	6.2	7.84	7.1	7.21	7.84	6.76	8.15	7.1	6.49	7.21
FC1	6.88	6.15	6.51	5.77	6.72	10.19	5.78	7.11	5.62	6.14	5.81	6.66	5.73	6.8	6.66	6.58	7.63	5.73	7.84	6.8
FC2	6.21	6.75	7.04	5.46	7.75	11.03	6.39	6.54	5.96	6.59	6.41	8.05	7.6	6.75	8.06	7.45	8.37	7.58	7.11	6.75
CP1	5.6	6.69	7.66	7.7	7.36	11.61	5.57	6.1	5.66	6.82	5.55	7.37	6.43	6.21	7.38	7.05	7.97	6.44	6.74	6.22
CP2	7.03	7.45	8.03	6.55	7.05	14.13	7.59	7.56	6.43	7	7.53	7.05	7.25	6.74	7.05	8.16	7.68	7.27	6.84	6.76

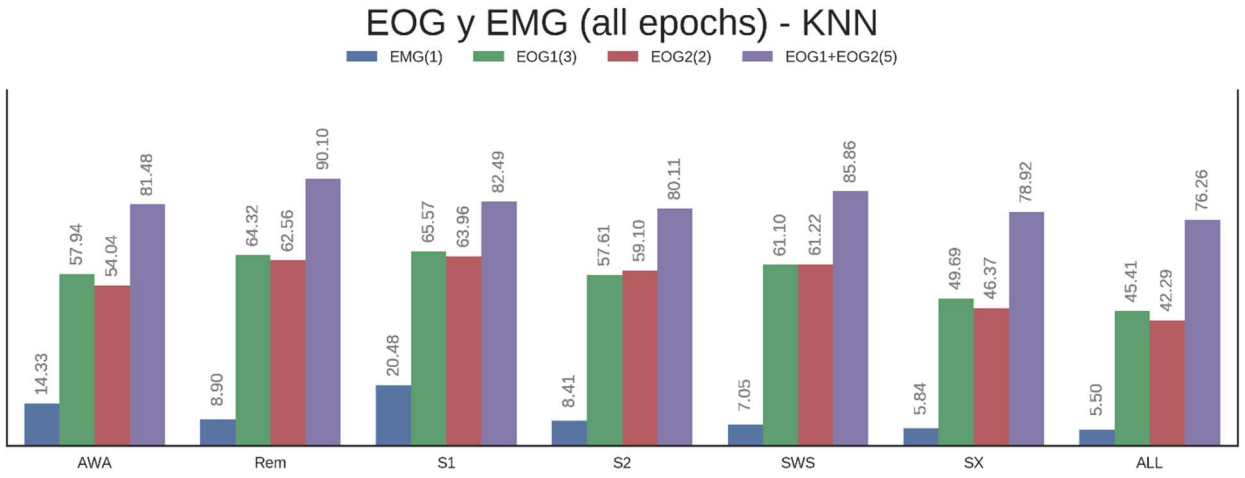


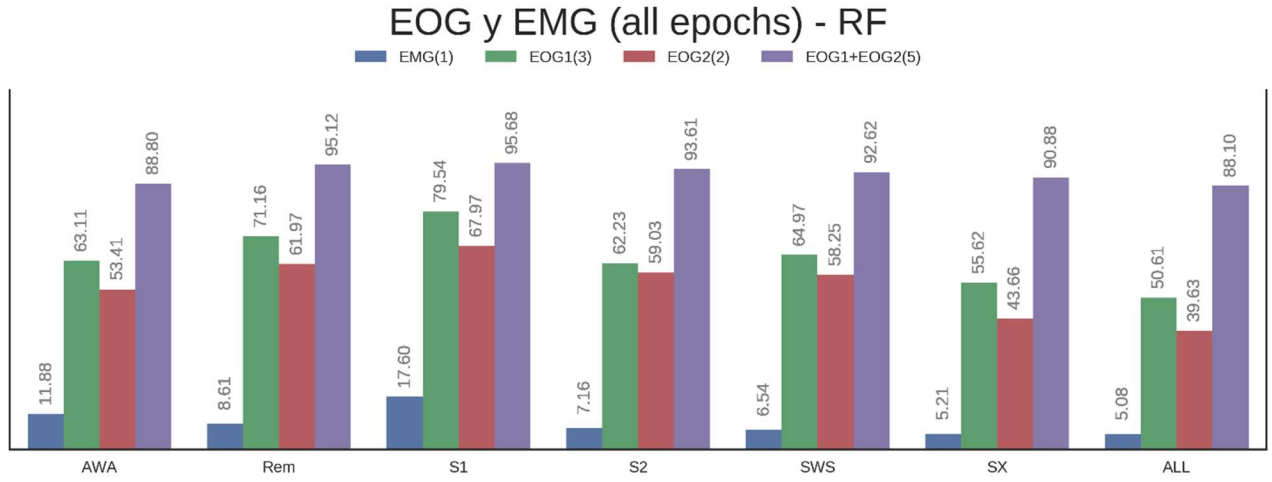
APPENDIX G : EOG AND EMG DECODING ACCURACIES

