2025 AI model to decode Imagined Speech Ing. Fis. Edgar Lara Arellano

Universidad Autónoma de Querétaro

Facultad de Ingeniería

AI model to decode Imagined Speech

Tesis

Que como parte de los requisitos para obtener el Grado de

Maestro en Ciencias en Inteligencia Artificial

Presenta

Ing. Fis. Edgar Lara Arellano

Dirigido por:
Dr. Andras Takacs

Querétaro, Qro. a Mayo de 2025

Universidad Autónoma de Querétaro

Facultad de Ingeniería

Maestría en Ciencias en Inteligencia Artificial

AI model to decode Imagined Speech

Tesis

Que como parte de los requisitos para obtener el Grado de

Maestro en Ciencias en Inteligencia Artificial

Presenta

Ing. Fis. Edgar Lara Arellano

Dirigido por

Dr. Andras Takacs

Dr. Andras Takacs
Presidente

Dr. Juvenal Rodríguez Reséndiz
Secretario

Dr. Saúl Tovar Arriaga
Vocal

M.C. Luis Roberto García Noguéz
Suplente

M.C. Luis Felipe Estrella Ibarra
Suplente

Centro Universitario, Querétaro, Qro. México
Fecha de aprobación por el Consejo Universitario (abril 2024)

I dedicate this thesis with all my heart to my mother, because without her,
I would not have been able to achieve it.

"If I have seen further, it is by standing on the shoulders of giants."
- Sir Isaac Newton

Acknowledgments

I would like to thank the following professors: MSc. Ramón Torres Alonso, Dr. Juan Manuel Ramos Arreguín, Dr. Arturo González Gutiérrez, Dr. Alberto Hernández Almada, Dr. Efrén Gorrostieta Hurtado, Dr. Jesús Carlos Pedraza Ortega, Dr. José-Emilio Vargas-Soto, Dr. Marco Antonio Aceves-Fernández, Dr. Sebastian Salazar Colores, for the wonderful lectures they taught during my Master's degree.

To the professors: Dr. András Takács, Dr. Juvenal Rodríguez Reséndiz, Dr. Saúl Tovar Arriaga, Msc. Luis Roberto García Noguéz, Msc. Luis Felipe Estrella Ibarra, for guiding me in my research project.

I would also like to thanks to Dr. Juvenal Rodríguez Reséndiz, for providing the Emotiv Epoc+ device used to record the signals.

I am also grateful to Dr. András Takács, for his financial support and assistance in acquiring spare parts necessary for the experiments.

I would also like to thank the researchers at the Laboratory of Neurobiology and Cellular Bioengineering: Dr. Hebert Luis Hernández Montiel, Dr. Liane Aguilar Fabré, Dr. René Francisco Rodríguez Valdés, Lic. María Eugenia Herrera Álvarez, for guiding me on neurological issues and providing me with advice.

I would like to thank the National Council of Humanities, Sciences and Technologies (CONAHCYT) for the financial support provided for my Master's studies.

To the people who participated directly or indirectly in the completion of this thesis, for their opinions and contributions to the completion of this work.

I thank that special person, who was by my side supporting me.

A sincere thank to my Thesis Director, Dr. András Tákacs, for sharing his knowledge and guidance during my Master's studies.

I would like to thank my friends, especially Eng. Oscar Eduardo Orozco Ruíz, for listening to me every time I told him about a setback or a breakthrough in this journey.

I would like to thank my mother with all my heart, without her and her unconditional love, I would not be where I am right now.

Abstract

This study introduces an innovative approach for classifying Electroencephalogram (EEG) signals produced when a person focuses on particular words, an area of research known as "Imagined Speech." EEG captures the brain's electrical activity through a set of sensors placed on the scalp, providing insights into neural processes related to thought and cognition. However, accurately classifying these signals poses significant challenges, as it requires machine learning (ML) models to learn the subtle and complex patterns associated with specific imagined words. Traditional classification techniques often struggle in this domain due to high inter-subject variability, low signal-to-noise ratio, and the non-stationary nature of EEG signals. In response, this work proposes a new method to enhance classification accuracy, centered around the creation of a specialized feature vector tailored to improve the model's ability to differentiate between imagined words. The core of this approach is the use of a genetic algorithm that optimizes feature selection by efficiently navigating through a vast feature space. This genetic algorithm is highly advantageous in this setting, as it iteratively refines the feature set to identify the most predictive features, maximizing classification accuracy. Compared to other feature selection methods, genetic algorithms are well-suited for EEG signal processing due to their ability to balance exploration and exploitation, ultimately yielding a feature combination that significantly outperforms existing techniques. This method has the potential to improve the reliability of imagined speech classification, making it a valuable contribution to EEG-based brain-computer interfaces and opening doors for advancements in assistive technology and neural communication.

keywords: EEG, feature extraction, genetic algorithm, classification, electrode selection.

Resumen

Este estudio presenta un enfoque innovador para clasificar las señales de electroencefalograma (EEG) que se producen cuando una persona se concentra en palabras específicas, un área de investigación conocida como "Habla Imaginada". El EEG captura la actividad eléctrica cerebral mediante un conjunto de sensores colocados en el cuero cabelludo, lo que proporciona información sobre los procesos neuronales relacionados con el pensamiento y la cognición. Sin embargo, clasificar con precisión estas señales plantea desafíos significativos, ya que requiere modelos de aprendizaje automático (ML, por sus siglas en lengua inglesa) para aprender los patrones sutiles y complejos asociados con palabras imaginadas específicas. Las técnicas de clasificación tradicionales suelen presentar dificultades en este ámbito debido a la alta variabilidad interindividual, la baja relación señal-ruido y la naturaleza no estacionaria de las señales de EEG. En respuesta, este trabajo propone un nuevo método para mejorar la precisión de la clasificación, centrado en la creación de un vector de características especializado, diseñado para mejorar la capacidad del modelo para diferenciar entre palabras imaginadas. La base de este enfoque es el uso de un algoritmo genético que optimiza la selección de características mediante la navegación eficiente a través de un amplio espacio de características. Este algoritmo genético resulta muy ventajoso en este contexto, ya que refina iterativamente el conjunto de características para identificar las más predictivas, maximizando así la precisión de la clasificación. En comparación con otros métodos de selección de características, los algoritmos genéticos son muy adecuados para el procesamiento de señales de EEG gracias a su capacidad para equilibrar la exploración y la explotación, lo que resulta en una combinación de características que supera significativamente a las técnicas existentes. Este método tiene el potencial de mejorar la fiabilidad de la clasificación del habla imaginada, lo que lo convierte en una valiosa contribución a las interfaces cerebro-computadora basadas en EEG y abre la puerta a avances en tecnología de asistencia y comunicación neuronal.

palabras clave: EEG, extracción de características, algoritmo genético, clasificación, selección de electrodos.

Contents

List of Tables

List of Figures

List of Abbreviations

ACC	Accuracy
ADA	AdaBoost
ANN	Artificial Neural Network
ACO	Ant Colony Optimization
CNN	Convolutional Neural Network
CSP	Common Spatial Patterns
CWT	Continuous Wavelet Transform
dB	decibel
DBN	Deep Belief Network
DL	Deep Learning
DNN	Deep Neural Networks
DT	Decision Tree
DWT	Discrete Wavelet Transform
EEG	Electroencephalogram
ELM	Extreme Learning Machine
FP	False positives
FN	False negatives
GA	Genetic Algorithm
GBM	Gradient Boosting Machine
GRU	Gated Recurrent Unit
Hz	Hertz
kNN	k-Nearest neighbours
LDA	Linear Discriminant Analysis
LogReg	Logistic Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
MLP	Multi Layer perceptron
mRVM	multiclass Relevance Vector Machine
NB	Naïve Bayes
PCA	Principal Component Analysis
QDA	Quadratic Discriminant Analysis
RF	Random Forest
RLDA	Restricted Linear Discriminant Analysis
RNN	Recurrent Neural Networks
SVM	Support Vector Machine
TL	Transfer Learning
TN	True negatives
TP	True positives

1 Introduction

Imagined speech, silent speech, or inner speech is defined as the thought of speech without movement of the mouth, tongue, lips, or hands, that is, thinking about one's own voice uttering words [2].

Understanding this process is important because it gives us a glimpse into internal thoughts and how the brain processes them. Understanding this processing allows for the development of communication skills through brain-computer interfaces (BCI) [3].

By activating neural mechanisms related to thought and memory, it is possible to improve cognitive processes through Imagined Speech. Recent work mentions that it is possible to strengthen neural circuits involved in speech processing through Imagined Speech. This is of particular interest, especially in processes that require computing and retaining information, that is, memory processes. [4].

Research and development in this branch of science have led to efficient therapeutic strategies for people with speech disorders, and it is even possible to target specific neural pathways for particular aspects of speech. These disorders primarily impair the formulation or creation of sounds, without which effective communication is complicated. There are different disorders: on the one hand, we have phonological, vocal, and resonance disorders, while on the other, we have articulatory disorders, which involve difficulty physically producing sounds. This is often related to deficient positioning of the tongue, lips, or jaw. While phonological disorders are caused by a cognitive aspect, where the individual has difficulties recognising the sound rules of their own language, which leads to certain speech lapses. These factors, as expected, affect the individual's quality of life. In particular, pitch, volume, and even problems with the vocal cords are affected by voice disorders, while impairments in sound clarity stem from abnormal airflow through cavities such as the mouth and nose; these are resonance disorders [5][6]

Approximately 1% of the world's population has some type of speech or language disorder. This 1% can be interpreted as millions of people with difficulty expressing themselves and/or understanding others, which affects the way they relate to the world. This difficulty or impairment in these skills is considered a significant obstacle that impacts how these people interact and integrate into society, as well as their educational experience and daily life. [7]

Advances in this line of research are enabling the development of therapeutic breakthroughs with significant improvements. Furthermore, advances in this field are enabling the development of new assistive technologies, such as the aforementioned BCIs, which have the potential to restore and improve communication skills in individuals with speech disorders. These advances remain a research challenge, yet they hold the promise of improving the life quelity of countless individuals affected by these language

and speech disorders [8].

Imagined speech is a phenomenon that has existed since the emergence of language. However, it is primarily associated with research using the detection and processing of EEG signals. This process is a difficult task due to several factors: because it demands cognitive resources, it can lead to fatigue and overload; this task can be interrupted by distractions from other thoughts, and sensory stimuli can add unnecessary noise; Furthermore, the assessment of coherence is affected by the lack of auditory and vocal feedback; differences in the coordination of brain regions among individuals, as well as disparities in the intensity of inner speech, affect the ability to imagine speech. All of the above poses a significant and complex challenge for the processing of imagined speech.

As already mentioned, EEG signals are extremely complex, making EEG signal classification difficult. For this reason, the databases we find are composed of simple vowel structures. For example, the authors [9, 10, 11] report the use of e, a base consisting of five vowels (a, e, i, o, u) in English; they reported an accuracy of 35.7%, 87%, and 79.7%, respectively. While the work by [12] proposes four simple and binary words (Yes, No) in English and (Haan, Na) in Hindi; this work reported a classification accuracy of 73.4%. On the other hand, the work of [13] used two English words (Yes, No) reporting an accuracy of 63.16%.

With this in mind, we can see that the simpler the set of words, the easier it will be for the classification model to perform the task, thus achieving greater accuracy. Conversely, the more complex the set of words, the more complicated the task, which translates into lower accuracy as mention in [1].

On the other hand, within the methodology used to perform this classification, we can find that the state of the art includes the use of various characteristics. For example, in the work of [10], four characteristics are used (mean, variance, standard deviation, and skewness), the work of [11] proposes the use of 32 characteristics, while in the work of [14], the use of up to 44 characteristics is reported.

Having already mentioned the presence of features, let's delve a little deeper into this topic. Due to the complexity of signals, it is extremely important, even necessary, to characterize these signals in order to obtain information about them. This is achieved using a technique known as "Feature Extraction." An EEG signal undergoes changes when one word or another is thought of; these changes are not directly identifiable, which is why the need to employ this characterization technique arises. It should also be noted that the features used will also depend on the signal's nature and the classification model. This method does not have a standard for type and quantity extraction of these features. There are both traditional methods and methods using deep learning techniques for extracting these features. Among the traditional techniques, we can find,

for example, works based on digital signal processing. In the case of Deep Learning techniques, we can find the work of [13] which uses Artificial Neural Networks (ANN) or, for example, the work of [15] which analyzes features using Deep Neural Networks (DNN). Furthermore, within the branch of genetic algorithms, we can find works such as that of [16], where the authors extract features from EEG signals using Ant Colony Optimization (ACO).

It is possible to optimize the selection of these features by using genetic algorithms (GAs). A review of the literature revealed that the work by [17] uses a GA for feature extraction. On the other hand, the work by [18] to achieve optimal classification of epileptic seizures with EEG signals combines GAs with wavelet analysis. Additionally, we can find the work by [19] in which, to identify people, a GA is used to select electrodes from EEG signals.

Among the uses of GAs, we can find in particular the work of [20], in which the authors proposed a modular framework that optimizes descriptor configurations using a GA to vary transformations such as illumination, scaling, and rotation, balancing size, efficiency, and invariance to these transformations. This implementation is achieved by using binary-coded genomes as parameters of the GA, which control the signal processing parameters and the coupling of these descriptors. This implementation also incorporates modules that are selectively activated, achieving a reduction in dimensionality while maintaining invariance to environmental changes in the images. This proposal demonstrated performance that outperformed standard descriptors in the task of image classification. This work, in particular, highlights the optimization and potential of GAs when extracting specific features in particular scenes.

Once a GA has been used to optimize feature selection, the next step is to use a classifier model. These algorithms can be machine learning (ML) or deep learning (DL). Among ML algorithms, we find works such as that of [21], in which the authors used support vector machines (SVMs), or that of [22] or that of [23], which used Naïve Bayes and Random Forest, respectively.

Among the works that report the use of deep learning (DL) algorithms, we find, for example, the work of [13], which reports the use of Recurrent Neural Networks (RNN), or that of [9], where Convolutional Neural Networks (CNN) were used in conjunction with transfer learning (TL). Another example is the work of [24], which reports the use of a combination of CNN+RNN.

Now, it is important to mention the work of [14], in which the author proposes the extraction of 44 statistical features extracted from EEG signals. This served as inspiration for the present work, which proposes the use of 214 features that allow for a more in-depth characterization of EEG signals in imagined speech. After feature extraction, a GA was used to select characteristics with the most substantial data for

entry into the classification algorithm. Within the literature review, it was found that current methodologies report classification accuracies goes from 35.7% to 87%, while the proposed approach shows a considerable increase as reported in chapter 7.

When reviewing the literature it is possible to find trends within the works related to Imagined Speech, this can be observed in figure 1.1 which represents a bibliometric analysis of these trends, in which we can see that the dominant classification models are CNN and SVM, in addition, this analysis shows a clear relationship linking feature extraction, signal processing and EEG signal classification, which indicates a clear tendency towards new and groundbreaking techniques to enhance classification techniques and methods in Imagined Speech.

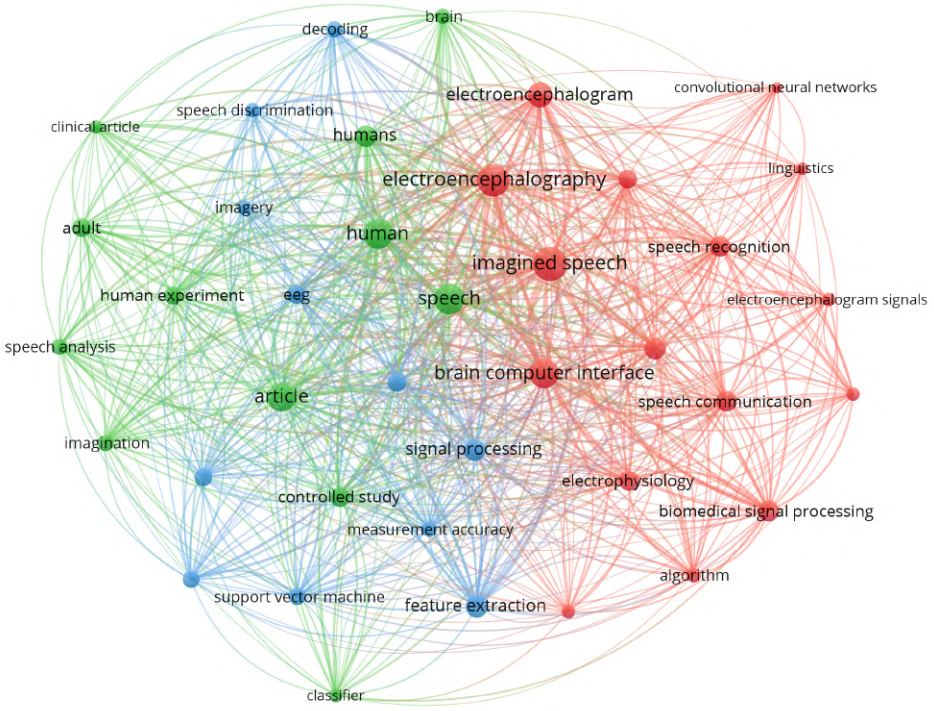


Figure 1.1: Bibliometric network made from Scopus papers that have kewords like "Imagined Speech", "classification" and "EEG".

This work contains the following main contributions:

- Extraction of information from EEG raw signals by generating features.
- The development of a GA from scratch to enable more efficient sorting by finding optimal electrode selection.
- The implementation of said GA with the objective of finding the most relevant features, that is, those that maximize the accuracy in signal classification.

- Ten classifier models were tested to compare their classification performance.

- A dataset was generated with the collected signals, which was published in a public Mendeley Data repository with the aim of enabling research in this branch of knowledge and allowing other techniques to be compared with the present work.

1.1 Justification

Although people around the world suffer from speech disorders caused by illnesses, whether neurological, trauma, or some type of congenital condition, they find that communication is a basic need. These disorders can affect and reduce individuals' quality of life by restricting their ability to work, study, and even socialize. Many individuals with cognitive or motor disorders are unable to provide clear physical information, therefore, text-to-speech systems, eye-tracking communication devices, and conventional assistive technologies are limited by their reliance on this clear physical information.

In particular, thanks to EEG signals, brain-computer interfaces (BCIs) have opened up new opportunities for direct brain-machine communication. Imagined Speech, however, offers an improvement over non-invasive communication systems, as it involves the internal articulation of words without actual vocalization. It is important to emphasize that EEG signals, especially those related to Imagined Speech, are known to be noisy and variable between subjects, which represents a major challenge, as it poses a technical difficulty for consistent classification.

That said, it is also worth mentioning that the success of an Imagined Speech speech recognition system will depend primarily on the efficiency of both feature extraction and selection. This is because these are complex signals, and traditional feature extraction methods are sometimes unable to extract the subtle and complex patterns of these signals. This is where Genetic Algorithms come in. These are a class of evolutionary algorithms motivated by natural selection and offer an interesting approach to feature generation and optimization in this context. Given their rapid search capability in large universes of solutions, they are capable of finding sets of features that maximize the classification accuracy of these signals.

For this reason, this work offers a new methodology for feature generation using a genetic algorithm for classifying imagined speech from EEG signals, with the aim of improving assisted communication for people with speech disorders. This work contributes to both scientific knowledge and practical applications, so the development of this project is situated at the intersection of AI algorithms, human-computer interaction, and biomedical signal processing.

1.2 Problem description

Despite increasing advances and research into imagined speech as a noninvasive method for BCIs, current EEG-based classification systems suffer from low accuracy, poor generalization, and limited practical application. These problems are mainly due to the difficulty in extracting features from EEG signals, since, as mentioned above, these signals are complex and non-stationary. It is important to mention that traditional feature extraction methods that have been implemented are often inflexible and tend to neglect or omit neural patterns associated with imagined speech.

Additionally, individuals with speech disorders may present altered brain activity patterns due to underlying neurological conditions, all of which further complicates the generalization of models. That said, it is critical to employ adaptive and intelligent feature extraction techniques that allow for the identification of unique patterns in EEG signals for optimal Imagined Speech classification.

While numerous studies have employed deep learning and traditional statistical methods for EEG signal processing, there is limited exploration of evolutionary computation, particularly genetic algorithms, to optimize features tailored to imagined speech activities. It is even rare that people focus their efforts on assistive technologies for speech disorders.

However, while numerous studies have implemented deep learning techniques in conjunction with traditional signal processing techniques for EEG signal classification, a very limited population implements evolutionary computation, particularly genetic algorithms, to optimize features tailored to imagined speech activities.

The challenge lies in the effective use of genetic algorithms to create and optimize feature extraction from EEG signals in a way that maximizes the accuracy of classifying these signals in imagined speech.

1.3 Organization

The next sections explain how to achieve classification signals on imagined speech tasks. The structure is the following:

- Chapter 2 includes the researches and previous works.
- Chapter 3 is the conceptual base and the literature review in which this scientific research is based.

  Chapters 2 and 3 contain resource material and literature review that can be similar with the previously published article [1]. That article was derived from the same research project as this thesis, in addition, the reused content has been

properly cited. The inclusion of this material follows institutional guidelines for the use of previously published work in academic theses.

- Chapter 4 is the conjecture underlying the present investigation.
- Chapter 5 includes the general objective and specific objectives as the goals to accomplish in the present work.
- Chapter 6 talks about the dataset, the computing power, the processing signals approaches, the develop of feature engineering, the optimisation of a GA, and the creation and use of classification algorithms.
- In chapter 7 there is an explanation of the evaluation metrics and an analysis of data classification results.
- In chapter 8 the results and forthcoming research on EEG systems for Imagined Speech classification problems are explained and discussed.
- Chapter 9 contains the key points summarized and the most relevant insights from the research are presented.
- Finally, chapter 10 includes documents, such as patient consent statement, social compensation letters and scientific articles published during the investigation.

2 State-of-the-Art

Processing imagined speech is essential due to it help us comprehend the mechanisms by which the brain formulates and regulates internal dialogue, enhances communicative abilities, and facilitates the advancement of brain-computer interfaces.

In the consulted works in the state-of-the-art, there are found that the employed datasets consists on individuals whose average age goes from 20 to 30 years. In the case of the number of individuals there is not an standard number, the average number of individuals is 12. The smallest dataset was used by [11] with only 3 males, while the largest one was used by [25] with 27 individuals (gender not specified).

Another relevant information of the datasets is the sample frequency of the EEG recordings, [26][21][27][28] used a 1kHz frequency. Only [29] employs 500Hz. The works of [12][24][30][10] uses 256Hz. Notice that [24] and [10] resampled his signals from 1kHz to 256Hz. And finally [25] and [9] uses 128Hz ([9] resampled from 1kHz to 128Hz).

In the case of the EEG devices, all the authors used a research device. Only [25] used a commercial device, the Emotiv Epoc+, the same device used in this thesis research project.

Finally, in the last column of Table 2.1 are listed the number of electrodes used by the classifier models. Some works such as [26], [27] and [10] used 64 electrodes. While other works used a device with less number of electrodes, such as [25], [29] and [9] with 14, 20 and 32, respectively. Works like [12], [11], [28], [30], [9] and [23] used a specific number of electrodes based on interest areas according to each author.

The information described is compacted in the Table 2.1.

The Imagined Speech classification task is challenging due to a variety of interrelated factors. It requires a substantial amount of cognitive resources, which can result in a possible overcharge. Interruptions from other thoughts can disturb it, and the inclusion of sensory experiences increase unwanted noise. The absence of auditory and spoken feedback has made measuring coherence difficult. This process is further complicated by the coordination of different brain regions. In addition, the complexity of imagined speech is also affected by how rich each person inner speech is, which leads to difficulty processing EEG signals related to this cognitive task.

Table 2.1: Information of datasets in the state-of-the-art.

Article	**Individuals age**	**Number of individuals**	**Sample frequency**	**EEG device**	**Electrodes**
[26]	27.4 ± 4	8 males 4 females	1 kHz	Neuroscan quik-cap	64
[12]	-	-	250 Hz	EGI Geodesic headset	F7,T7,P7,P3,C3
[29]	-	5	500 Hz	Enebrio dry electrode	20
[21]	22.9 ± 2.3	15 males	1 kHz	EEG-1200 Nihon Kohden	65
[11]	22	3 males	-	g.SAHARABox	FC3,FC5,C5, T7,CP3,CP5, TP7,P5,A2,Nz
[24]	-	14	1 kHz (resampled to 256 Hz)	-	-
[27]	27.6 ± 3.2	20	1 kHz	BrainAmp DC amplifier	64
[25]	-	27	128 Hz	Emotiv Epoc+	14
[28]	-	-	1 kHz	SynAmps RT	C4,FC3,FC1,F5, C3,F7,FT7,Cz, P3,T7,C5
[30]	22-32	11 males 4 females	1 kHz (resampled to 256 Hz)	ActiCHamp	FC5,CP5,F5,FT7, FC3,TP7,CP3,P5
[31]	30 ± 10	5 males 3 females	-	-	32
[9]	-	15	1 kHz (resampled to 128 Hz)	Grass analog 8-18-36	F3,F4,C3, C4,P3,P4
[23]	24.43 ± 1.13	6 males 1 female	-	64ch EEG cap active Ag/AgCl	F3,FC5, FC3,T7, C5,C3,CP5,CP3, P5,FCz,FPz
[10]	28.25 ± 2.71	5 males	1 kHz (resampled to 250 Hz)	Hydrocel geodesic sensor net 64ch	64

This complexity on EEG signals in imagined speech makes the classification in an ardours task. As a result, signal databases consist of simple, imagined speech forms. Authors of [10], [9], [30] and [11] employed a dataset of five vowels (a,e,i,o,u) in English, (except for [9], the vowels are in Spanish) reporting a classification accuracy of -%, 35.68%, 49.9% and 79.7%, correspondingly.

In addition, [12] proposes four words, (Yes, No) in English and (Haan, Na) in Hindi, reporting 87.69%.

The database of [27] reports (Yes, No), achieving a 63.16% accuracy.

There are works that such as [30] and [23], that uses a set of long and complex words, they achieved 66.2% and 34.2%, respectively.

Some works, such as [29] and [31] can be biased, because the individuals did not thought in words. in the case of [29] the experiment consists in think in directional arrows (up, down left, right). In the case of [31], the individuals think in geometric shapes (circle, square, trapezoid). These works achieved an accuracy of 82.5% and 35.78%, respectively.

Table 2.2 concisely presents the state-of-the-art consulted for Imagined Speech task. The table is sorted based on the accuracy (last column) in descent order. The more straightforward the words, the more the classifier model can categorise them and so the more accurate they are. Otherwise, the more complicated the words, the more challenging the job and, hence, the lower the accuracy.

Simple or complex the words, it is necessary to preprocess the signals before apply a classifier model. Therefore, it's common to find works using band-pass filters such as [26], [21], [11], [27], [28], [9] and [23]. Instead of band-pass filter, some works such as [12] and [11] uses a notch filter. Only the work of [24] employs a low-pass filter.

Another common preprocessing technique are the Transformations. The work of [29] presents the use of the Gabor Discrete Transform. In the case of [30] the author uses the Morlet Wavelet Transform. In the case of [27] and [31] the authors employs the Discrete Wavelet Transform.

In the case of the classifiers, the majority of authors employs classic Machine Learning models, such as Support Vector Machines, which is the top used model. Another models employed are Random Forest, k-Nearest Neighbours, Naïve Bayes, Decision Trees and Linear Discriminant Analysis.

In the case of Deep Learning models, there just a few. [26] used a Deep Belief Network (DBN), [12] used an Artificial Neural Network. In the case of Convolutional Neural Networks (CNN), [24] employed a CNN connected to a Long-Short Term Memory (LSTM), while [9] employed a CNN with Transfer Learning (TL). In the case of [27], a Multi Layer Perceptron (MLP) was the model employed. Finally, [28] used a Deep Neural Network (DNN).

The above information is presented in a concise way in the Table 2.2.

In addition to Imagined Speech , there are more mental tasks related to EEG classification, such as motor imagery, epileptic Seizures, drowsiness detection, among others.

Due to the complexity of the signals, it is necessary to characterize them to have information about them, this is possible using the "Feature Extraction" technique. A signal changes when thinking about one word or another, however, these changes cannot be identified directly. Furthermore, the characteristics vary based on the nature of the signals and the algorithm used by the classifier; hence, there is no standard for the sort of features to extract or the number of features to extract.

Taking into consideration the current state of the art, we examine the proposed approaches for accomplishing the categorisation. The authors provide a range of characteristics for use. For instance, the work of [11], [26], [10], [32], [33] and [14] utilized 4, 13, 4, 11, 10 and 43 statistical features, respectively. These statistical features where calculated from the raw EEG signals.

On the other hand, the works of [18], [34], [17] and [35] extracted the features from transformations.
[18] extract features from a 4th level wavelet packet decomposition, while [34] from a db10 wavelet packet transform with a haar wavelet packet transform. The work of [17] used a Wavelet analysis, Hilbert-Huang Transform, and Common Spatial Patterns. Finally, the work of [35] extracted 31 features from a 5-level DWT Daebuchies wavelet.

The above information is presented in a concise way in the Table 2.3.

This thesis research project aims to employ a Genetic Algorithm to optimise feature selection and electrode selection, with the objective of maximising the classifier model's accuracy. This will be detailed in depth in the Chapter 6.

Table 2.2: Relevant information for the classification models in the state-of-the-art.

Article	Inputs (words)	Preprocessing	Models	Accuracy (%)
[26]	/iy/, /uw/, /piy/, /tiy/, /diy/, /m/, /n/)	Band-pass filter, Laplacian	DBN, SVM	88
[12]	'Yes', 'No' (English) 'Haan', 'Na' (Hindi)	Notch Filter	NN, SVM, RF, ADA	87.69
[29]	up, down, left, right (arrows, not words)*	Notch filter, ICA, Gabor Discrete Transform	LDA	82.5
[21]	'ue', 'shita', 'hidari', 'migi', 'mae', 'ushiro' (Japanese)	Band-pass filter, Power Spectral Density	quadratic kernel SVM	81.3
[11]	'a','e','i','o','u', mute word	Band-pass filter, Notch filter, Four statistical features	SVM, DT, LDA-PCA, QDA-PCA	79.70
[24]	'iy', 'piy', 'tiy', 'diy', 'uw', 'm', 'n', 'pat', 'pot', 'knew', 'gnaw'	Low-pass filter	CNN, LSTM, XGBoost	73.45-75.55
[27]	'Yes', 'No'	Butterworth band-pass, ICA, Discrete Wavelet Transform	LDA, SVM, 5-NN, NB, MLP	63.16
[25]	'arriba', 'abajo', 'izquierda', 'derecha', 'seleccionar'*	Butterworth filters	k-Means, Naïve Bayes, TL	62.79
[28]	'iy', 'piy', 'tiy', 'diy', 'uw', 'm', 'n', 'pat', 'pot', 'knew', 'gnaw'	Band-pass filter, Laplacian spatial filter, ocular artifact removal	DNN	57.15
[30]	'in', 'out', 'up', 'cooperate', 'independent', 'a', 'e', 'i', 'o', 'u'	Morlet wavelet transform, ICA, FFT, Common spatial patterns	mRVM	49.9 (vowels) 50.1 (shorts) 66.2 (longs) 80.21 (shorts +longs)
[31]	/circle/,/square/,/trapezoid/*	IIR, DWT, RWE	SVM, RF	35.78
[9]	'a','e','i','o','u' (spanish)	Band-pass , ICA	CNN, TL	35.68
[23]	'ambulance', 'clock', 'hello', 'help me', 'light', 'pain', 'stop', 'thank you', 'toilet', 'TV', 'water', 'yes'	Butterworth Band-pass, CSP	RLDA, Random Forest	34.2
[10]	'a','e','i','o','u'	Four statistical features, sparse regression model	ELM, ELM-R, ELM-L, LDA, SVM-R	-

Table 2.3: Features employed for EEG classification in the state-of-the-art.

Article	Task	Features
[11]	Silent Speech	Average, dispersion, standard differential, skewness
[26]	Articulated and imagined speech	Mean, median, standard deviation, variance, maximum, minimum, maximum ± minimum, sum, energy, spectral entropy, skewness, kurtosis
[10]	Imagined Speech	mean, variance, standard deviation, skewness
[18]	Epileptic seizures	fourth-level wavelet packet decomposition
[34]	Drowsiness detection	db10 wavelet packet transform, haar wavelet packet transform
[32]	Mental imagined tasks	Minimum, Maximum, Mean, Median, Mode, Standard deviation, First quartile, Third quartile, Inter quartile range, skewness, kurtosis
[17]	Motor imagery	Wavelet analysis, Hilbert-Huang Transform, Common Spatial Patterns
[35]	Epileptic seizure	31-features extracted from 5-level DWT Daubechies wavelet
[33]	Emotion classification	Standard Deviation, Mean Absolute Deviation, Median Absolute Deviation, L1-Norm, L2-Norm, Max Norm, Fractional Dimension, Spectral Norm, Power Spectral Density, Spectral Entropy
[14]	Epileptic seizures	43 statistical features

As mentioned above, Genetic Algorithms can be used in EEG signals. Within the state-of-the-art, we can find works like [17], which implements genetic algorithms for feature extraction. One more research is [18] where the author develop a GA combined with wavelet analysis in order to achieve efficient EEG epileptic seizures signal classification. Another work is [34] where the GA is employed to find most accurate cut-off's frequency limits that allow to detect drowsiness. Additional work is [19], authors developed EEG electrode selection for person recognition trough a GA. Table 2.4.

The present work employs EEG signals recorded with an Emotiv Epoc+ EEG device recording with a sample frequency of 128Hz. The feature extraction technique presented in this thesis research project was inspired by [14].

The research of [14] suggests a method of analysing EEG information that extracts 44 statistical features. The present work employs 214 variables, detailed in the methods chapter, that effectively characterise EEG signals from imagined speech with more depth.

A Genetic Algorithm was subsequently employed to determine the most informative characteristics for the classifier model and to identify the optimal electrodes to enhance accuracy.

In the methodology chapter is describe the detailed steps used to classify signals in imagined speech.

Table 2.4: Genetic Algorithm in EEG classification in the state-of-the-art.

Article	**GA utilization**
[17]	Optimize features from Wavelet analysis and Hilbert-Huang Transform using Common Spatial Patterns.
[18]	Find the optimal feature subset from the wavelet coefficients.
[34]	Select the most accurate rhythm (frequency limits) for drowsiness detection.
[19]	Determine the minimum number of electrodes that provide the maximum accuracy for EEG-based person identification.

3 Theoretical Framework

3.1 Disorder speech

A speech disorder refers to a condition where an individual experiences difficulty producing or shaping the speech sounds required for effective communication with others. Recent studies show that up to 1% of the world's population suffers from some degree of difficulty in speech, language or communication needs. There is a neural correlation of imagined speech and its relationship with articulated speech. [36]

The Broca area is a region in the brain, situated in the frontal lobe of the left hemisphere, that houses neurons responsible for speech functions. It serves a vital role in the generation of articulate speech. The Wernicke area, found in the posterior part of the superior temporal gyrus within the dominant hemisphere, includes the auditory cortex situated along the lateral sulcus. (Fig. 3.1).

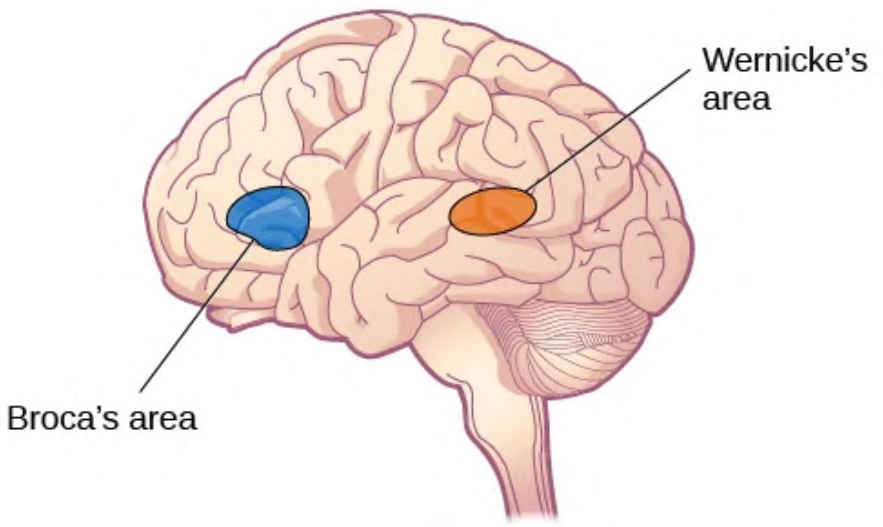


Figure 3.1: Broca's and Wernicke's areas. Source: [36]

3.2 Electroencephalogram

An electroencephalogram (EEG) is a diagnostic test that monitors and records the brain's electrical activity. It involves placing special sensors, known as electrodes, on the head, which are linked to a computer via wires. [37]

3.2.1 Sampling rate

Refers to how quickly the EEG equipment records and displays the brain's electrical activity. This sweep speed determines the amount of time it takes to display a segment of the EEG signal on the screen of the recording device. Sampling rate is commonly

expressed in terms of time per division (T/D), which indicates how much time each division spans on the horizontal axis of the EEG screen. Slower rates can be useful for examining fine details of brain activity, while faster rates can provide a more general view of brain activity over a shorter period of time. [38]

3.2.2 Sensitivity

Sensitivity in the context of electroencephalography (EEG) refers to the ability of EEG recording equipment to detect and record electrical signals generated by brain activity. High sensitivity means that the equipment is able to record even small variations in the brain's electrical activity, while low sensitivity can make these signals less noticeable or even undetectable. It is important to note that too high a sensitivity can result in the capture of electrical noise or artifacts, making accurate interpretation of the EEG signal difficult. On the other hand, too low a sensitivity can miss valuable information from brain activity, affecting the quality of EEG results. [39]

3.3 Brain rhythms

Raw EEG has generally been described in terms of frequency bands (Fig. 3.2) [40], such as:

a. **Gamma** (greater than 30Hz)

b. **Beta** (13-30Hz)

c. **Alpha** (8-13 Hz)

d. **Theta** (4-8 Hz)

e. **Delta** (less than 4 Hz)

3.3.1 Alpha band

Alpha waves have been associated with extroversion (with lower levels observed in introverts), creativity (appearing in creative individuals when solving problems), and mental effort.

- Distribution: Typically regional, often encompassing an entire lobe; strongest in the occipital region when the eyes are closed.
- Subjective Feeling States: Relaxed and calm, yet alert and not drowsy.
- Associated Tasks and Behaviours: Linked to meditation and states of inactivity.

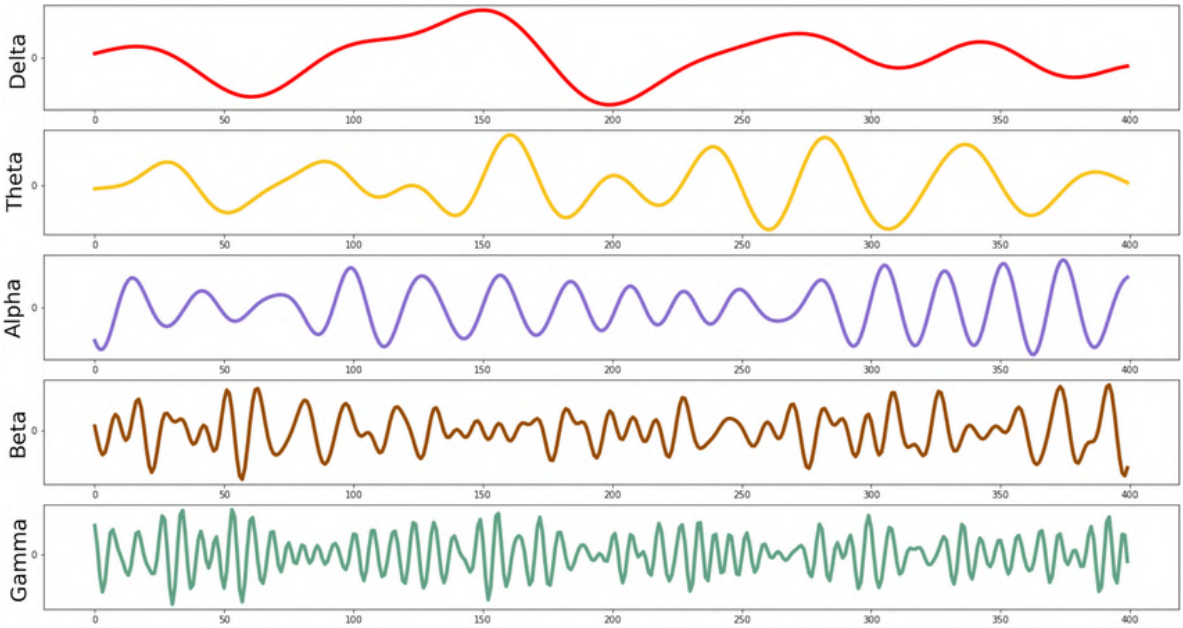


Figure 3.2: Diagram of brain rhythms. Source: [41]

3.3.2 Beta band

The beta band typically appears symmetrically on both sides of the brain, with the most prominent activity in the frontal regions. It may be diminished or absent in areas with cortical damage. Beta activity dominates when the brain is engaged, such as when the eyes are open, during listening, analytical problem-solving, judgment, decision-making, and processing information about the environment.

- Distribution: Localized by hemisphere and lobe (e.g., frontal, occipital).
- Subjective Feeling States: Relaxed yet focused.
- Associated Tasks and Behaviours: Low sensorimotor rhythm (SMR) activity may indicate ADD or difficulties with sustained attention.

Variants of beta bands: Middle Beta (15-18 Hz)

- Distribution: Localized, covering multiple regions, sometimes concentrated at a single electrode.
- Subjective Feeling States: Active thinking, with heightened self-awareness and awareness of surroundings.
- Associated tasks and behaviors: Mental activity
- Physiological correlates: Alert and active, but not agitated

High Beta (above 18 Hz)

- Distribution: Localized, often highly focused.
- Subjective Feeling States: Heightened alertness, sometimes agitation.
- Associated Tasks and Behaviors: Intense mental activity, such as performing calculations or planning.
- Physiological Correlates: General activation of both mental and physical functions.

3.3.3 Gamma band

Gamma activity occurs when the brain processes information from multiple regions simultaneously. It is hypothesized that 40 Hz activity plays a role in coordinating these areas for efficient processing. Well-regulated 40 Hz activity is linked to good memory, whereas deficiencies in 40 Hz activity are associated with learning difficulties.

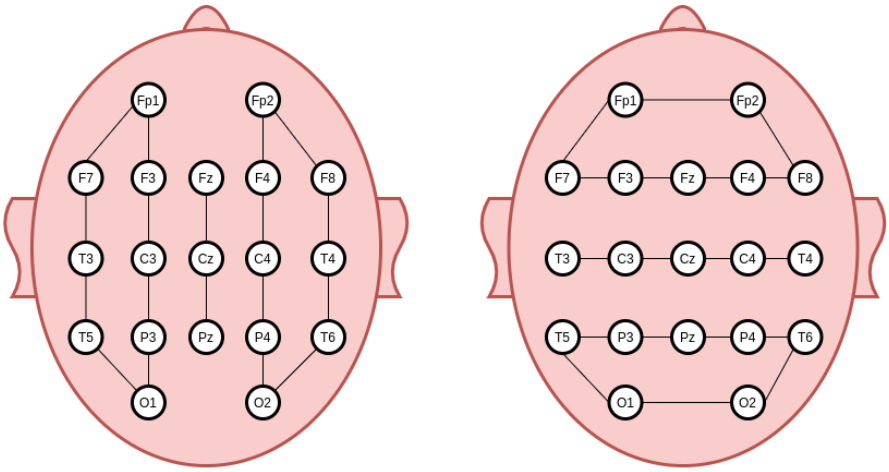
- Distribution: Localized and can be highly focused.
- Subjective Feeling States: Heightened alertness or agitation.
- Associated Tasks and Behaviours: Complex mental activities such as mathematical problem-solving or planning.
- Physiological Correlates: Broad activation of both mental and physical functions.

3.4 Brain activity

As previously mentioned, an electroencephalogram (EEG) is a diagnostic tool that monitors and records the brain's electrical activity. Special sensors, known as electrodes, are placed on the scalp and connected to a computer via wires.

3.4.1 Normal brain activity

In awake adults, the EEG typically displays alpha and beta waves. Both hemispheres of the brain exhibit similar electrical activity patterns, with no abnormal bursts or slow brain waves detected. If flashing lights are used during the test, the occipital region of the brain may briefly respond to each flash, but the overall brain wave activity remains normal.



h

Figure 3.3: A. Longitudinal bipolar montage. B. Transverse bipolar montage. Source: [43]

3.4.2 Abnormal brain activity

When the two hemispheres of the brain show differing electrical patterns, it may indicate an issue in one specific region or side of the brain. The EEG may reveal sudden bursts of electrical activity, known as spikes, or a noticeable slowing of brain waves.

These irregularities can be caused by various conditions, including brain tumors, infections, injury, stroke, or epilepsy.

Abnormal changes in brain waves may also involve more than one area of the brain. Widespread disruptions in brain activity can occur due to conditions like drug poisoning, infections (such as encephalitis), or metabolic disorders (like diabetic ketoacidosis).

Additionally, the EEG may show delta waves or an excess of theta waves in awake adults, which can suggest brain injury or disease. Certain medications can also trigger these patterns. [42]

3.5 Conventional montages

- Bipolar montage (Fig. 3.3):
  - In this montage, each pair of adjacent electrodes is connected so that each serves as a reference for the other. This type of montage highlights the potential difference between nearby points, which can be useful for detecting focal activity and comparing adjacent brain areas.
  - Example: Fp1-F7, F7-T3, T3-T5, etc.
- Reference montage (Fig. 3.4):
  - In this montage, all electrodes are referenced to a single reference electrode or a combination of electrodes (common reference). This allows for a direct

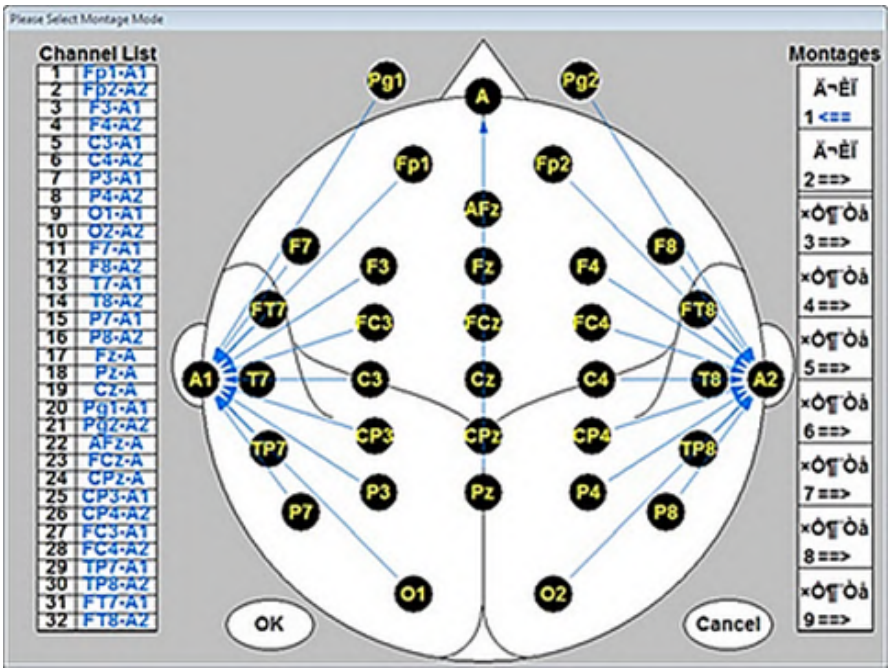
h 

Figure 3.4: Reference montage. Source: [45]

comparison of activity in different brain areas relative to a constant reference point.