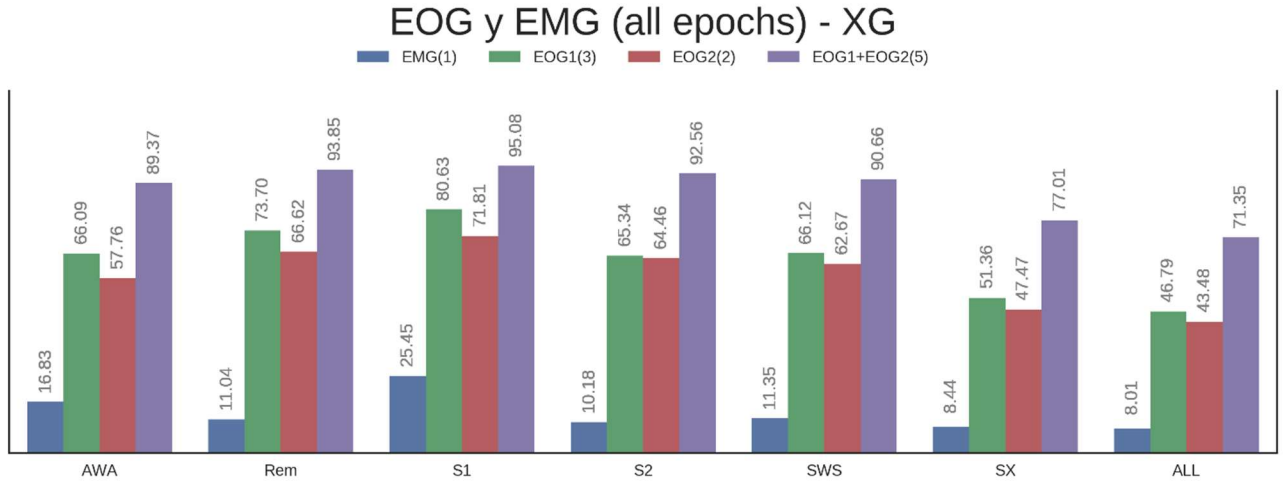
Feature [0] = petropy (EMG), Feature [2] = kurtosis (EOG1), Feature [4] = kurtosis (EOG2),

Feature [1] = petropy (EOG), Feature [3] = percentile (EOG1), Feature [5] = percentile (EOG2)






EOG y EMG (all epochs) - RF
EMG(1)
EOG1(3)
EOG2(2)
EOG1+EOG2(5)
11.88
63.11
53.41
88.80
8.61
71.16
61.97
95.12
17.60
79.54
67.97
95.68
7.16
62.23
59.03
93.61
6.54
64.97
58.25
92.62
5.21
55.62
43.66
90.88
5.08
50.61
39.63
88.10
AWA
Rem
S1
S2
SWS
SX
ALL


EOG y EMG (all epochs) - XG
EMG(1)
EOG1(3)
EOG2(2)
EOG1+EOG2(5)
16.83
66.09
57.76
89.37
11.04
73.70
66.62
93.85
25.45
80.63
71.81
95.08
10.18
65.34
64.46
92.56
11.35
66.12
62.67
90.66
8.44
51.36
47.47
77.01
8.01
46.79
43.48
71.35
AWA
Rem
S1
S2
SWS
SX
ALL