- Example: Fp1-Ref, Fp2-Ref, C3-Ref, etc., where "Ref" can be a point such as the A1 electrode (left atrial). [44]

3.6 Standard montage

- International 10-20 System (Fig. 3.5A):
  - This is the most commonly used system for electrode placement in EEG. Electrodes are placed at specific points that are separated by 10% or 20% of the total distance between marked anatomical points (nasion, inion, and preauricular).
  - Electrode positions: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, T3, T4, T5, T6, Fz, Cz, Pz.
- International 10-10 System (Fig. 3.5B):
  - This is an extension of the 10-20 system that includes more electrodes for more detailed recording of brain activity. Intermediate points are used, increasing the number of electrodes and improving spatial resolution.
  - Electrode positions: In addition to those of the 10-20 system, positions such as AF3, AF4, FC1, FC2, CP1, CP2, etc. are included. [46]

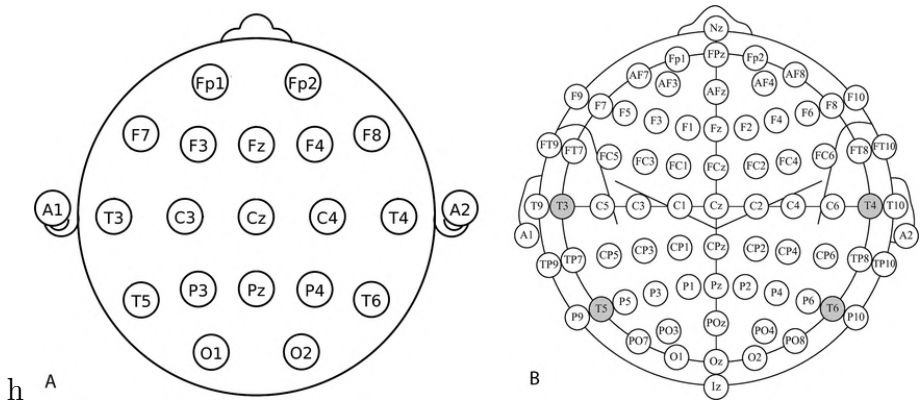


Figure 3.5: A. 10-20 system electrode positions. B. 10-10 system electrode positions. Source: [47]

3.7 10-20 system

The International 10-20 system (Fig. 3.6) is a globally recognized method used to standardize the placement of electrodes in electroencephalography (EEG) based on anatomical landmarks.

Introduced by Herbert Jasper at the 4th International EEG Congress in Brussels in 1957, the system aimed to standardize EEG electrode positioning. The "10" and "20" in the system refer to the distances between adjacent electrodes, which are either 10% or 20% of the total skull distance (anterior-posterior or right-left). This total distance is measured from specific anatomical points on the scalp: the nasion and inion (for the anterior-posterior direction) and the two preauricular points (for the right-left direction).

Each electrode is identified by a name that typically consists of one or more letters and a number. [48]

3.7.1 10-20 system: letters meaning

The letter assigned to each electrode reflects the general brain region it corresponds to. The electrode labeling from front to back is as follows: Fp (prefrontal or frontal pole), F (frontal), C (midline of the brain), T (temporal), P (parietal), and O (occipital).

Additionally, M and A are sometimes used to refer to the mastoids and earlobes, respectively. These locations are often included as reference points for off-line signal analysis.

3.7.2 10-20 system: numbers meaning

The electrode number indicates its distance from the brain's midline. Electrodes placed at the midline are labeled with a "z" to represent zero. As you move away from the midline, the electrode numbers increase. Odd-numbered electrodes are located in the left hemisphere, while even-numbered ones are in the right hemisphere. (Fig. 3.7) [37]

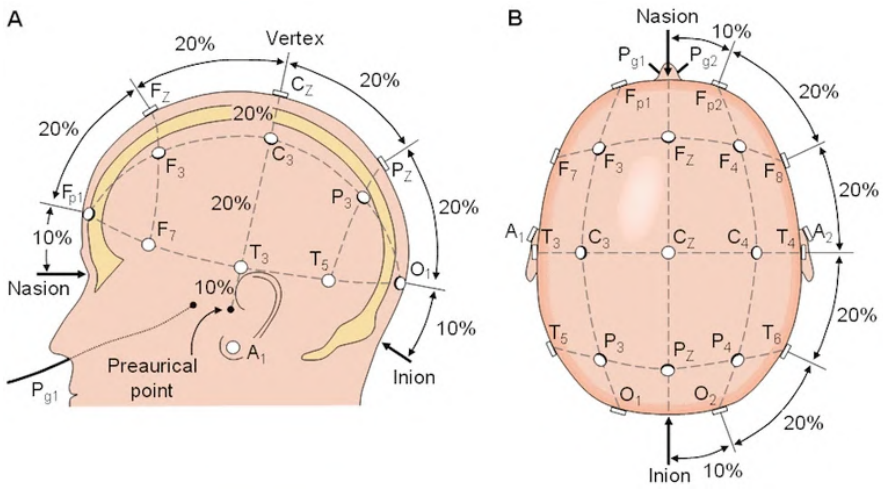


Figure 3.6: Diagram of 10-20 system. Source: [37]

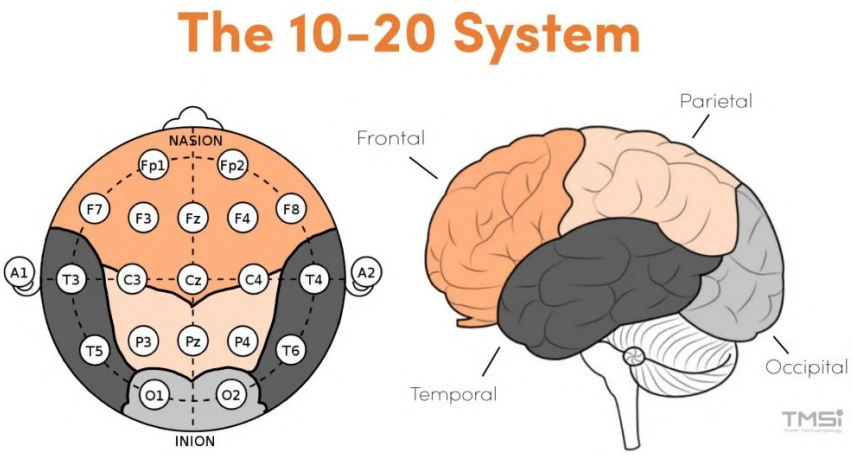


Figure 3.7: Letters and numbers in 10-20 system. Source: [37]

3.8 Neuroimaging devices

There are different devices such as:

- Electroencephalogram (EEG)
- Electrocorticography (ECoG)
- Functional magnetic resonance imaging (fMRI)
- Functional Near Infrared Spectroscopy (fNIR)

3.8.1 EEG's pros

a. It is cheaper and non-invasive

b. Has good temporal resolution, although Electrocorticography (ECoG) has better

c. Setup preparation time can be reduced by using dry electrodes instead of gel-based electrodes

Despite the positive points mentioned above, it is important to mention the weak points.

3.8.2 EEG's cons

a. A low signal-to-noise ratio compared to other modalities

b. Limited spectral and spatial resolution

c. It is difficult to record for a long time using gel or saline-based electrodes because they can dry out

d. A trained person is required to place the device

3.9 EEG device and software

3.9.1 EEG device

The signals were recorded using a Emotiv Epoc+, this device is a headset (Fig. 3.8) and have the following specifications:

- EEG signals
  - A Sinc digital filter of 5th order is included
  - Employs a single ADC in order to generate sequential samples.
  - The samples are can be 128/256 SPS (128 sps was the allowed by the software license bought), the internal sampling rate is 2048 SPS.
  - $8400\mu V(pp)$ as dynamic range
  - At 50Hz and 60Hz a nothc filter is included, while in 0.16Hz to 43Hz range is the bandwidth.
- Sensitivity
  - Resolution: 14 bits with 1 LSB = $0.51\mu V$ (16 bit ADC, 2 bits instrumental noise floor omitted), or 16 bits (user-configurable)
  - The device's sensitivity can detect voltage changes as small as a half microvolt approximately, which is suitable for capturing EEG signals that are usually between 10 and 100 μV.
- Connectivity

- Wireless: Low Energy Bluetooth
- USB dongle: employs a frequency band of 2.4GHz

- EEG sensors:
  - 14 channels: AF4, F8, F4, FC6, T8, P8, O2, O1, P7, T7, FC5, F3, F7, AF3
  - references: P4, P3
  - Sensor material: Felt cushions soaked in saline solution

Figure 3.8: Emotiv Epoc+ headset

3.9.2 EEG software

To connect the Emotiv Epoc+ headsset it is necessary to employ an specific woftware, the one provided by Emotiv. The Emotiv Pro software (Fig. 3.9), which standard licence was bought thanks to the CONAHCYT scholarship.

3.10 Artifacts

Artifacts in an electroencephalogram (EEG) are unwanted signals that are superimposed on the true brain signals, making their correct interpretation difficult. These artifacts can originate from a variety of sources, both internal (physiological) and external (non-physiological). [49][50]

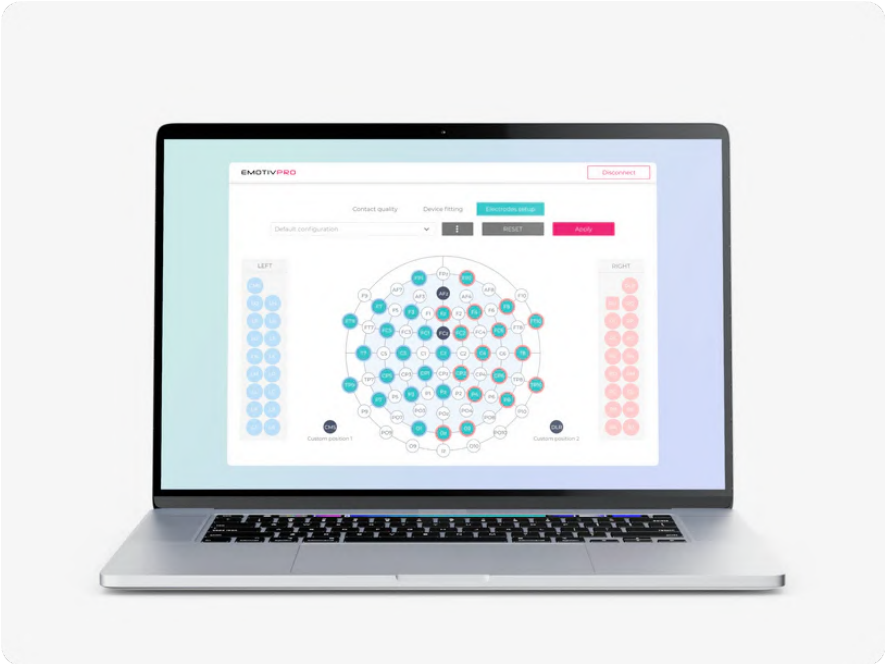


Figure 3.9: Emotiv PRO software

3.10.1 Physiological artifacts

- Eye movements (Electrooculograms, EOG): Blinking movements: Produce large waves in the frontal regions of the EEG. Lateral eye movements: Generate signals in the temporal leads.
- Muscle activity (Electromyograms, EMG): Jaw movement, chewing, or muscle tension in the face and neck: May cause rapid, high-frequency signals.
- Cardiac activity (Electrocardiograms, ECG): Heartbeats: May induce small rhythmic waves, especially in recordings near the neck or chest.
- Breathing: Movements of the chest and abdomen during breathing: May cause base shifts and oscillations in the EEG signals.

3.10.2 Non-physiological artifacts

- Electrical interference: 60 Hz (or 50 Hz in some regions) from power grids: Produces high-frequency waves, usually recognizable by their periodicity and specific frequency.
- Electrodes: Poor electrode contact: Results in motion artifacts or fluctuations in impedance, which can generate random noise or spikes.
- Cable movements: Electricity cable movement: May cause signal fluctuations.

- Electronic devices: Medical equipment and other electronic devices nearby: May generate electromagnetic interference. [51]

3.10.3 Thermal Artifacts

- Environmental Conditions: Hot Environment: High ambient temperature may cause excessive sweating in the patient, which may affect the impedance of the electrodes and create artifacts.
- Monitoring Equipment: Nearby Heat-Emitting Medical Devices: Heat-generating equipment near the EEG recording area may introduce thermal noise.
- Body Temperature Fluctuations: Fever or Hyperthermia: Changes in the patient's body temperature may alter the electrical properties of the skin and electrodes.

3.11 Identification and Management of Artifacts

- Patient Preparation: Inform the patient about the importance of minimizing movement and relaxing muscles during recording.
- Proper Electrode Placement: Ensure good electrode contact and stability by cleaning the skin and using conductive gels.
- Signal Filtering and Processing: Use low-pass, band-pass, and notch filters to eliminate specific unwanted frequencies.
- Visual and Automated Analysis: Technicians and algorithms can identify and correct artifacts, sometimes even removing them using signal processing techniques. [52]

3.12 Laterality

- Left hemisphere dominance:
  - Language processing: In language-related tasks, the left hemisphere is dominant, both for most right-handed people and many left-handed people. This includes both speech production (Broca's area) and comprehension (Wernicke's area).
  - Electrode focus: Emphasize electrode placement over the left hemisphere, particularly F3, FC5, and T7, to capture activity related to inner speech. [53]
- Right hemisphere contribution:

- Prosody and context: The right hemisphere plays a role in processing prosodic (intonation and rhythm) and contextual aspects of language, which may also be relevant to inner speech.
- Balanced focus: While focusing on the left hemisphere, also include electrodes over the right hemisphere (F4, FC6, T8) to capture a complete picture of brain activity. [54]

3.13 Digital Filters

Digital filters are algorithms or electronic systems that process digital signals to modify the information contained in them. Unlike analog filters, which manipulate continuous electrical signals, digital filters work with discrete samples of the signal. [55]

3.13.1 Low-pass filter

- It is a type of electronic filter that allows signals with a frequency lower than a certain cut-off frequency to pass, while attenuating or blocking signals with a frequency higher than that frequency.
- It is used to remove or reduce the high-frequency components of a signal, allowing only the lower frequencies to pass through.
- It is commonly used in audio applications to remove high-frequency noise or to smooth out signals. [55]

3.13.2 High-pass filter

- Unlike the low-frequency filter, the high-frequency filter allows frequencies above a certain cutoff frequency to pass and attenuates lower frequencies.
- It is used to attenuate the low-frequency components of a signal, allowing only the higher frequencies to pass through.
- It is useful in applications such as low-frequency noise removal or signal separation. [55]

3.13.3 Band-pass filter

Is an electronic or digital filter that allows signals within a certain frequency range to pass through, while attenuating frequencies outside this range. Essentially, it transmits frequencies between two specified cut-off points, known as the lower and upper cutoff frequencies, while filtering out the frequencies below the lower cutoff (low frequencies) and above the upper cutoff (high frequencies).

Key Characteristics:

- Center Frequency: The frequency in the middle of the passband, which is most effectively transmitted.

- Bandwidth: The range between the lower and upper cutoff frequencies, defining the width of the passband.

- Applications: Band-pass filters are commonly used in communication systems, audio processing, and biomedical devices like ECG monitors to isolate particular frequency components of a signal. [56]

3.13.4 Band-stop filter

A stop-band filter, also known as a band-stop filter, is designed to reject or attenuate frequencies within a specific range, while allowing frequencies outside this range to pass. It is the inverse of a band-pass filter.

Key Characteristics:

- Stopband: The range of frequencies that are rejected.

- Passbands: The frequencies outside the stopband that are allowed to pass through.

- Applications: Stop-band filters are often used in noise reduction and interference rejection. For example, they are used in audio systems to eliminate unwanted noise at certain frequencies. [57]

3.13.5 Notch filter

A notch filter, also known as a band-reject filter or band-stop filter, is a type of filter that attenuates a very narrow range of frequencies while allowing all other frequencies to pass. It is specifically designed to eliminate a narrow band of unwanted frequencies, such as a single harmonic or noise source, without affecting the rest of the signal spectrum.

Key Characteristics:

- Narrow Bandwidth: The frequency range it attenuates is much narrower compared to standard band-stop filters.

- Sharp Attenuation: The filter's transition between attenuation and passband is sharp, meaning it only affects the specified frequency.

- Applications: Notch filters are commonly used to remove power line interference in biomedical signals or to eliminate narrowband interference in communication systems. [58][59]

3.14 Time-series Characterisation

Is the process of identifying and describing the fundamental properties of a time-series, in this case, an EEG signal. The properties are such as frequency, amplitude, phase, and temporal behaviour. In signal processing, signals are typically classified based on various parameters like time-domain, frequency-domain, periodicity, and determinism, those parameters help in understanding, analysing and processing them.

3.14.1 Time-domain characterisation

In the time domain, signals are analysed with respect to how they vary over time. Common characteristics include:

- Amplitude: The maximum or peak value of the signal. It indicates the strength or intensity of the signal at any given time.
- Duration: The length of time for which the signal exists. This could be finite (e.g., a pulse) or infinite (e.g., a continuous waveform).
- Waveform: The shape of the signal as it evolves over time. Common waveforms include sinusoidal, square, triangular, and pulse signals.
- Energy and Power:
  - Energy signals: Signals that have finite energy but infinite duration.
  - Power signals: Signals with infinite energy but finite power over a period of time.

Mathematically, the energy E and power P of a signal $x(t)$ are defined as Eqs. (1-2) [55]:

$$E = \int_{-\infty}^{\infty} |x(t)|^2 \, dt \tag{1}$$

$$P = lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |x(t)|^2 \, dt \tag{2}$$

3.14.2 Frequency-domain characterisation

In this domain, signals are analysed based on their frequency components. This domain provides information related to the signal's periodicity and frequency content.

- Frequency Spectrum: The decomposition of a signal into its constituent sinusoidal components using Fourier Transform. It reveals the different frequencies present in the signal and their respective magnitudes.

- Bandwidth: The range of frequencies that a signal occupies. Signals with higher bandwidth can carry more information but may require more power or be prone to noise.
- Phase: The relative shift in the time domain of a sinusoidal signal. Phase shifts can indicate delays or alignments between different frequency components.
- Harmonics: Signals, especially periodic ones, often have harmonic components integer multiples of a fundamental frequency. The presence of harmonics affects the signal's complexity and spectral content. [56]

3.14.3 Deterministic vs. Stochastic Signals

- Deterministic Signals: Can be described mathematically with precision. Their behaviour is predictable over time, and they have no randomness involved.
- Stochastic (or Random) Signals: Exhibit some degree of randomness, and their exact behaviour cannot be predicted precisely. However, their statistical properties, such as mean and variance, can be described. [57]

3.14.4 Stationary vs. Non-Stationary signals

- Stationary Signals: Their statistical properties (e.g., mean, variance) do not change over time. Stationary signals are often easier to analyse because their behaviour is consistent over time. Example: White noise, a constant sine wave.
- Non-Stationary Signals: Their statistical properties change over time, making them more challenging to analyse. Many real-world signals, such as speech or ECG, fall into this category. Example: Speech signals, heart rate signals (ECG). [60]

3.15 Features to characterise signals

This section presents key equations used to extract relevant information form the signals, enabling their effective characterisation for input into classification models. Classification models are described in Chapter 3.16.

3.15.1 Mean

The mean, Eq. (3), is a metric in the field of statistics that states the average value of a data set. [61]

$$\text{Mean} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3}$$

where: each frame sample is x_i and the amount of frames is N.

3.15.2 Standard deviation

Standard deviation, Eq. (4), represents a quantification that measures the variability in a data set around its mean, also known as quantifying the dispersion of the data set.[62]

$$\text{Standard Deviation} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \text{Mean})^2} \tag{4}$$

where: frame sample is x_i, the amount of frames N and the average is Mean.

3.15.3 Coefficient variation

The coefficient of variation, Eq. (5) represents the dispersion of the data points, normalized to the mean. [63]

$$\text{Coefficient of Variation (CV)} = \frac{\text{Standard Deviation}}{\text{Mean}} \tag{5}$$

3.15.4 Median

The median, Eq. (6), in an organized set constitute the central data point . [62]

$$\text{Median} = \begin{cases} \text{middle value} & \text{if the number of data points is odd,} \\ \frac{\text{value at position } \frac{N}{2}+1+\text{value at position } \frac{N}{2}}{2} & \text{if the number of data points is even.} \end{cases} \tag{6}$$

3.15.5 Mode

The mode, Eq. (7), represents the data point that has the greatest incidence in a data set. [62]

$$\text{Mode} = \text{Value with greatest incident} \tag{7}$$

3.15.6 Max

The max (maximum), Eq. (8), a value that represents the highest value.[64]

$$\text{Max} = \max(x_1, x_2, \ldots, x_N) \tag{8}$$

3.15.7 Min

The min (minimum), Eq. (9), a value that represents the lowest value. [64]

$$\text{Min} = \min(x_1, x_2, \ldots, x_N) \tag{9}$$

3.15.8 First quartile

The first quartile (Q1), Eq. (10), it refers to the value that is below 25% in an ordered data set. [65]

$$Q1 = \text{Median of the lower half of the dataset} \tag{10}$$

3.15.9 Third quartile

The third quartile (Q3), Eq. (11), it refers to the value that is above the 25% in an ordered data set. [65]

$$Q3 = \text{Median of the upper half of the dataset} \tag{11}$$

3.15.10 Inter quartile range

The interquartile range (IQR), Eq. (12), this is the difference between Q1 and Q3.[65]

$$IQR = Q3 - Q1 \tag{12}$$

3.15.11 Kurtosis

Kurtosis, Eq. (13), this is a statistical metric that illustrates the distribution of data between the data center and the queues. [62]

$$\text{Kurtosis} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \text{Mean}}{\text{Standard Deviation}}\right)^4 - 3 \tag{13}$$

3.15.12 Skewness

Skewness, Eq. (14), it indicates the presence of bias in a data set by measuring the asymmetry in the distribution of a data set. [62]

$$\text{Skewness} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \text{Mean}}{\text{Standard Deviation}}\right)^3 \tag{14}$$

3.15.13 Detrended Fluctuation Analysis

Detrended Fluctuation Analysis (DFA), Eq. (15), allows the identification of signals autocorrelation.

$$F(\tau) \sim \tau^H \tag{15}$$

where:

- $F(\tau)$ is the fluctuation function that describes the signal's variance over a window of size τ,
- H is the Hurst parameter, which characterizes the scaling behaviour,
- τ is the window size over which the fluctuations are calculated. [66]

3.15.14 Activity Hjorth param

Activity, Eq. (16), one of the parameters established by Hjorth, quantifies the intensity of the signal, representing brain activity.

$$\text{Activity} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 \tag{16}$$

where: signal amplitude at one point is x_i while the frames in a signal is denoted by N. [67]

3.15.15 Mobility Hjorth param

Motility, Eq. (17), a parameter described by Hjorth, quantifies the average frequency and establishes correlations between brain activity and neuronal processes.

$$\text{Mobility} = \frac{\sqrt{\text{Var}(dx)}}{\sqrt{\text{Var}(x)}} \tag{17}$$

where: the EEG signal is x, first derivative is dx, variance is $\text{Var}(x)$ and first derivative variance is $\text{Var}(dx)$. [67]

3.15.16 Complexity Hjorth param

Complexity, Eq, (18), one of the parameters established by Hjorth, quantifies the variation in frequency.

$$\text{Complexity} = \frac{\text{Mobility}(dx)}{\text{Mobility}(x)} \tag{18}$$

where: the EEG signal is x, first derivative is dx and Mobility corresponds to the signal's mobility. [67]

3.15.17 Permutation entropy

Permutation Entropy (PE), Eq. (19), captures the values relations, facilitating to extract a probability distribution of the ordinal patters.

$$PE = -\sum_{p} P(\pi_p)\log(P(\pi_p)) \tag{19}$$

where: π_p is an ordinal pattern and the probability of each pattern is $P(\pi_p)$. [68]

3.15.18 Approximate entropy

Approximate entropy (ApEn), Eq. (20), assesses the degree of stability and uncertainty of variations in a set of time series data.

Equation: For a signal of length N, with m as the embedding dimension, and r the tolerance, the Approximate Entropy $ApEn$ is denoted as follows:

$$ApEn(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \tag{20}$$

Where $\phi^m(r)$ is:

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log\left(\frac{\text{number of patterns within tolerance } r \text{ of } X_i^m}{N-m+1}\right) \tag{21}$$

where: at the i-time-series starting position, an m-length segment is denoted by X_i^m. [69]

3.15.19 Spectral entropy

Spectral entropy (SE), Eq. (22), refers to the calculation of the distribution and predictability of spectral power, based on Shannon entropy.

$$SE = -\sum_{i=1}^{N} P(f_i) \log(P(f_i)) \tag{22}$$

where $P(f)$ is denoted as the Power Spectral Density:

$$P(f) = \frac{|X(f)|^2}{\sum_{i=1}^{N} |X(f_i)|^2} \tag{23}$$

[70]

3.15.20 Higuchi fractal dimmension

The Higuchi Fractal Dimension (HFD) method, Eq. (24), is used to count boxes in the dimension of a time series graph.

$$HFD = \text{slope of the line fitting } (\log(k), \log(L(k))) \tag{24}$$

where:

$$L(k) = \frac{1}{k} \sum_{m=1}^{k} L_m(k) \tag{25}$$

in particular:

$$L_m(k) = \frac{1}{\left(\frac{N-m}{k}\right) k} \sum_{i=1}^{\left(\frac{N-m}{k}\right)} |x_{m+i\cdot k} - x_{m+(i-1)\cdot k}| \tag{26}$$

[71]

3.15.21 Total power spectral density

Total Power Spectral Density (PSD), as defined in Eq. (27), quantifies the power in a time-series or a signal in relation to frequency.

$$\mathrm{PSD}_{\mathrm{total}} = \sum_{i=1}^{N} P(f_i) \tag{27}$$

where: - $P(f_i)$ denotes the power at each frequency f_i, - N signifies the total count of frequency bins. [72]

3.15.22 Centroid power spectral density

The centroid power spectral density (centroid PSD), as defined in Eq. (28), denotes the position of the centre of mass within the associated power spectrum.

$$f_{\mathrm{centroid}} = \frac{\sum_{i=1}^{N} f_i \cdot P(f_i)}{\sum_{i=1}^{N} P(f_i)} \tag{28}$$

where:

- N indicates the total number of frequency bins
- $P(f_i)$ signifies the power at frequency f_i
- f_i denotes the i-th bin frequency,

[73]

3.15.23 Determinism

Determinism (DET), es defined in Eq. (29), assesses the extent of repeating structures within a time series data sample and reveals patterns arranged in a systematic order.

$$DET = \frac{\sum_{l=l_{\min}}^{N} l \cdot P(l)}{\sum_{i,j} R(i,j)} \tag{29}$$

[74]

3.15.24 Trapping time

Trapping Time (TT), Eq. (30), assesses the duration during which the signal remains in a repetitive state, i.e., occurrence, which is beneficial for understanding brain functions, i.e. neural activities.

$$TT = \frac{1}{M} \sum_{i=1}^{M} T_i \tag{30}$$

[74]

3.15.25 Diagonal line entropy

Diagonal Line Entropy (DLE), as defined in Eq. (31), quantifies the complexity of cerebral by assessing repeated patterns. It calculates the entropy of the distribution of diagonal line lengths.

$$DLE = -\sum_{l=1}^{L} P(l) \log(P(l)) \tag{31}$$

[74]

3.15.26 Average diagonal line length

Average Diagonal Line Length (ADLL), Eq. (32), represented by the equation (32), assesses the constancy or duration of specific patterns of activity in the brain.

$$ADLL = \frac{1}{M} \sum_{i=1}^{M} l_i \tag{32}$$

[74]

3.15.27 Recurrence rate

Recurrence Rate (RR), Eq. (33), assesses the consistency of signals by determining the frequency with which the system reverts to analogous conditions.

$$RR = \frac{\sum_{i,j} R(i,j)}{N^2} \tag{33}$$

[74]

3.15.28 Spectral edge frequency 25

Spectral Edge Frequency 25 (SEF 25), Eq. (34), indicates the frequency that includes 25% of the total power of a signal.

$$\int_0^{f_{SEF25}} P(f)df = 0.25 \sum_{f=0}^{f_{\max}} P(f) \tag{34}$$

[75]

3.15.29 Spectral edge frequency 50

Spectral Edge Frequency 50 (SEF 50), Eq. (35), denotes the frequency at which 50% of the total power of a signal is encompassed.

$$\int_0^{f_{SEF50}} P(f)df = 0.50 \sum_{f=0}^{f_{\max}} P(f) \tag{35}$$

[75]

3.15.30 Spectral edge frequency 75

Spectral Edge Frequency 75 (SEF 75), Eq. (36), denotes the frequency encompassing 75% of overall power of a signal.

$$\int_0^{f_{SEF75}} P(f)df = 0.75 \sum_{f=0}^{f_{\max}} P(f) \tag{36}$$

[75]

3.15.31 Hurst exponent

The Hurst exponent (H), Eq. (37), is obtained by analysing autocorrelations in a sample of time series data to assess the persistence, unpredictability, or anti-persistence inherent in a signal.

$$R/S \sim n^H \tag{37}$$

The cumulative sum range is denoted by$R(n)$, for a time-serie window size n the standard deviation is denoted by $S(n)$, and finally, the slope of the log-log plot of R/S versus n is denoted as the Hurst exponent H. [76]

3.15.32 Singular valued decomposition entropy

Singular Value Decomposition Entropy (SVDE), Eq. (**??**), evaluates the information structure of a signal by measuring the uncertainty or disorder in the distribution of a time series data collection.

$$SVDE = -\sum_{i=1}^{k} p(\sigma_i) \log(p(\sigma_i)) \tag{38}$$

where $p(\sigma_i)$ denotes the normalised probability distribution of the singular values σ_i, and k represents the quantity of singular values included in the decomposition, often aligning with the rank of the matrix X). [77]

3.15.33 Petrosian fractal dimension

The Petrosian fractal dimension (PFD), represented by Eq. (39), evaluates the characteristics of brain activity, specifically recognizing and differentiating states of physiological functioning.

$$PFD = \log_2 \left(\frac{N_{sc}}{N - N_{sc}} \right) + 1 \tag{39}$$

where:

$$N_{sc} = \sum_{i=2}^{N} \mathbf{1}_{\{\text{sgn}(x_i - x_{i-1}) \neq \text{sgn}(x_{i-1} - x_{i-2})\}} \tag{40}$$

N_{sc} denotes the frequency with which the signal intersects its local mean. [78]

3.15.34 Katz fractal dimension

The Katz fractal dimension (KFD), represented by Eq. (41), evaluates and contrasts complex waveforms by quantifying the lack of uniformity or sophistication in brain activity, making it easier to differentiate between various mental states.

$$KFD = \log_{10} \left(\frac{L}{d_0} \right) / \log_{10}(N) \tag{41}$$

where:

$$L = \sum_{i=2}^{N} |x_i - x_{i-1}| \tag{42}$$

L represents thee cumulative length of the time series. [79]

3.15.35 Relative band power

Relative band power, represented by Eq. (43), assesses how power is distributed across different brain wave frequency bands associated with diverse cognitive and mental states.

$$R_b = \frac{P_b}{P_{\text{total}}} \tag{43}$$

where:

- R_b deenotes the relative band power for band b
- P_b represents the power inside frequency band b
- P_{total} signifies the total power across all frequencies.

[80]

3.15.36 Band amplitude

The measurement of the time series in a specific frequency interval is evaluated by the bandwidth, and is represented by Eq. (45).

$$A_b(t) = \sqrt{x_b(t)^2 + \dot{x_b}(t)^2} \tag{44}$$

$$A_b = \frac{1}{T}\int_0^T A_b(t)dt \tag{45}$$

where $A_b(t)$ represents the instantaneous amplitude of the signal, while A_b denotes the average amplitude within frequency band b. [81]

3.16 Introduction to Machine Learning

According to [82], Machine Learning (ML) is how to build algorithms that are able to learn from experience, in other words, that automatically improve through observing patterns in the data.

3.16.1 Key Components of Machine Learning

Describe the building blocks of ML models:

- Data: The Machine Learning algorithms are highly correlated to the quality and the quantity of data. The algorithm won´t perform well if the data is noisy, incomplete or irrelevant, no matter how sophisticated the algorithm is. The role of data preprocessing is as important as the algorithm selection, it involves cleaning, normalizing, and filtering data to ensure quality. These steps are critical for effective learning. Better data generally leads to better models, making it crucial to have accurate, representative, and relevant data.

- Features: Also known as feature extraction or feature engineering. It refers to the process of taking raw data, transforming it into a format that can be more easily understood by ML algorithms. Feature engineering is crucial for improving model performance. Feature extraction involves selecting the features (or variables) that capture the essence of the problem. Good features can make it easier for the learning algorithm to discover patterns. The process can be manual or automated, but the main idea is that not all features in the data are equally important, and extracting or creating informative ones improves model accuracy. [82]

- Types of learning algorithms:

  - Supervised Learning: The machine is trained using labelled data, meaning the algorithm is provided with input-output pairs. Here the goal of the algorithm is to learn a function that maps inputs to desired outputs. Examples include classification (like decision trees, neural networks) and regression algorithms (e.g., linear regression).
  - Unsupervised Learning: Learning from data without labeled outputs. The goal is to discover hidden structures within the data. Clustering algorithms like K-means and dimensionality reduction techniques like Principal Component Analysis (PCA) are examples of unsupervised learning algorithms.
  - Reinforcement Learning: This type of learning involves an agent that learns to make decisions by interacting with an environment. The agent is rewarded or penalized for its actions, and its goal is to maximize cumulative rewards. Reinforcement learning involves exploration and exploitation trade-offs and is more complex than supervised or unsupervised learning. [82]

- Model Evaluation: It is essential in Machine Learning to evaluate how well a model generalizes to new data. Accuracy (Eq. (46)) is the most basic measure, but there are more detailed methods like:

  - Precision, Recall, and F1-Score: These are other common metrics used to evaluate models, especially in classification tasks. (Eqs. (47-49))
  - Cross-validation: A technique to assess model performance by dividing the data into training and test sets multiple times.
  - Overfitting and Underfitting: Concepts in Machine Learning, where overfit is when the algorithm memorize the training data too closely, while underfit is when the model fail to capture underlying patterns. These concepts are crucial for model evaluation. [82]

3.17 Introduction to Deep Learning

Deep learning (DL) is a branch of machine learning that focuses on the use of neural networks with different layers that automatically learn patterns and features from large

amounts of data. Unlike traditional ML models that rely heavily on manual feature extraction, deep learning models can learn hierarchical representations, where higher layers capture characteristics of the input data. This makes DL algorithms more effective in complex tasks such as signal and image recognition, as well as natural language processing and autonomous systems. With the increased availability of large datasets and computing power (e.g., GPUs), deep learning has become the driving force behind major advances in Artificial Intelligence. [83]

3.17.1 Structure of Neural Networks

- Neural Networks: Neural networks are computational models inspired by the human brain's structure. The simplest neural network is the Perceptron, a single-layer network used for binary classification. However, modern applications require multi-layer networks (also known as multi-layer perceptrons or MLPs), which consist of multiple layers of interconnected neurons. Each layer processes the output of the previous layer, allowing the network to learn increasingly complex patterns. The first layer takes raw input (like images or text), and successive layers (hidden layers) extract features automatically, culminating in the final output layer, which makes predictions. These multi-layer networks are foundational to deep learning architectures.

- Activation Functions: Activation functions in neural networks determine the output of neurons and introduce non-linearity, enabling the network to learn complex mappings. Some commonly used activation functions include ReLU (Rectified Linear Unit), which outputs the input if positive and zero otherwise. ReLU is widely used because it helps mitigate the vanishing gradient problem and speeds up training. Another function is the Sigmoid, which squashes outputs between 0 and 1, making it useful for binary classification but prone to vanishing gradients in deep networks. Other functions like Tanh (hyperbolic tangent), Leaky ReLU, and Softmax are also employed in various layers depending on the task and network architecture.

- Training Process: Training neural networks involves updating the weights of connections between neurons to minimize the error in predictions. This is achieved using backpropagation, a method where errors are propagated backward from the output layer to adjust the weights layer by layer. Backpropagation calculates the gradient of the loss function concerning each weight using the chain rule. These gradients are then used by gradient descent, an optimization algorithm that updates the weights in the direction that reduces the error. Variants of gradient descent, such as stochastic gradient descent (SGD) and mini-batch gradient descent, are often used to make the process faster and more efficient, especially when dealing with large datasets. [83]

3.17.2 Popular Deep Learning Architectures

- CNNs: Convolutional Neural Networks (CNNs) are a specialized deep learning architecture designed to process structured grid-like data, such as images. CNNs use convolutional layers to automatically learn spatial hierarchies by applying filters to capture features like edges, textures, and more complex patterns at higher layers. Key components of CNNs include convolutional layers, pooling layers (which down-sample the data to reduce dimensionality), and fully connected layers that output the final predictions. CNNs have become the standard architecture for tasks like image classification, object detection, and visual recognition, as they are highly effective at learning spatial relationships in data.

- RNNs: Recurrent Neural Networks are Deep Learning architectures designed for sequential data, such as time series, speech, or text. Unlike traditional feedforward networks, RNNs have connections that loop back, allowing them to retain information from previous inputs. This makes RNNs powerful for tasks where context or memory of prior inputs is essential, such as language modeling or speech recognition. However, standard RNNs suffer from vanishing gradients when dealing with long sequences. To address this, variants like LSTMs (Long Short-Term Memory networks) and GRUs (Gated Recurrent Units) were developed, allowing RNNs to capture long-term dependencies in sequences more effectively. These models are widely used in machine translation, time series forecasting, and natural language processing tasks. [83]

3.18 One-to-one classification

In ML, One-to-One classification is a strategy used to extened binary classification problems to handled classic problems. In this strategy, a separate classifier is trained for every possible pair of classes in the dataset. For a problem with n classes there will be $\frac{n(n-1)}{2}$ binary classifiers. For this research project, for 9 classes there will be 36 binary classifiers. In a One-vs-One approach, each classifier cast a vote for one of the two classes it was trained on. The class that receives the majority of votes across all pairwise classifiers is assigned as the final prediction. [84]

3.19 Confusion matrix

In [85], the author introduced concepts related to statistical classification, which laid the groundwork for what we now call the confusion matrix. The matrix gained more formal usage and adoption in machine learning, statistics, and signal processing communities in the mid-20th century. Over time, it became a standard tool for evaluating classification models, particularly as computational methods and classification techniques evolved.

The confusion matrix (Fig. 3.10) summarizes the performance of a classification algorithm by displaying counts of:

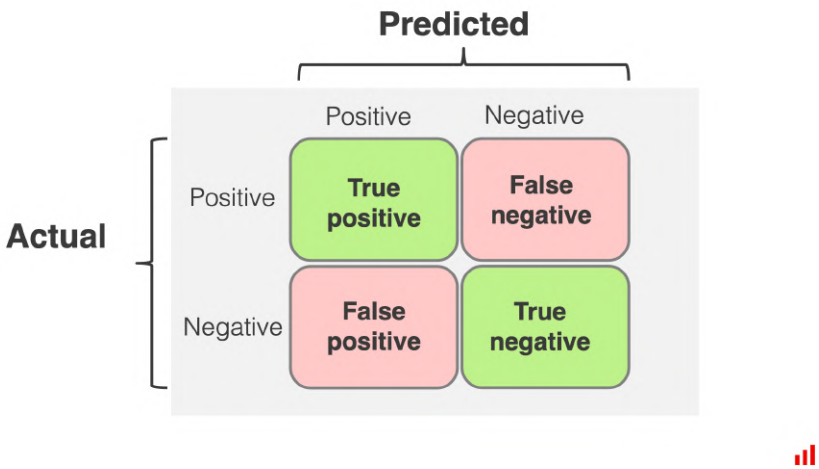


Figure 3.10: How to interpret a confusion matrix. Source: [86]

- True Positives (TP): Positive instances correctly predicted.
- True negatives (TN): Negative instances correctly predicted.
- False Positives (FP): Positive instances incorrectly predicted.
- False Negatives (FN): Negative instances incorrectly predicted.

The confusion matrix remains useful in the evaluation of deep learning-based classifiers, especially when discussing performance metrics for binary or multiclass classification problems.

3.20 Performance metrics

In this section are described metrics that can be calculated from the confusion matrix to evaluate the performance of the model based on the predictions that has made. The metrics that are going to be used in this work are listed below.

3.20.1 Accuracy

Proportion of correct predictions over the total predictions. Eq (46)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{46}$$

3.20.2 Precision

Proportion of true positives about the total positive predictions. Eq (47)

$$Precision = \frac{TP}{TP + TN} \tag{47}$$

3.20.3 Recall

Proportion of true positives about the total real positive cases. Eq (48)

$$Recall = \frac{TP}{TP + FN} \tag{48}$$

3.20.4 F1-score

Proportion of true positives about the total real positive cases. Eq (49)

$$F1 - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{49}$$

3.21 Genetic Algorithms

A Genetic Algorithm (GA) belong to the branch of Evolutionary Algorithms, which are inspired by biological evolution. This algorithm is a meta-heuristic, i.e. an algorithm use to solve complex problems, where it effectively identifies a solution that is close to the best one in a very vast search space. The description of how a genetic algorithm works is as follows:

3.21.1 Population initialization

It starts by creating a population, an initial set of possible solutions (individuals) to the problem you are trying to solve. Each individual represents a possible solution, they are made up of a set of genes or parameters.

3.21.2 Fitness evaluation

You evaluate the fitness of the population, i.e. the fitness of each individual. This function indicates how well each individual solves the problem. Individuals with greater fitness values are better solutions.

3.21.3 Selection

Selects individuals from the current population to become parents of the next generation. Selection is biased towards individuals with the highest fitness, increasing the chances that 'good' solutions will be passed on to the next generation. As selection methods can be found:

- Roulette Selection: Individuals are chosen with a probability proportional to their fitness. You can imagine it like spinning a roulette wheel, where the size of each individual's slice is proportional to its fitness value. [87]

- Tournament Selection: Selects a few individuals at random and chooses the best of them to be a parent. The size of the tournament controls the level of selection pressure. [88]

- Rank Selection: Assigns ranks to individuals based on their fitness and selects individuals based on their rank. This can reduce the influence of outliers on fitness.

- Monogamous Random Selection: A settler is chosen at random, marked and the same procedure is carried out with another settler, so that when marking them they are not chosen again.

- Polygamous Random Selection: Similar to the previous one, they are chosen at random but they are not marked, so that they can be chosen again

3.21.4 Crossover

Combines genetic material from two parent individuals to create one or more offspring. Crossing may involve exchanging genes between parents at specific crossing points such as 1-way or 2-way or even n-way crosses, or homogeneous crosses.

3.21.5 Mutation

Introduces random changes (mutations) into the genetic material of offspring to promote diversity in the population. Mutations help the algorithm to explore new regions of the search space. 2 mutation techniques were implemented in this paper:

- Bit-swap mutation: Technique used in genetic algorithms and evolution-inspired optimization algorithms to explore new solutions. In simple terms, it involves swapping positions between bits in the binary representation of a genome or a candidate solution.

    a. Selecting two random positions within the chromosome

    b. Taking the genes from these positions

    c. Swapping the genes from these positions [89]

3.21.6 Elitist criterion

Replace the population of the current and previous generations with the best settlers generated. This step ensures that the population evolves in successive generations since only the fittest individuals survive.

3.21.7 Termination

The selection, crossing, mutation and replacement steps are repeated for a certain number of generations or until a termination criterion is reached. In this case a hybrid criterion was used:

- number of generations
- termination by δ
- termination by ϵ

Selection is crucial to maintain diversity in the population while favoring promising solutions. The balance between exploration (diversity) and exploitation (fitness improvement) is a key consideration when designing effective genetic algorithms. [90][91]

4 Hypothesis

The application of signal processing and machine learning techniques to EEG signals enables the recognition of Imagined Speech from simple words with accuracy exceeding the current state-of-the-art threshold of 87%, furthermore, by incorporating genetic algorithms for feature selection, the classification process can be further optimized, leading to additional improvements in the recognition of simple Spanish words in Imagined Speech.

5 Objectives

5.1 General objective

Develop a model that decodes imagined speech by classifying EEG signals using AI techniques to assist patients with speech disorders.

5.2 Specific objectives

- Obtain and label EEG signals to generate a database.
- The database contain signals from spanish words.
- Implement an AI model to evaluate and classify signals.

6 Methodology

This chapter describes in detail the steps and environment to record the data, the preprocessing data techniques, optimization with a Genetic Algorithm and the classifier models tested. In addition, in the present thesis research project, the data was collected by own hand due to the lack of Spanish imagined speech EEG datasets. Following sections present the steps to build the dataset.

6.1 Develop Graphic User Interface to record

It is necessary a program that label the signals at the moment they are being recorded. The Emotiv Software Pro allow us to add markers while recording, nevertheless, it was required to add particular markers (a label indicating the corresponding word thought) in the exact moment. Therefore, a **Graphic User Interface (GUI)** was coded using the Streamlit package, which is an open-source Python framework for data scientists.

In the builded GUI (Fig. 6.1), one is abble to specify the metadata of the individual, the fields are:

- Recording time (set to always 120s seconds)
- Change word every n(s) (set to always 10 seconds)
- Individual age (20 to 30 years)
- Individual gender (Male, Female)
- individual ID (s00x_n, where x is the subject number and n is the number of the subject recording)

After the GUI menu was builded, the eight words mentioned before ('Si', 'No', 'Baño', 'Hambre', 'Sed', 'Ayuda', 'Dolor', 'Gracias') where converted to mp3 files using Google Text-to-Speech (gTTS), which is a Python library and CLI tool to interface with Google Translate's text-to-speech API. This mp3 files are going to be reproduced according to the recording methodology depicted in Fig. 6.5.

Once the user click on the "Start recording" button, the GUI will behave exactly as the Fig. 6.5. It will count the seconds to play the 'beep' sound and the mp3 words files trough the speakers until complete the specified 120 seconds. Also, in the GUI will be shown the current word in which the individual has to think as shown by Fig. 6.2.

Once the 120 seconds have elapsed, it will be play a final mp3 file saying "Gracias. Hemos terminado." (Thank you, we have finished).

Also, it will be generated a JSON file containing:

- Individual metadata: to associate the file with its corresponding EEG record.

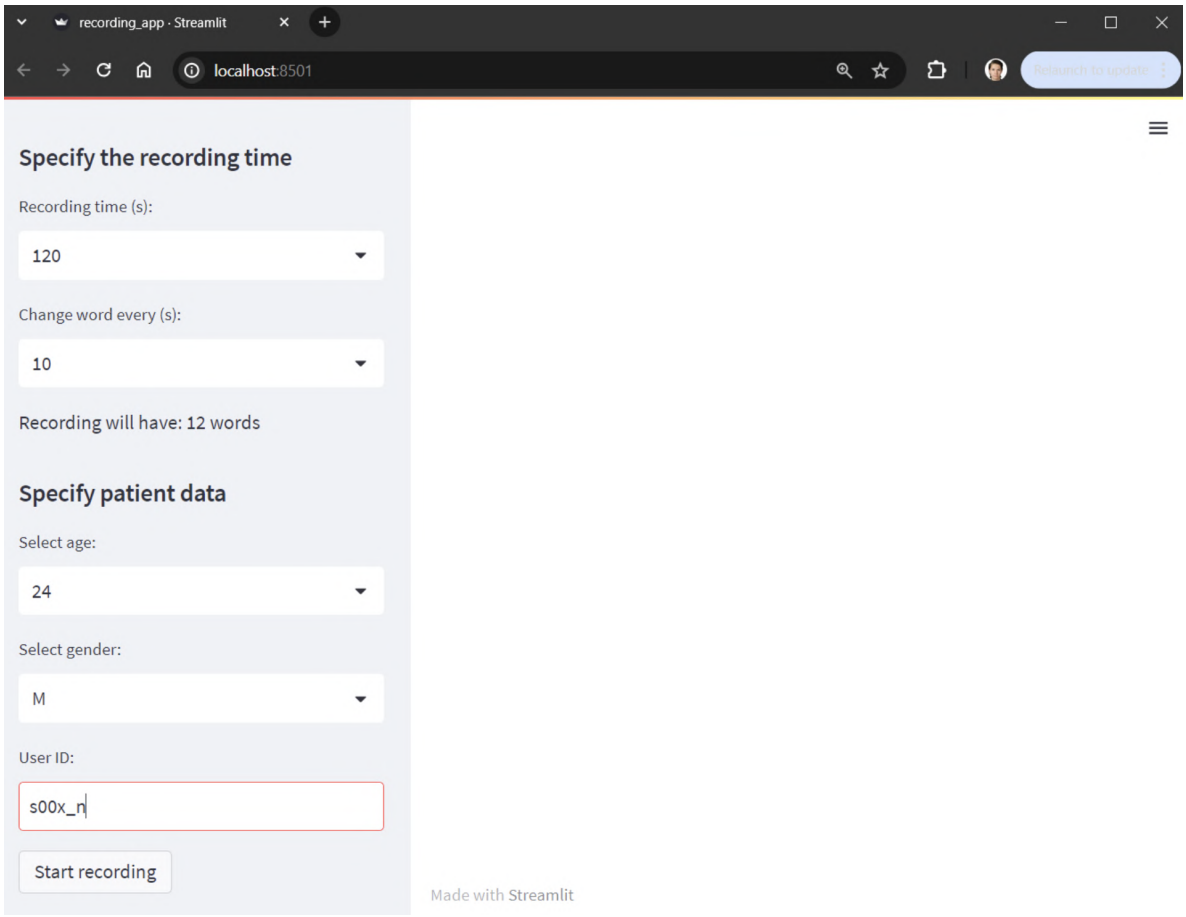


Figure 6.1: Builded GUI menu. Own authorship.

- labels: the words that appear during the recording, necessary due to the words appear randomly in each recording.

An example of a JSON file is in Fig. 6.3.

Once the user clicks the "Start Recording" button, they must start recording at the same time in the Emotiv Software, so that the audios, recordings and labels are correctly aligned temporally.

In the Fig. 6.4 is described the steps that were followed in this research. Each step will be described in a specific section on this chapter.

1. The first topic to be addressed is the **development of the Interface** to record data.

2. Once the interface was developed, the next step was **record data**. This was done running at the same time the interface en the Emotiv software.

3. After generating the records, the recording files were cleaned, labelled and re-shaped into chunks.

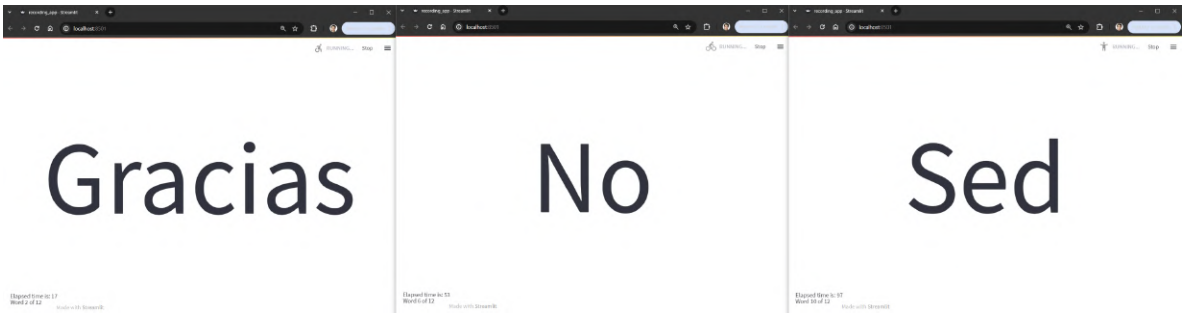


Figure 6.2: Example of current words presented in the screen by the GUI. Own authorship.

Figure 6.3: JSON file generated by the GUI. Own authorship.

4. Once chunks were generated, they go trough several steps in order to extract features and characterize the signals generating feature maps that describes the behaviour, nature and patterns in the signals.

5. After characterizing signals, the features maps enter to the classifier models. The models have the goal of find patterns in order to correlate with the label. In other words, finding patterns in the signals to be able to classify them with the imagined word. In addition, a Genetic Algorithm is used to search for the combination of electrodes and features that allows the model to predict with greater accuracy.

6. Finally, after finish all this process, the output will be the best classifier model and the best combination of electrodes and features that allows to identify the imagined word from the EEG signals.

6.2 Record data

6.2.1 Inclusion criteria

Individuals must have the following characteristics:

- be right-handed
- not taking any medicine that affects the nervous system

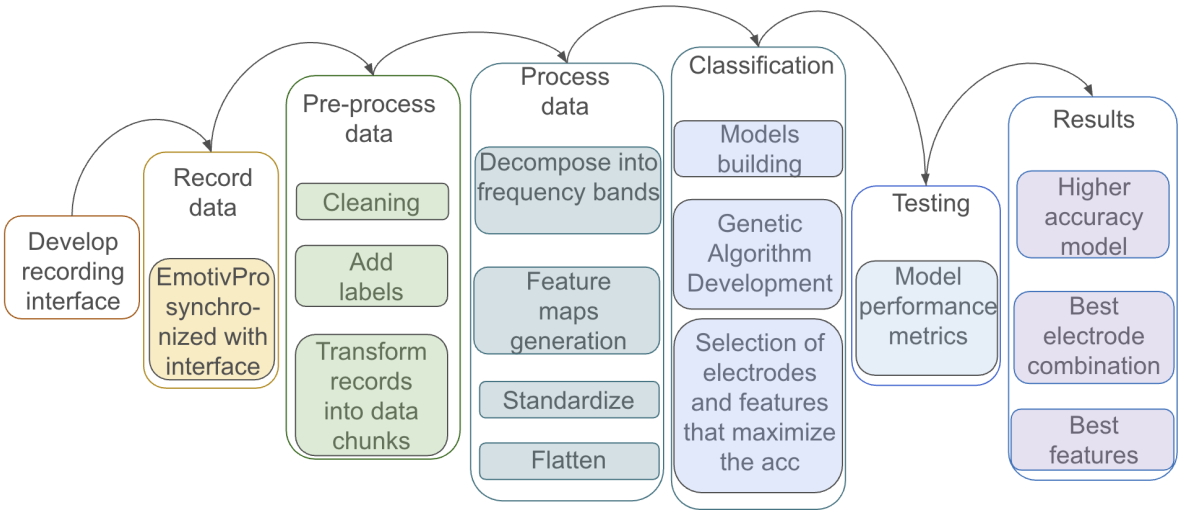


Figure 6.4: Methodology diagram.

- age between 20-30 years
- indistinct gender
- not to be a smoker
- not to drink alcohol
- not to use drugs
- not having epileptic seizures

6.2.2 Informed consent

Informed consent was used in this research to ensure ethical standards were met and to protect the rights, privacy, and autonomy of all participants. By obtaining informed consent, participants were fully informed about the purpose, procedures, and benefits of the study, enabling them to make a voluntary and informed decision to participate. This process also established transparency and trust between the researcher and participants, allowing individuals to understand their role in the study and affirming their freedom to withdraw at any stage without consequence. Informed consent is essential in upholding ethical research practices, aligning with institutional guidelines and broader ethical standards in research. In the annexes section can be found the informed consent employed in this research.

6.2.3 Individual indications

Each individual was programmed to attend to the recording session. The individual should attend this indispensable indications for the recording day. The indications are

listed below:

- Having had breakfast.
- Clean and dry hair, washed only with shampoo.
- Do not use creams or rinses, do not apply gel or spray.
- Do not stay up late.
- Do not stop taking medication and bring doses of these.
- The patient should not have a cough, flu or fever.

6.2.4 Words to classify

[23], in which twenty commonly used words for patients were selected from a communication board widely used in hospitals around the world for patients with paralysis/aphasia.
Inspired by [23], the present thesis research project employs eight complex words plus the mute word, which is an state where the individual should not think in any word. The eight words are the following:

- 'Si'
- 'No'
- 'Baño'
- 'Hambre'
- 'Sed'
- 'Ayuda'
- 'Dolor'
- 'Gracias'

6.2.5 Recording methodology

Following a methodology similar to the consulted in the state-of-the-art. In this project, 5 recordings will be made for each subject. Each recording will last 120 seconds (2 minutes).

Each recording will have the following structure:

a. A *beep* sound trough the speakers for 1 second.

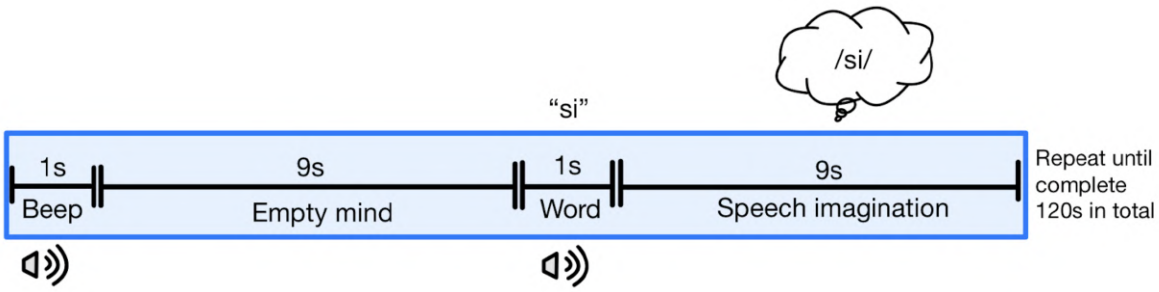


Figure 6.5: Recording methodology. Own authorship.

b. The following 9 seconds, the individual have to try to keep his mind clear and not think in any word.

c. A random word (from eight words menu mention before) sound trough the speaker for 1 second.

d. The individual have think in say the corresponding word without actually saying it.

e. This entire process repeats until complete the 120 seconds.

6.2.6 Place the headset

After the GUI have been build, it was time to record the signals. For that, the first step is to properly set the headset. The steps were the following:

1. Prepare the sensors:
    a. Hydrate the Sensors: The EPOC+ uses saline-soaked felt pads. Start by placing the felt pads into each sensor, and apply a few drops of saline solution to each one. The pads should be moist but not dripping.
    b. Attach Sensors: Insert each hydrated sensor into the headset.
2. Adjust the headset fit:
    a. Position the Headset: Place the headset so the sensors are properly positioned on the scalp and behind the ears. The headset should sit comfortably, resting on the head without slipping. (Fig. 6.6)
3. Start the software connection:
    a. Power on and Pair with Device: Once adjusted, the headset was turned on and paired with the computer.
4. Position the sensors:
    a. Align the Sensors: The sensors should be in contact with the scalp, as good contact improves signal quality. The headset has specific placements to match with the head's 10-20 EEG system points. (Fig. 6.7)
    b. Ensure Contact: All sensors were checked to make firm contact with the scalp. Sometimes it might need to move aside some hair or adjust individual sensors to ensure good contact. (Fig. 6.8A)
5. Check for proper fit:
    a. The EPOC+ device has an indicator in the app showing if each sensor has a good connection. It was adjusted as needed for optimal contact, using the device software to confirm connection quality. (Example of bad contact quality in Fig. 6.8B). (Example of good contact quality in Fig. 6.8C).

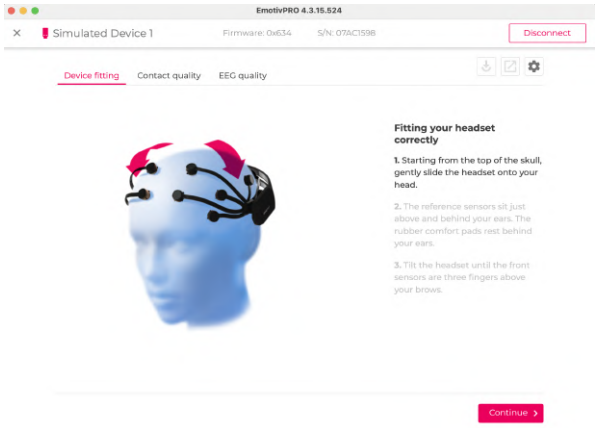


Figure 6.6: Fitting the Epoc+ headset properly.

Figure 6.7: Example of Epoc+ headset placement in an individual. Own authorship.

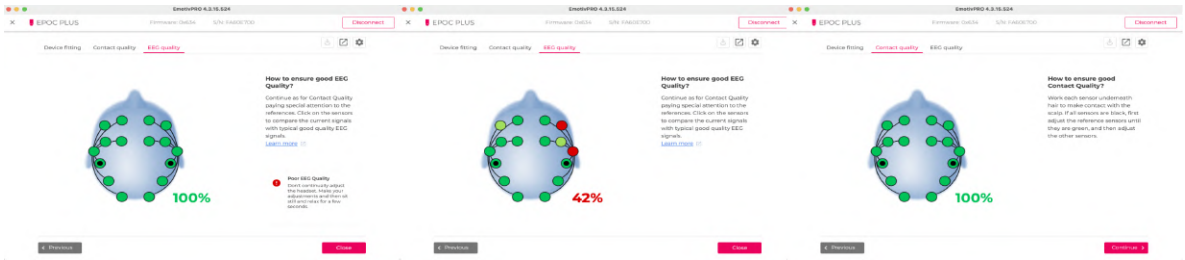


Figure 6.8: A-Ensure good contact quality. B-Example of bad contact in Epoc+ headset. C-Example of bad contact in Epoc+ headset. Own authorship.

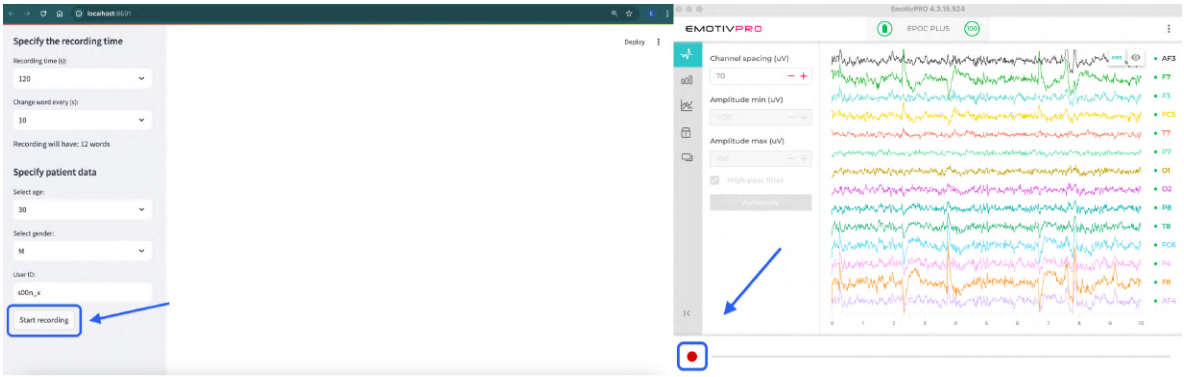


Figure 6.9: A-Record button in Emotiv software. B-Record button in GUI.

6.2.7 Record signals

Once the connection was verified, then proceed to fill the metadata blank in GUI (Fig. 6.1) and Emotiv software (Fig. 6.10). After that, proceed to record the signals. To do this, first click on the "Start_recording" button (Fig. 6.9A), once the first "beep" sounds through the speaker, click on the record button in the Emotiv software (Fig. 6.9B).

Once the signals have been recorded, we can open the record files and visualize the plots. (Fig. 6.11)

6.2.8 Export and store data

After signals have been record, it is necessary to export the data. The first step is to select the files and then proceed to click on the "Export" button as shown in fig. 6.12A. After that, a specifications window will open, unmark boxes if there are so, mark the "CSV" box and then click on the "Export" button, as shown in fig. 6.12B.

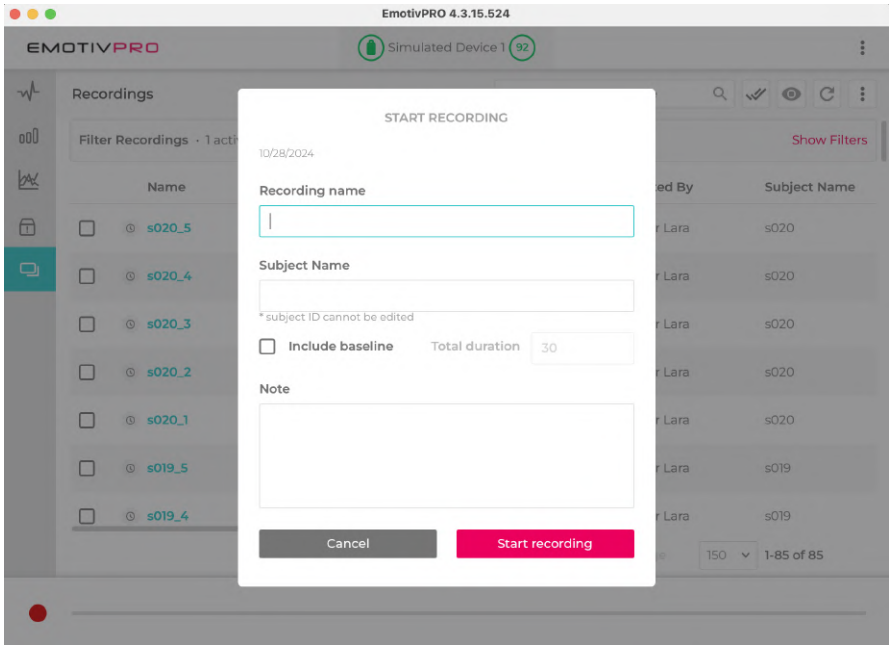


Figure 6.10: Specify metadata before start recording. Own authorship.

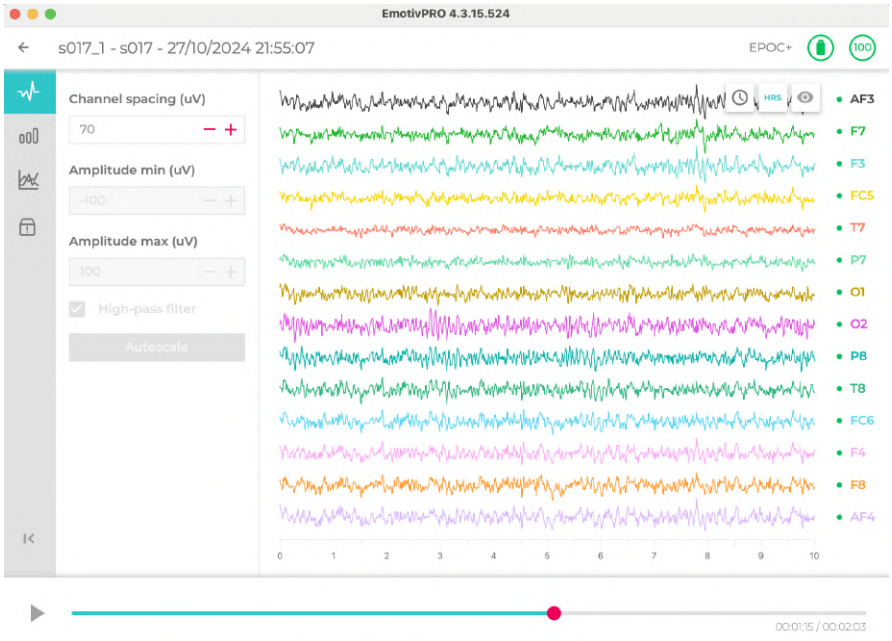


Figure 6.11: Example of a recorded signal. Own authorship.

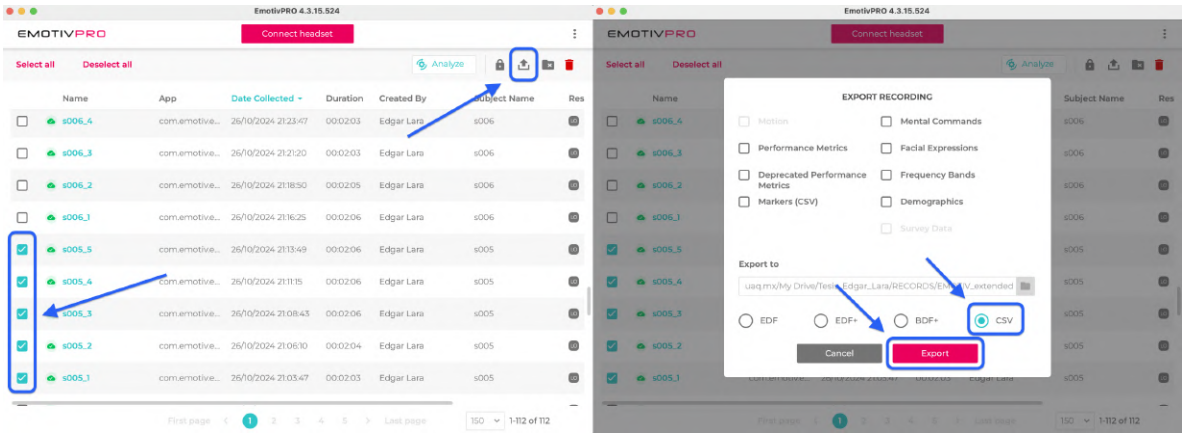


Figure 6.12: A- Select files to export. B- Export files specifications.

6.3 Pre-process data

6.3.1 Data cleaning

The exported CSV contains several columns with metadata related to the device. nevertheless, the ones containing the signals are the ones named and renamed as follows:

- EEG.AF3: AF3
- EEG.F7: F7
- EEG.F3: F3
- EEG.FC5: FC5
- EEG.T7: T7
- EEG.P7: P7
- EEG.O1: O1
- EEG.O2: 02
- EEG.P8: P8
- EEG.T8: T8
- EEG.FC6: FC6
- EEG.F4: F4
- EEG.F8: F8
- EEG.AF4: AF4

Additionally, in this step each JSON file was opened to get the labels and inject them into its corresponding CSV files. The labels were transform using the following categorical encoder:

- 0: '...'
- 8: 'Si'
- 6: 'No'
- 2: 'Baño'
- 5: 'Hambre'
- 7: 'Sed'
- 1: 'Ayuda'
- 3: 'Dolor'
- 4: 'Gracias'

The dataset was made public in the Mendeley data repository and can be found as EEGIS - Electroencephalogram Imagined Speech dataset in [92].

6.3.2 Datasets

After this cleaning process, it was created a second dataset by simply doing a query in the words 'Sí' and 'No'. This was done in order to verify if there are differences in the classification of Simple Spanish words vs Complex Spanish words. The answer can be found in the Results section. The following steps in the methodology were applied to both datasets.

Once the data was cleaned, the next step was to create separate windows of data. Each window contains 10 seconds signals related to only one word thought instead of the entire 120 seconds recording. From each window were removed the first two seconds keeping only the remaining eight seconds, the 1st second correspond to the "beep" sound or the random word played by the speaker. The 2nd second was removed to ensure that the individual have began to think in that word. As shown in Fig. 6.13

Therefore the data shape is (N, 14, 1024) Where:

- N: number of instances
- 14: number of electrodes
- 2014: frames of the window

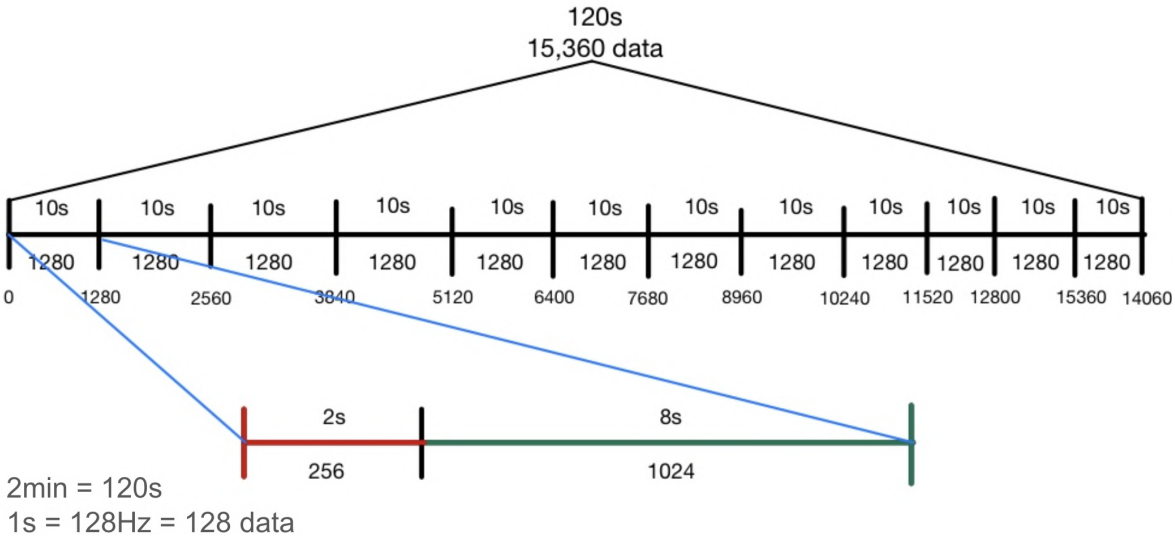


Figure 6.13: Generate data windows. Own authorship.

6.3.3 Make data chunks

After split the data into windows, it is necessary to generate smaller windows, which will be called chunks. The data chunks consist of data associated to only 1 second of recordings, i.e 128 frames. For that reason the windows where split into chunks of 128 frames with an overlap of 16 frames, resulting in 9 chunks for every window.

An illustrative way to explain this can be observed in the following pairs of indexes, that explain how the chunks were generated from each window.

$$[0, 128]$$

$$[112, 240]$$

$$[224, 352]$$

$$[336, 464]$$

$$[448, 576]$$

$$[560, 688]$$

$$[672, 800]$$

$$[784, 912]$$

$$[896, 1024]$$

As a result, the data shape now is (N, 14, 128)

- N: number of chunks
- 14: number of electrodes
- 128: frames of the chunk

Examples of chunks extracted from a window can be seen in Fig. 6.14.

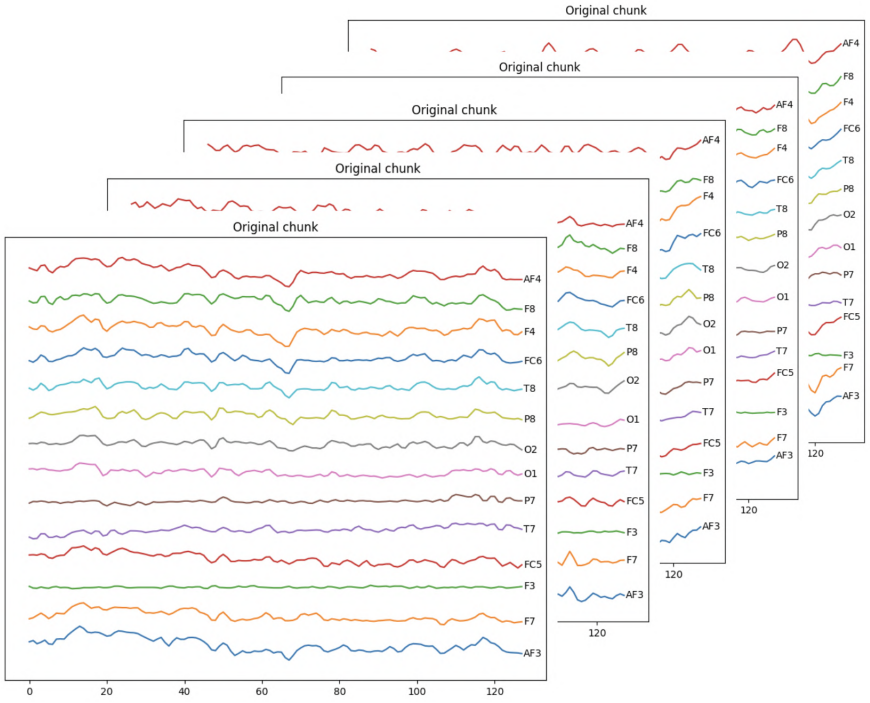


Figure 6.14: Example of chunks. Own authorship.

6.4 Process data

6.4.1 Class balance

As observed in Fig. 6.15a, due to the occurrence of the state empty mind during the recording stage, the class '...' appears much more than the rest of the words. Therefore, a class balance is necessary in order to perform in further steps a proper **classification**. It was taken the mean occurrence of the rest of the classes, and then the '...' was reduced, randomly, to the mean value. As observed in Fig. 6.15b, the class balance was perform.

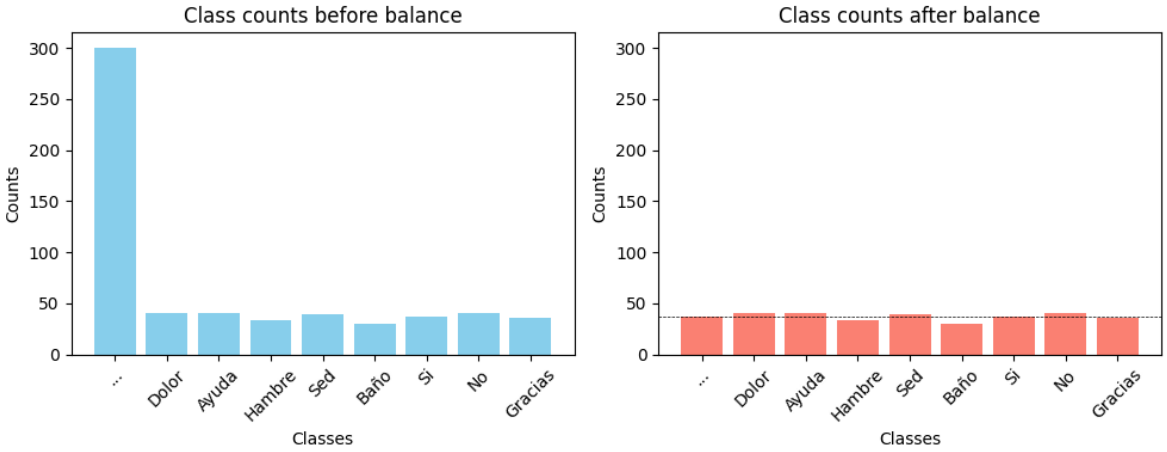


Figure 6.15: a. Unbalanced data. b. Balanced data

6.4.2 Split signals into frequency bands

Imagined speech involves silently "saying" words or sentences within the mind without vocalizing them. This process activates specific brain regions and generates neural oscillations that can be detected through EEG. By filtering EEG data into delta, theta, alpha, beta, and gamma bands, it can be effectively study the neural correlates of imagined speech, as each frequency band reveals different aspects of cognitive and neural processing related to this internal task.

- Delta band
  - Role in Imagined Speech: Delta oscillations are associated with foundational processes like memory encoding and retrieval, both of which play a part in generating internal speech, especially if one is recalling phrases or constructing sentences.
  - Potential for Detecting Imagined Speech: While delta frequencies are generally less prominent during active cognition, slow-wave oscillations may still be useful for studying the rhythmic pattern of silent articulation, especially in longer imagined sentences.
- Theta band
  - Role in Imagined Speech: Theta waves are strongly linked with working memory and the cognitive control needed to structure thoughts into coherent inner speech. They are also associated with language processing and can be more prominent in the frontal and temporal lobes during inner speech tasks.
  - Potential for Detecting Imagined Speech: Theta band power is often elevated during verbal working memory tasks, such as keeping a sentence in mind while imagining speaking it. This makes theta oscillations a potential marker for identifying imagined speech, especially in experiments where subjects are mentally rehearsing sentences.
- Alpha band
  - Role in Imagined Speech: Alpha oscillations, especially in the posterior regions, are linked to attentional control and mental inhibition. During imagined speech, alpha power may increase to help inhibit external stimuli, allowing for focused inner dialogue.
  - Potential for Detecting Imagined Speech: Alpha band activity has been observed to decrease in brain areas associated with language processing (like the left temporal lobe) during verbal thinking tasks. Monitoring changes in alpha power, especially reductions in task-related regions, could serve as a marker of imagined speech activity.
- Beta band

- Role in Imagined Speech: Beta waves are associated with motor control, attention, and active cognitive engagement. During imagined speech, beta activity often reflects the activation of motor-related areas (like the sensorimotor cortex) even though no actual movement occurs.
- Potential for Detecting Imagined Speech: Imagining speech often activates the motor cortex in a similar pattern as real speech, producing beta oscillations in this area. Monitoring beta activity, especially in the motor and premotor areas, can help detect when a person is engaged in internal speech. This makes beta a key frequency band in brain-computer interface (BCI) applications aiming to interpret imagined speech.

- Gamma band
  - Role in Imagined Speech: Gamma oscillations are involved in high-level cognitive processes like binding information across brain regions, including integrating semantic, syntactic, and phonological elements of language during speech.
  - Potential for Detecting Imagined Speech: Gamma activity is often elevated during complex cognitive tasks that involve language comprehension or production. During imagined speech, gamma synchronization in areas like the left temporal and frontal lobes may reflect the neural integration necessary for constructing sentences. Monitoring gamma could help capture the specific content or semantic processing involved in inner speech, potentially distinguishing between types of imagined phrases.

Therefore, it is necessary to split the signals into its respective frequency bands.

That was achieved using a 4th order Butterworth band-pass filter. These filter characteristics were chosen based on:

- Butterworth Filter Characteristics
  - Smooth Frequency Response: A Butterworth filter provides a maximally flat frequency response in the passband, meaning it has no ripples. This smooth response is particularly advantageous for EEG signal processing, where abrupt changes could distort the delicate signal components across bands.
  - Avoidance of Distortion: Butterworth filters avoid distortion within the passband frequencies, which is crucial when analysing EEG because any non-linearities could introduce unwanted artifacts or alter the physiological signals. This choice minimizes interference with the true nature of the EEG signal.

- Choice of Filter Order (4th Order)

- Balance Between Sharpness and Computational Efficiency: Higher-order filters provide sharper cutoffs, but they also introduce increased computational complexity and phase distortion. A 4th-order filter is often chosen as a balance, providing adequate roll-off at the band edges without excessive computational cost or phase distortion.
- Sufficient Selectivity for EEG Bands: EEG bands are defined over relatively broad frequency ranges (e.g., Delta at 0.5-4 Hz, Beta at 13-30 Hz), and a 4th-order filter offers enough roll-off to effectively isolate these bands without significant overlap.
- Reduced Phase Distortion: Lower-order filters, such as a 4th-order, generally induce less phase distortion compared to higher orders, which is beneficial since EEG analysis can be sensitive to phase relationships across frequencies.

The cut-off frequencies were:

- Gamma: 30 Hz - 100 Hz
- Beta: 13 Hz - 30 Hz
- Alpha: 8 Hz - 13 Hz
- Theta: 4 Hz - 8 Hz
- Delta: 0.5 Hz - 4 Hz

The five frequency filters were applied to each chunk for every electrode.
After this step, the data shape was (N, 14, 128, 6) where:

- N: number of chunks
- 14: number of electrodes
- 128: frames of the chunk
- 6: channels (original and five filters)

In the Fig. 6.16 can be observed in a illustrative way how every electrode from each chunk was filtered into the five frequency bands mentioned before.

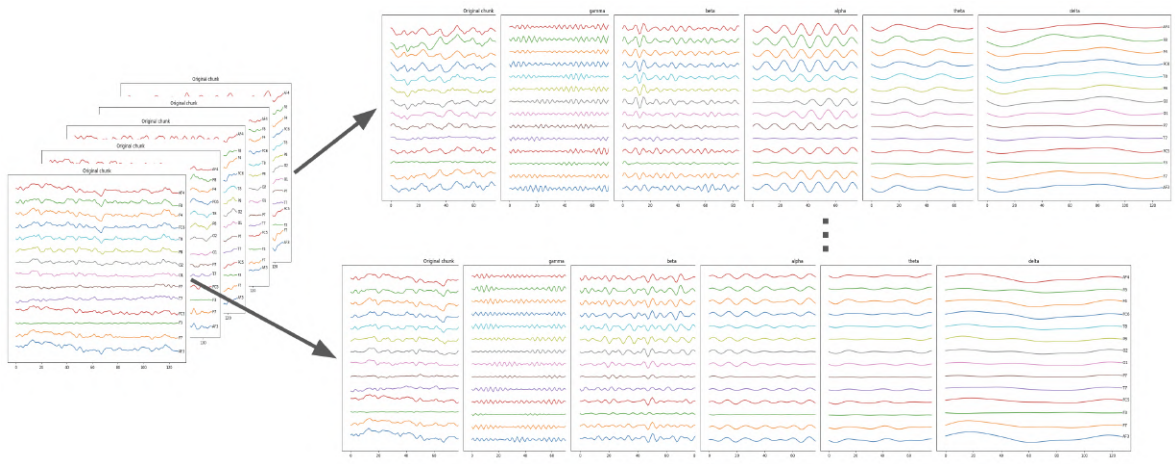


Figure 6.16: Filtered chunks

6.4.3 Feature Extraction and proposed features

As mention before, the GA will be feed with the features maps which contain info related to the characteristics and thee electrodes, those feature maps are generated with the feature extraction technique, therefore, before explaining the GA implementation is necessary to explain about the feature extraction technique and the corresponding generation of feature maps.

After applying filters to the raw EEG signals, it is necessary to extract features. This necessity rely on that most classification models depend on well-defined, relevant features rather than raw signal data to perform optimally. By extracting features, we are providing the model with concise, high-quality inputs that improve its ability to discriminate between different EEG patterns or states.

In addition, raw EEG signals are highly complex and noisy, with large amounts of data points across time. Directly analysing the raw signal can be inefficient and often does not yield meaningful insights. Feature extraction condenses this raw data into a more informative set of parameters that capture the key characteristics of the signals.

The proposed features have a diversity types such as:

- Statistical Features: Statistical characteristics (e.g., mean, variance, skewness, kurtosis) offer a summary of the signal's overall distribution and variability. These are essential for capturing basic but fundamental aspects of EEG activity, such as the general energy level in each band, the presence of bursts or spikes, and the symmetry of the distribution. Statistical features are especially useful because they provide a global overview of signal properties.

- Time-Series Metrics: Time-domain features, such as peak-to-peak amplitude, zero-crossing rate, and signal entropy, capture temporal dynamics and fluctuations within each EEG band. These metrics reveal details about how the EEG signal changes over time, offering insights into the presence of specific waveforms

(e.g., alpha waves, beta bursts) or rhythm patterns. These are crucial for tasks like mental state classification or identifying transient events within the EEG.

- Fractal and Nonlinear Features: Fractal dimensions and other nonlinear metrics capture the complexity and irregularity of the EEG signal. EEG data often exhibit chaotic and fractal properties, especially in certain cognitive or neurological states. Fractal features are powerful in distinguishing states of consciousness or identifying neurological anomalies because they quantify the signal's self-similarity and structural complexity, which are hard to capture with traditional linear metrics.

By using a combination of statistical, time-series, and fractal features, it can be covered a broader spectrum of EEG signal characteristics. This multi-faceted approach captures both the general distribution (through statistical features), the dynamic fluctuations (through time-series metrics), and the underlying complexity of the signal (through fractal dimensions). This comprehensive representation yield to more robust and reliable insights into the EEG data.

Different types of features can complement each other, offering more discriminative information. For instance, while statistical features might capture the overall power level in a certain band, fractal features can provide information about cognitive states associated with more complex brain activity. This diversity in feature types helps in differentiating between various cognitive or neurological states with higher accuracy.

The features employed are the following:

- relative band power [80]
- band amplitude [81]
- mean [61]
- standard deviation [62]
- coefficient variation [63]
- singular valued decomposition entropy [77]
- Petrosian fractal dimension [78]
- Katz fractal dimension [79]
- Higuchi fractal dimension [71]
- median [62]
- mode [62]
- max [64]

- min [64]
- spectral edge frequency 25 [75]
- spectral edge frequency 50 [75]
- spectral edge frequency 75 [75]
- kurtosis [62]
- skewness [62]
- detrended fluctuation analysis [66]
- determinism [74]
- trapping time [74]
- diagonal line entropy [74]
- average diagonal line length [74]
- recurrence rate [74]
- activity Hjorth param [67]
- mobility Hjorth param [67]
- complexity Hjorth param [67]
- Hurst exponent [76]
- total power spectral density [72]
- centroid power spectral density [73]
- permutation entropy [68]
- approximate entropy [69]
- spectral entropy [70]
- first quartile [65]
- third quartile [65]
- inter quartile range [65]

In the Chapter 3.15 were explained one by one the 36 proposed features.

6.4.4 Generation of feature maps

Once defined the 36 features, they were extracted from the signals

a. The first 34 features were extracted from the raw EEG chunk.

b. All the 36 were extracted from each of 5 the band filtered chunks.

c. Therefore, for a single chunk the total features extracted were 214.

d. This was done for each of the 14 electrodes.

As a result, the feature maps have a shape of:

$$(14, 214)$$

Obtaining a feature tensor, Fig. 6.17, where each instance is a feature map that correspond to each chunk. Once the feature maps were generated, they can enter to

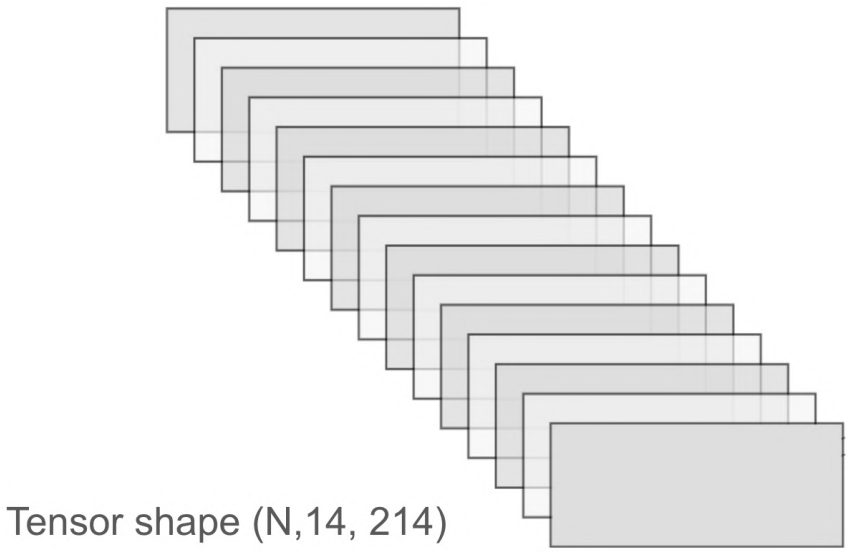


Figure 6.17: Feature maps diagram.

the ML models in order to classify the EEG signals for Imagined Speech. Nevertheless, this feature maps can be reduced i.e. reduce the electrodes and the amount of features extracted. This can be done using a Genetic Algorithm (GA) were using the optimal

combination of electrodes.

The GA aims to select electrodes and combine features to optimise the accuracy of a Machine Learning model. The vast number of possible combinations necessitates the use of a genetic algorithm, a meta-heuristic designed to optimise feature selection effectively. The genetic algorithm will create a population of genomes for reproduction and evolution across generations. The genomes will consist of binary vectors of length 224:

- the first 14 elements of the vectors show which electrodes will be selected.
- the remaining 214 items show which features need to be selected in order to enhance.

Ten different traditional ML classification models will be used to improve the electrodes and features. The model's accuracy, as defined in Eq. (46), will be used as the objective function, turning this into a maximisation problem.

This process is described in Fig. 6.18.

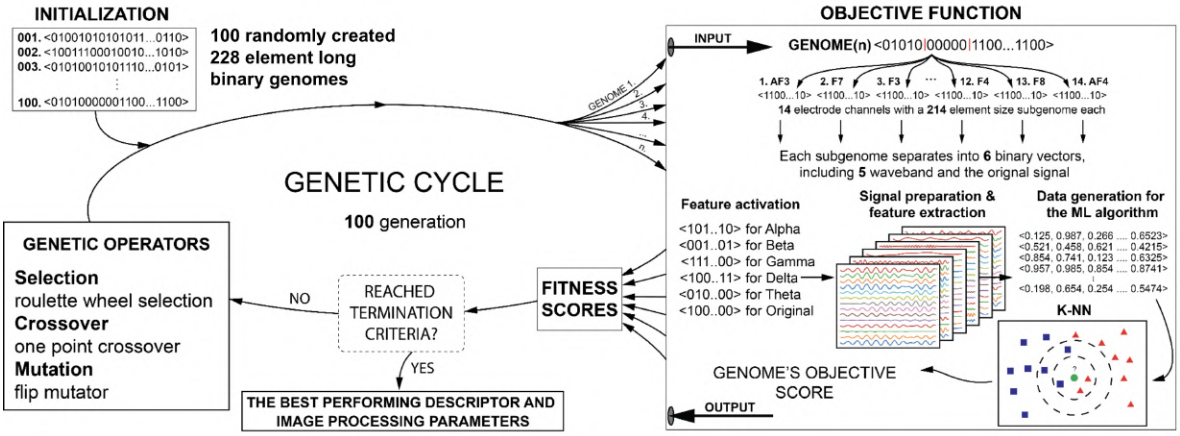


Figure 6.18: Genetic Algorithm cycle diagram. Source: [1]

The parameters for the GA are the followings:

- Type of gen: dichotomic 0,1
- Number individuals: 100
- Number of genes: 214
- Type of selection: rank
- Elitist criterion: Genetic competition
- Type of crossover: single point
- Fitness function: Accuracy's model
- Mutation: bit flip
- Termination criteria: 100 generations

6.4.5 Prepare input data for classifier models

Ten different Machine Learning models (listed in the next section) were employed in order to classify. For that reason, the following steps were done in order to have a proper input data:

6.4.6 Normalize data

Data processing, normalisation, or standardisation is essential prior to inputting into the ML model to improve performance. The models to be used use data standardization (Eq. (50)).

$$z = \frac{x_i - \mu}{\sigma} \tag{50}$$

Let x_i represent the instance, i.e. data point, to be standardised, μ denote the mean of all instances in the column, and σ signify the data column standard deviation. All data attributes must be standardised on a common scale to preserve the integrity of value ranges and maintain information fidelity.

6.4.7 Flatten data

The shape of the feature matrices is (14,214), as previously stated. The GA identifies the optimal features and electrodes that provide the most pertinent information. Consequently, the matrices are structured as (N, M), where N represents the number of electrodes selected by the genome and M denotes the number of features specified by the genome. Classical ML models require a data vector as entry, consequently, data points require be reshaped by flattening the feature matrices from

$$(N, M)$$

to

$$(1, NxM)$$

feature vectors. This is illustrated more clearly in the diagram of Fig. 6.19.

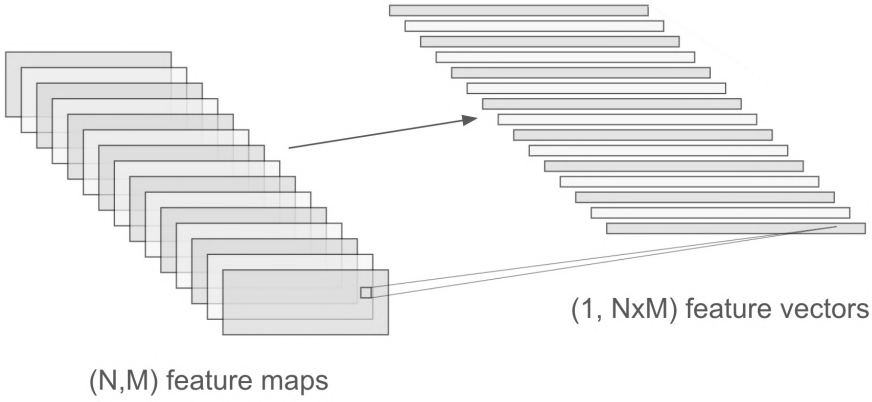


Figure 6.19: Flatten feature maps.

6.5 Classifier models

Ten different Machine Learning models were employed in order to classify the EEG signals. TThe models are listed below, and were described in depth in Chapter 3.

a. Support Vector Machine

b. Multi Layer Perceptron

c. Linear Discriminant Analysis

d. Naïve Bayes

e. AdaBoost

f. Gradient Boosting Machine

g. Decision Tree

h. Random Forest

i. Logistic Regression

j. kNN

6.5.1 GA and entries explanation

The GA, a metaheuristic search algorithm which is inspired by Darwinian evolution theory , will optimise electrode and feature selection to maximise classification accuracy as previously mentioned. Nevertheless, the search space in this instance is too extensive for an exhaustive computational search, as it consists of 228 factorial possible combinations of electrodes and features. Consequently, the GA will effectively identify an approximate optimal solution within this extensive search space.

The term "individuals" must be clarified when discussing the GA, as it does not specifically refer to the ten subjects in our dataset. Rather, each "individual" in this context is a potential solution. Specifically, a binary vector of length 228 signifies whether or not specific features and electrodes are selected (1) or not (0).

The genetic algorithm works as follows:

1. Initialization:

    a. Binary vector population of 100 individuals is randomly generated. 100 is a standard population size in GAs.

2. Selection and Reproduction:

a. Through pairing, called selection in the context of GA, and recombination, called crossover, these 100 individuals are crossed to create offspring. Population increase twice, i.e. to 200 temporarily.

3. Evaluation and Survival of the Fittest:

   a. A different subset of possible solution of features and electrodes is specified by each binary vector, this will be used in classification steps.

   b. Based on the classifier accuracy the individuals are ordered.

   c. The top 100 individuals are preserved for the subsequent generation, while thee bottom 100 are eliminated.

4. Iterative Optimization:

   a. This strategy iteratively enhances the selection of electrodes and attributes across multiple generations until the algorithm reaches an optimal subset.

5. Final Solution Selection:

   a. Once the algorithm reaches convergence, the final solution is the individual (binary vector) that achieved the highest classification in the last generation. This vector indicates which electrodes and features provide the model with the most useful information, with values of 0 and 1.

By using a GA, it is possible to avoid computationally expensive calculations, the GA allows us to efficiently investigate the large search space to find the most suitable features and electrodes.

6.5.2 Hyperparameter optimization

In classification models, hyperparameter tuning plays a crucial role in classifier performance in terms of accuracy. In this work, default parameters were used for all classifiers to ensure a fair comparison, as this is a common approach in evaluating ML classifier models. However, it is important to mention that hyperparameter optimization could improve accuracy performance.

Despite hyperparameter tuning was not applied uniformly across all classifiers due to computational limitations, a heuristic search for the optimal number of k-neighbors was performed for the kNN classifier model, which likely contributed to achieve superior accuracy compared to the other classifiers.

Although hyperparameters were not fine-tuned, feature set optimization helped improve model performance.

A key factor in this approach was computational efficiency. Given the required preprocessing steps, as well as the genetic algorithm used (100 individuals, 100 generations, 10 models, and a KNN search on 7 k-values), this resulted in 700,000 different model training runs, making hyperparameter tuning across all classifiers unfeasible with the described setup due to its high computational cost and time consumption.

This work recognizes that hyperparameter tuning could further refine model performance; therefore, we plan to explore this in future work to provide a more comprehensive comparison.

6.6 Additional approaches

The methodology described in this section was not the only one proposed or approved. Recurrent neural networks (RNN), Long Short Term Memory (LSTM) networks, Gated Recurrent Unit (GRU), and a CNN+LSTM combination were also implemented.

Unfortunately, this approach was not successful for multi-class classification because the results showed low accuracy, there was excessive bias, and the models were unable to discern between classes. This was because this approach was unable to decipher patterns and generalize the tiny but existing features in the signals.

Not only were variants of recurrent networks worked with, but simple deep networks were also implemented, as well as variations of CNNs where the final layers were connected to Deep Neural Networks (DNN) or LSTM networks. However, as already mentioned, this approach was not able to find patterns in the signals and perform correct classification.

Another approach was to implement transfer learning using networks such as ResNet and VGG, in which the shape of the input data were matrices (electrodes x frames x 1), these matrices were subjected to transformations and reshapes in order to adjust them to the input shape of the aforementioned networks, once this was done, the header of said networks was frozen and the output layers were modified with the aim of predicting the nine classes present in the dataset.

7 Results

7.1 Binary classification for /Sí/ and /No/ words

In this section, you can find graphs of the GA performance for each of the ten classifiers. As can be seen, each of these graphs starts with relatively low accuracy values since the GA begins with randomly initialized possible solutions. However, with each generation, the population average and the best individual increase, as expected. In this section, we can also find the genome of the best individual, that is, the specific binary vector that indicates which electrodes and features are activated to achieve the best accuracy for each of the models.

In addition, this section includes the performance of the genome of the best individual found by the genetic algorithm will be evaluated for each of the 10 classifier models. This will be evaluated using the confusion matrix, from which a classification report will be generated that includes the metrics accuracy, precision, recall, and f1-score. The decoding of the genome, that is, the binary vector, is also shown, displaying the electrodes as well as the features found by the GA that maximized the accuracy for that particular model.

7.1.1 kNN-GA performance

Fig. 7.1 corresponds to the GA combined with the kNN classifier model.

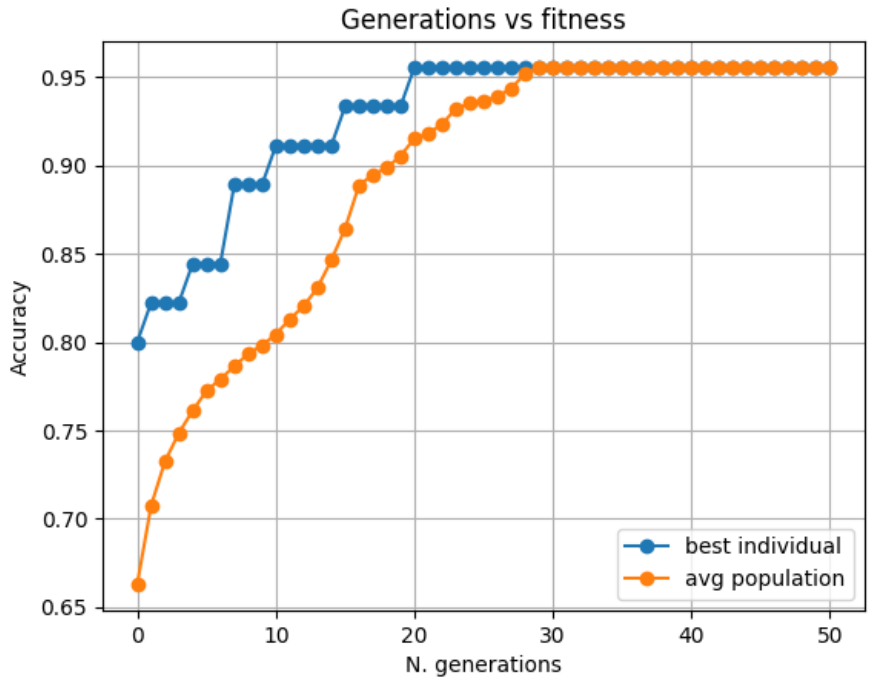


Figure 7.1: kNN-GA training performance.

best genome =
(0,0,0,0,1,0,1,1,1,0,1,0,1,0,0,0,0,1,1,0,1,1,0,0,0,0,0,0,1,1,0,0,0,1,0,1,0,0,0,1,
0,1,0,0,1,1,1,1,1,0,1,0,1,1,0,1,1,1,1,1,1,1,0,0,1,0,0,1,1,0,1,0,0,1,0,1,0,0,1,0,1,1,
0,0,0,1,0,0,0,1,1,0,0,1,0,0,0,1,1,1,1,1,1,0,1,0,0,0,0,1,1,0,0,0,1,1,1,0,1,1,1,0,1,1,
0,1,1,0,0,0,0,0,0,1,1,0,0,1,0,1,0,1,0,0,0,1,1,0,0,0,0,1,0,1,0,0,0,1,1,0,0,1,1,0,0,
0,0,0,1,0,1,0,1,1,0,1,1,1,1,0,0,1,0,1,1,1,1,1,1,1,0,1,1,1,0,1,1,1,1,1,0,0,0,0,0,1,
1,0,0,1,1,0,0,1,0,1,1,1,0,0,1,1,0,0,1,0,0,1,0,1)

7.1.2 Decode genome for kNN

8 selected electrodes: AF3, T7, O1, O2, P8, T8, F4, F8

15 features for raw data: std, coeff var, median, mode, third quartile, skewness, activity hjorth param, complexity hjorth param, permutation entropy, approximate entropy, total power spectral density, centroid power spectral density, spectral edge frequency 50, spectral edge frequency 75, katz fractal dimension

23 features for delta band: mean, std, median, mode, max, first quartile, inter quartile range, skewness, detrended fluctuation analysis, activity hjorth param,

mobility hjorth param, permutation entropy, approximate entropy, spectral entropy, higuchi fractal dimmension, centroid power spectral density, trapping time, diagonal line entropy, average diagonal line length, spectral edge frequency 50, singular valued decomposition entropy, petrosian fractal dimension, band amplitude

16 features for theta band: coeff var, max, min, inter quartile range, skewness, detrended fluctuation analysis, activity hjorth param, mobility hjorth param, complexity hjorth param, permutation entropy, spectral entropy, total power spectral density, spectral edge frequency 75, petrosian fractal dimension, katz fractal dimension, band amplitude

14 features for alpha band: std, mode, max, inter quartile range, skewness, detrended fluctuation analysis, complexity hjorth param, permutation entropy, approximate entropy, higuchi fractal dimmension, diagonal line entropy, spectral edge frequency 75, katz fractal dimension, band amplitude

20 features for beta band: mean, median, first quartile, third quartile, inter quartile range, skewness, activity hjorth param, mobility hjorth param, complexity hjorth param, permutation entropy, spectral entropy, centroid power spectral density, diagonal line entropy, average diagonal line length, compute recurrence rate, spectral edge frequency 50, spectral edge frequency 75, hurst exponent, relative band power, band amplitude

15 features for gamma band: mean, max, min, detrended fluctuation analysis, permutation entropy, total power spectral density, centroid power spectral density, determinism, trapping time, diagonal line entropy, compute recurrence rate, spectral edge frequency 25, spectral edge frequency 50, singular valued decomposition entropy, band amplitude

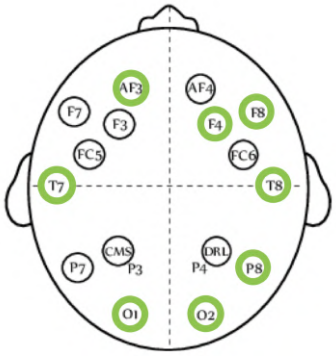


Figure 7.2: Optimal electrodes found for kNN.

7.1.3 kNN evaluation

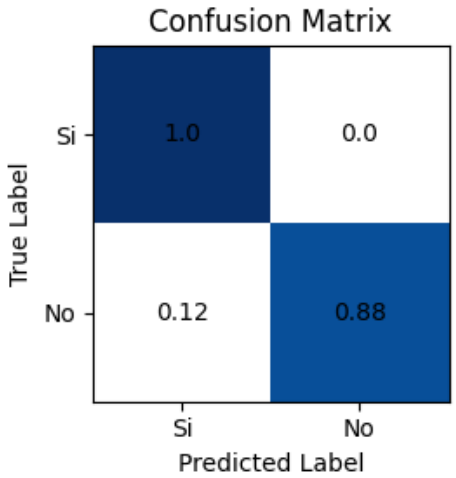


Figure 7.3: kNN confusion matrix.

	precision	recall	f1-score	support
0	0.93	1.00	0.97	28
1	1.00	0.88	0.94	17
accuracy			0.96	45
macro avg	0.97	0.94	0.95	45
weighted avg	0.96	0.96	0.95	45

Figure 7.4: kNN performance metrics.

7.1.4 LogReg-GA performance

Fig. 7.5 corresponds to the GA combined with the Logistic Regression classifier model. best genome =

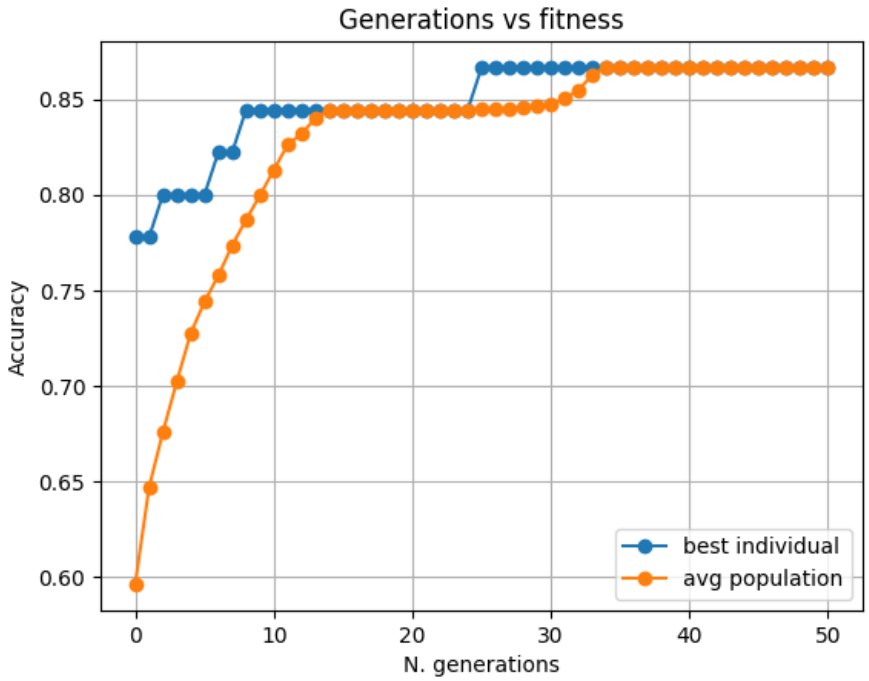


Figure 7.5: LogReg-GA training performance.

(0,0,0,1,0,0,1,1,1,0,0,0,1,0,1,1,1,1,1,0,1,0,1,1,0,1,1,0,1,0,1,1,1,0,0,1,0,1,1,0,
0,0,0,0,0,1,1,1,0,0,1,0,0,0,0,1,0,0,0,0,0,0,1,1,1,1,1,0,0,0,1,0,1,1,0,1,1,1,0,1,
1,0,1,1,0,1,0,1,0,1,1,0,0,0,1,0,0,1,0,0,1,0,1,0,0,0,0,1,1,0,1,1,1,0,1,1,1,1,1,1,0,
1,0,1,0,0,1,1,1,0,1,1,0,1,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,1,1,1,1,1,0,1,0,0,0,0,0,0,
0,0,1,1,1,0,0,0,1,0,1,0,1,1,1,0,1,0,1,0,0,1,0,1,0,1,1,0,1,0,1,0,0,0,1,1,1,0,1,1,1,
1,1,1,0,0,0,0,1,1,1,0,1,0,0,1,0,0,0,0,1,1,1,0,1,1)

7.1.5 Decode genome for LogReg

5 selected electrodes: FC5, O1, O2, P8, F8

20 features for raw data: mean, std, coeff var, median, mode, min, third quartile, inter quartile range, skewness, detrended fluctuation analysis, mobility hjorth param, permutation entropy, approximate entropy, spectral entropy, centroid power spectral density, trapping time, diagonal line entropy, singular valued decomposition entropy, petrosian fractal dimension, katz fractal dimension

17 features for delta band: coeff var, first quartile, mobility hjorth param, complexity hjorth param, permutation entropy, approximate entropy, spectral entropy, determinism, diagonal line entropy, average diagonal line length, spectral edge frequency 25, spectral edge frequency 50, spectral edge frequency 75, singular valued decomposition entropy, petrosian fractal dimension, relative band power, band amplitude

19 features for theta band: std, median, max, min, kurtosis, activity hjorth param, permutation entropy, spectral entropy, trapping time, diagonal line entropy, compute recurrence rate, spectral edge frequency 25, spectral edge frequency 50, hurst exponent, singular valued decomposition entropy, petrosian fractal dimension, katz fractal dimension, relative band power, band amplitude

16 features for alpha band: std, median, min, first quartile, third quartile, kurtosis, skewness, activity hjorth param, average diagonal line length, compute recurrence rate, spectral edge frequency 75, hurst exponent, singular valued decomposition entropy, petrosian fractal dimension, katz fractal dimension, band amplitude

15 features for beta band: third quartile, inter quartile range, kurtosis, mobility hjorth param, permutation entropy, spectral entropy, higuchi fractal dimmension, total power spectral density, determinism, diagonal line entropy, spectral edge frequency 25, spectral edge frequency 75, singular valued decomposition entropy, petrosian fractal dimension, relative band power

20 features for gamma band: mean, mode, max, min, third quartile, inter quartile range, kurtosis, skewness, detrended fluctuation analysis, activity hjorth param, spectral entropy, higuchi fractal dimmension, total power spectral density, determinism, average diagonal line length, hurst exponent, singular valued decomposition entropy, petrosian fractal dimension, relative band power, band amplitude

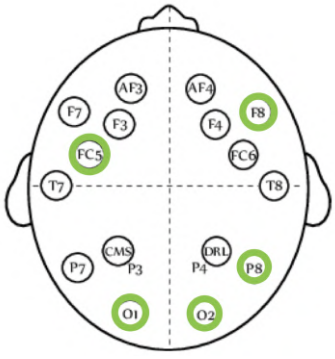


Figure 7.6: Optimal electrodes found for LogReg.

7.1.6 LogReg evaluation

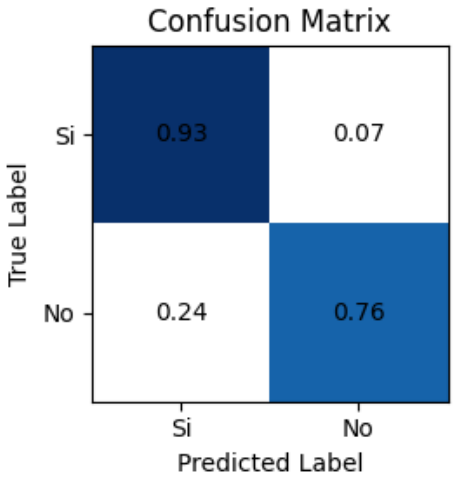


Figure 7.7: LogReg confusion matrix.

	precision	recall	f1-score	support
0	0.87	0.93	0.90	28
1	0.87	0.76	0.81	17
accuracy			0.87	45
macro avg	0.87	0.85	0.85	45
weighted avg	0.87	0.87	0.86	45

Figure 7.8: LogReg performance metrics.

7.1.7 RF-GA performance

Fig. 7.9 corresponds to the GA combined with the Random Forest classifier model. best genome =

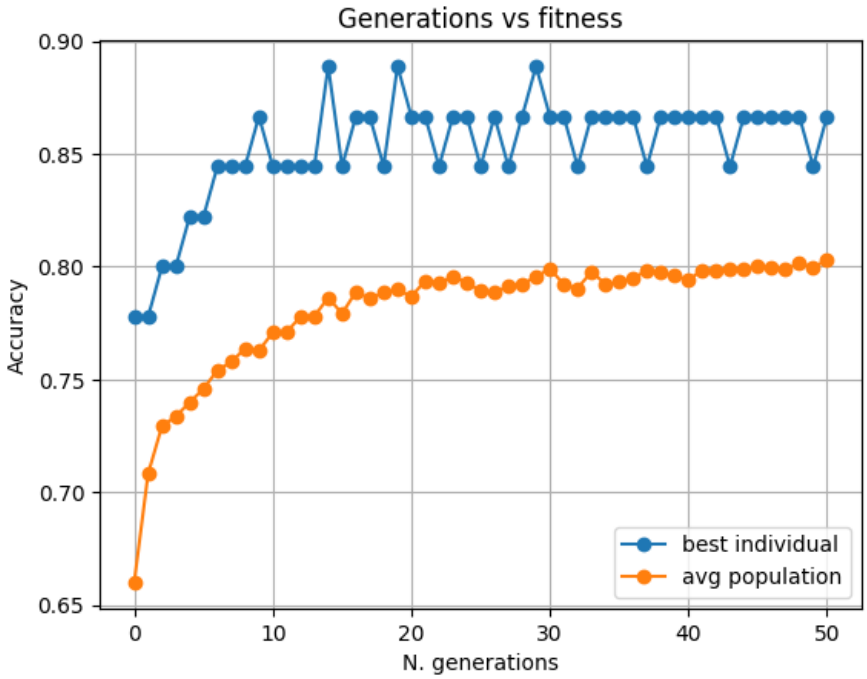


Figure 7.9: RF-GA training performance.

(0,0,1,0,1,1,1,0,1,0,0,0,1,1,0,1,1,0,1,0,0,0,1,0,1,1,0,0,1,0,0,0,0,1,1,0,0,0,0,0,0,
0,0,1,1,0,1,1,0,0,1,0,1,1,1,0,0,0,1,1,0,0,1,1,0,0,1,0,1,0,0,1,1,0,1,1,1,1,1,1,1,0,1,
0,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,1,0,0,1,0,1,0,1,1,0,0,1,0,1,1,0,1,0,0,1,0,1,1,0,0,1,
0,1,0,1,1,0,1,1,1,1,1,1,1,1,0,1,0,0,0,0,1,0,0,1,1,1,0,1,0,0,1,0,1,0,1,0,0,0,1,0,0,1,
0,0,1,1,0,1,0,0,1,1,0,1,0,0,1,1,1,1,1,0,0,1,0,1,1,0,0,0,0,0,0,1,0,0,1,1,1,0,1,1,0,1,0,
1,0,1,0,1,1,1,0,0,1,0,0,1,1,0,1,1,1,0)

7.1.8 Decode genome for RF

7 selected electrodes: F3, T7, P7, O1, P8, F8, AF4

13 features for raw data: std, coeff var, mode, third quartile, kurtosis, skewness, mobility hjorth param, higuchi fractal dimmension, total power spectral density, spectral edge frequency 75, hurst exponent, petrosian fractal dimension, katz fractal dimension

20 features for delta band: coeff var, mode, max, min, kurtosis, skewness, mobility hjorth param, complexity hjorth param, spectral entropy, total power spectral

density, trapping time, diagonal line entropy, compute recurrence rate, spectral edge frequency 25, spectral edge frequency 50, spectral edge frequency 75, hurst exponent, singular valued decomposition entropy, petrosian fractal dimension, relative band power

17 features for theta band: coeff var, max, min, first quartile, third quartile, skewness, activity hjorth param, complexity hjorth param, spectral entropy, total power spectral density, determinism, trapping time, compute recurrence rate, spectral edge frequency 50, spectral edge frequency 75, singular valued decomposition entropy, relative band power

21 features for alpha band: mean, std, mode, min, third quartile, inter quartile range, skewness, detrended fluctuation analysis, activity hjorth param, mobility hjorth param, complexity hjorth param, permutation entropy, approximate entropy, spectral entropy, total power spectral density, average diagonal line length, spectral edge frequency 50, spectral edge frequency 75, hurst exponent, petrosian fractal dimension, band amplitude

17 features for beta band: std, median, first quartile, kurtosis, activity hjorth param, mobility hjorth param, permutation entropy, higuchi fractal dimmension, total power spectral density, determinism, average diagonal line length, compute recurrence rate, spectral edge frequency 25, spectral edge frequency 50, singular valued decomposition entropy, katz fractal dimension, relative band power

18 features for gamma band: max, third quartile, inter quartile range, kurtosis, detrended fluctuation analysis, activity hjorth param, complexity hjorth param, approximate entropy, higuchi fractal dimmension, centroid power spectral density, determinism, trapping time, compute recurrence rate, spectral edge frequency 75, hurst exponent, petrosian fractal dimension, katz fractal dimension, relative band power

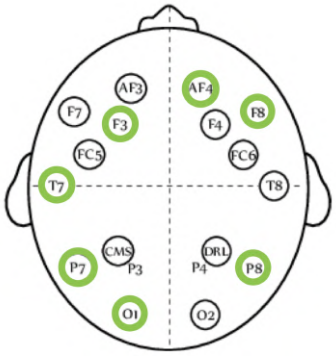


Figure 7.10: Optimal electrodes found for RF.

7.1.9 RF evaluation

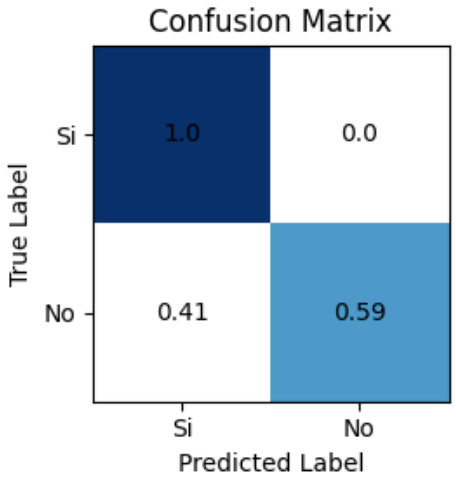


Figure 7.11: RF confusion matrix.

	precision	recall	f1-score	support
0	0.80	1.00	0.89	28
1	1.00	0.59	0.74	17
accuracy			0.84	45
macro avg	0.90	0.79	0.81	45
weighted avg	0.88	0.84	0.83	45

Figure 7.12: RF performance metrics.

7.1.10 DT-GA performance

Fig. 7.13 corresponds to the GA combined with the Decision Tree classifier model. best

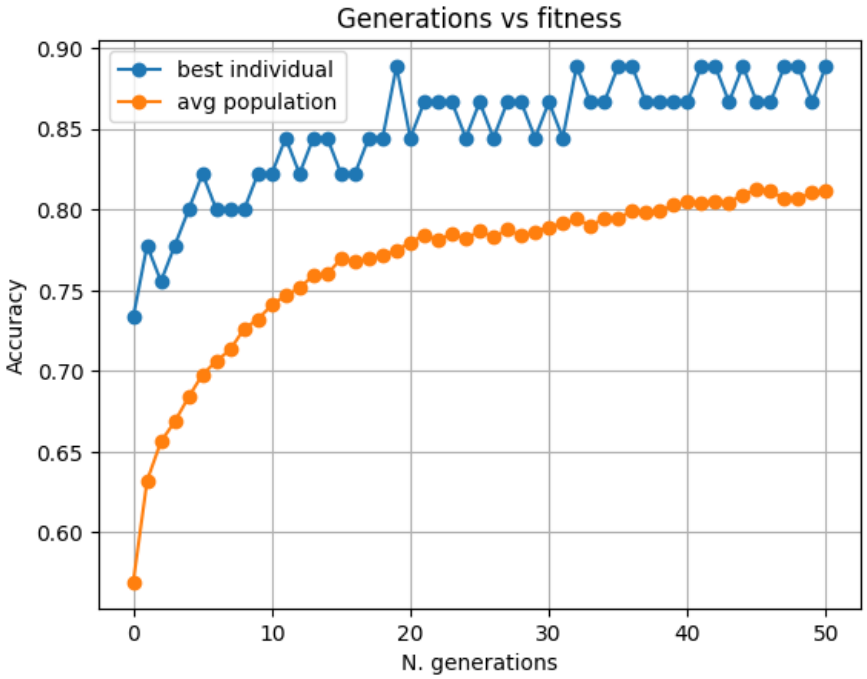


Figure 7.13: DT-GA training performance.

genome =
(1,0,0,0,0,0,1,1,1,1,1,1,1,0,1,1,0,1,0,1,1,1,1,1,1,0,0,0,1,0,0,0,1,1,0,0,0,1,0,0,0,
0,1,0,1,1,1,1,0,0,0,1,0,0,1,1,1,1,1,1,1,1,0,1,1,0,0,0,1,0,0,0,0,1,1,0,0,0,1,1,0,1,0,0,
0,0,0,1,1,1,0,0,0,0,0,1,1,1,1,1,0,0,0,1,0,0,0,0,1,1,1,1,0,0,0,0,0,0,1,0,0,0,1,0,1,1,
1,0,1,0,1,0,1,0,0,0,0,1,0,0,0,1,0,1,0,1,1,0,1,1,0,0,0,1,1,1,0,0,1,0,1,0,1,1,0,1,0,1,
0,1,0,1,0,0,1,1,0,1,1,0,0,0,1,1,1,1,0,1,0,1,0,1,1,0,1,0,0,1,0,0,1,1,1,0,0,0,0,0,1,0,
0,1,1,0,1,1,1,1,1,0,1,0,1,0,1,0,0,0,0)

7.1.11 Decode genome for DT

8 selected electrodes: AF3, O1, O2, P8, T8, FC6, F4, F8

18 features for raw data: mean, std, median, max, min, first quartile, third quartile, inter quartile range, kurtosis, mobility hjorth param, spectral entropy, higuchi fractal dimmension, trapping time, spectral edge frequency 50, hurst exponent, singular valued decomposition entropy, petrosian fractal dimension, katz fractal dimension

16 features for delta band: median, min, first quartile, third quartile, inter quartile range, kurtosis, skewness, detrended fluctuation analysis, mobility hjorth param,

complexity hjorth param, higuchi fractal dimmension, diagonal line entropy, average diagonal line length, spectral edge frequency 75, hurst exponent, petrosian fractal dimension

14 features for theta band: coeff var, median, mode, kurtosis, skewness, detrended fluctuation analysis, activity hjorth param, mobility hjorth param, spectral entropy, trapping time, diagonal line entropy, average diagonal line length, compute recurrence rate, katz fractal dimension

17 features for alpha band: std, median, mode, max, first quartile, inter quartile range, skewness, permutation entropy, total power spectral density, determinism, diagonal line entropy, average diagonal line length, spectral edge frequency 25, spectral edge frequency 50, petrosian fractal dimension, katz fractal dimension, relative band power

20 features for beta band: std, median, max, min, third quartile, kurtosis, detrended fluctuation analysis, mobility hjorth param, approximate entropy, spectral entropy, total power spectral density, centroid power spectral density, average diagonal line length, compute recurrence rate, spectral edge frequency 25, spectral edge frequency 50, hurst exponent, petrosian fractal dimension, relative band power, band amplitude

16 features for gamma band: std, mode, first quartile, third quartile, inter quartile range, complexity hjorth param, spectral entropy, higuchi fractal dimmension, centroid power spectral density, determinism, trapping time, diagonal line entropy, average diagonal line length, spectral edge frequency 25, spectral edge frequency 75, singular valued decomposition entropy

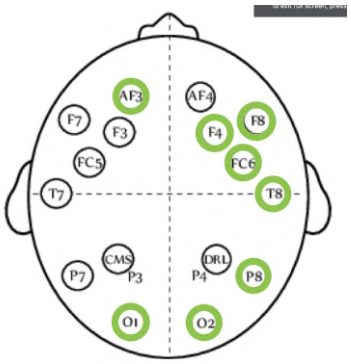


Figure 7.14: Optimal electrodes found for DT.

7.1.12 DT evaluation

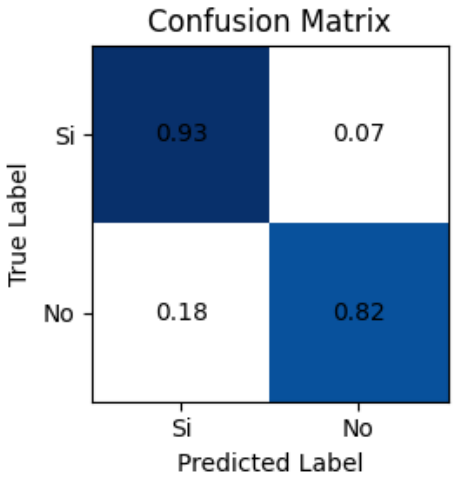


Figure 7.15: DT confusion matrix.

	precision	recall	f1-score	support
0	0.90	0.93	0.91	28
1	0.88	0.82	0.85	17
accuracy			0.89	45
macro avg	0.89	0.88	0.88	45
weighted avg	0.89	0.89	0.89	45

Figure 7.16: DT performance metrics.

7.1.13 GBM-GA performance

Fig. 7.17 corresponds to the GA combined with the Graadient Boosting Machine classifier model. best genome =

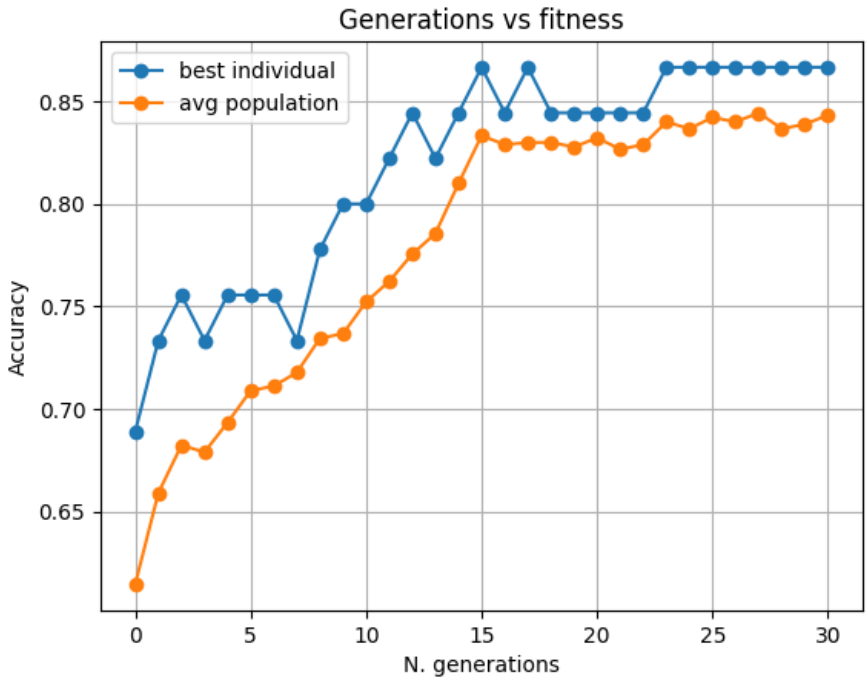


Figure 7.17: GBM-GA training performance.

(0,0,1,0,0,0,1,1,1,0,0,1,1,0,0,1,1,0,1,0,1,1,0,1,1,0,0,1,0,1,1,1,0,1,0,0,1,0,1,1,
0,1,1,1,0,0,1,1,0,0,1,1,0,0,0,1,0,1,0,1,0,0,1,0,0,0,0,1,1,0,1,1,0,1,0,0,1,0,1,0,
1,1,0,1,1,0,1,0,0,1,0,1,0,1,0,1,1,1,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,1,0,
0,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,1,1,0,0,0,1,1,1,0,1,1,1,1,1,0,1,1,1,1,0,0,1,1,0,0,
1,0,1,0,1,0,0,1,1,1,0,0,1,0,1,0,1,1,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,0,0,0,0,1,
1,0,1,1,1,1,0,0,1,0,1,0,0,1,1,0,0,1,0,0,1,0,0,0)

7.1.14 Decode genome for GBM

6 selected electrodes: F3, O1, O2, P8, F4, F8

20 features for raw data: std, coeff var, mode, min, first quartile, inter quartile range, kurtosis, activity hjorth param, complexity hjorth param, permutation entropy, approximate entropy, higuchi fractal dimmension, determinism, diagonal line entropy, average diagonal line length, spectral edge frequency 25, spectral edge frequency 50, spectral edge frequency 75, petrosian fractal dimension, katz fractal dimension

15 features for delta band: coeff var, median, first quartile, inter quartile range, skewness, mobility hjorth param, higuchi fractal dimmension, total power spectral density, determinism, trapping time, average diagonal line length, spectral edge frequency 50, hurst exponent, katz fractal dimension, relative band power

14 features for theta band: mean, std, median, min, third quartile, kurtosis, detrended fluctuation analysis, activity hjorth param, mobility hjorth param, permutation entropy, approximate entropy, spectral entropy, singular valued decomposition entropy, petrosian fractal dimension

16 features for alpha band: mean, third quartile, kurtosis, spectral entropy, higuchi fractal dimmension, trapping time, diagonal line entropy, average diagonal line length, spectral edge frequency 25, spectral edge frequency 50, spectral edge frequency 75, hurst exponent, singular valued decomposition entropy, katz fractal dimension, relative band power, band amplitude

19 features for beta band: mean, median, mode, first quartile, inter quartile range, skewness, mobility hjorth param, complexity hjorth param, permutation entropy, higuchi fractal dimmension, centroid power spectral density, trapping time, diagonal line entropy, compute recurrence rate, spectral edge frequency 25, hurst exponent, singular valued decomposition entropy, relative band power, band amplitude

15 features for gamma band: coeff var, median, min, skewness, detrended fluctuation analysis, mobility hjorth param, complexity hjorth param, permutation entropy, approximate entropy, total power spectral density, determinism, average diagonal line length, compute recurrence rate, spectral edge frequency 75, petrosian fractal dimension

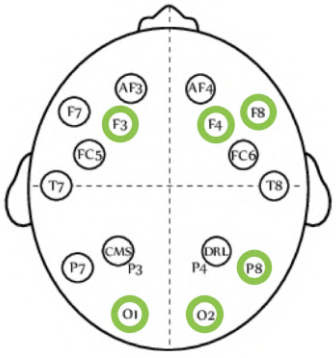


Figure 7.18: Optimal electrodes found for GBM.

7.1.15 GBM evaluation

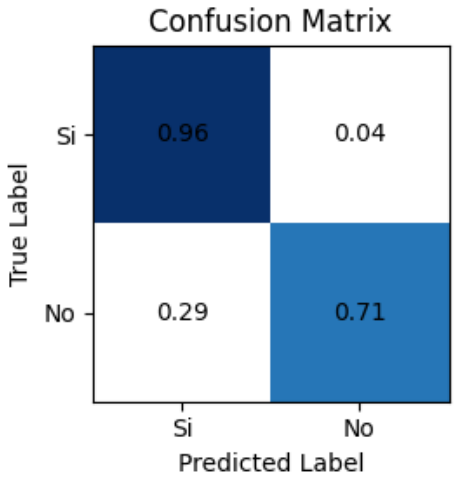


Figure 7.19: NB confusion matrix.

	precision	recall	f1-score	support
0	0.84	0.96	0.90	28
1	0.92	0.71	0.80	17
accuracy			0.87	45
macro avg	0.88	0.84	0.85	45
weighted avg	0.87	0.87	0.86	45

Figure 7.20: NB performance metrics.

7.1.16 Ada-GA performance

Fig. 7.21 corresponds to the GA combined with the Ada Boost classifier model. best

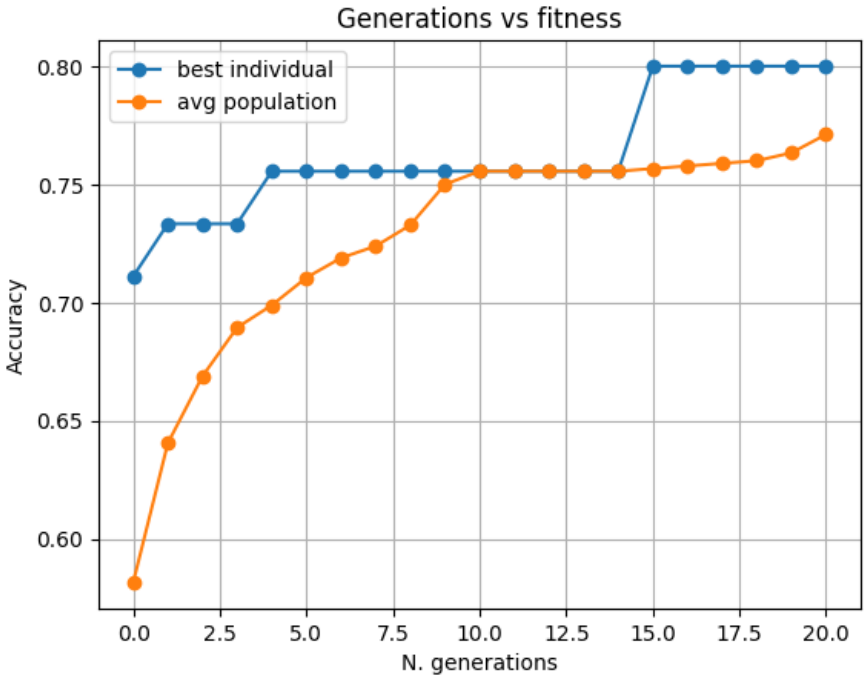


Figure 7.21: AdaBoost-GA training performance.

genome =
(0,0,1,0,0,1,0,1,1,1,0,0,0,0,1,1,0,0,0,0,0,0,1,0,1,1,0,1,1,0,1,0,1,0,1,1,0,1,0,0,0,
0,0,0,1,1,0,1,1,1,0,0,1,1,0,1,0,0,0,0,0,0,0,1,1,1,1,0,0,0,0,0,0,1,0,0,1,1,0,0,1,0,1,
1,0,0,1,1,1,0,1,1,0,0,1,1,0,1,0,1,1,1,0,0,1,1,1,1,0,0,1,0,1,1,1,0,1,0,1,1,1,1,1,1,1,
1,0,0,1,0,1,1,0,0,1,0,1,1,1,0,0,1,0,1,1,0,0,0,1,0,0,1,1,1,1,0,1,0,1,0,0,1,0,1,1,1,0,
1,0,0,0,0,1,1,0,0,0,1,0,1,0,0,0,1,0,1,1,0,0,0,1,1,0,0,1,0,0,1,1,1,1,0,0,1,0,0,1,1,0,
1,0,0,1,1,1,1,0,1,1,1,0,1,1,0,0,0,0,0)

7.1.17 Decode genome for Ada

5 selected electrodes: F3, P7, O2, P8, T8

15 features for raw data: mean, std, third quartile, kurtosis, skewness, activity hjorth param, mobility hjorth param, permutation entropy, spectral entropy, total power spectral density, centroid power spectral density, trapping time, hurst exponent, singular valued decomposition entropy, katz fractal dimension

15 features for delta band: mean, std, mode, max, first quartile, complexity hjorth param, permutation entropy, approximate entropy, spectral entropy, average diagonal line length, spectral edge frequency 50, spectral edge frequency 75, petrosian

fractal dimension, relative band power, band amplitude

22 features for theta band: coeff var, median, mode, min, first quartile, kurtosis, skewness, activity hjorth param, complexity hjorth param, permutation entropy, approximate entropy, total power spectral density, centroid power spectral density, determinism, trapping time, compute recurrence rate, spectral edge frequency 50, spectral edge frequency 75, hurst exponent, petrosian fractal dimension, relative band power, band amplitude

21 features for alpha band: mean, std, coeff var, median, mode, max, third quartile, kurtosis, skewness, mobility hjorth param, permutation entropy, approximate entropy, spectral entropy, centroid power spectral density, trapping time, diagonal line entropy, spectral edge frequency 50, singular valued decomposition entropy, petrosian fractal dimension, katz fractal dimension, relative band power

16 features for beta band: mean, coeff var, max, first quartile, third quartile, inter quartile range, skewness, permutation entropy, approximate entropy, centroid power spectral density, trapping time, spectral edge frequency 25, spectral edge frequency 75, hurst exponent, relative band power, band amplitude

18 features for gamma band: coeff var, max, min, first quartile, third quartile, skewness, mobility hjorth param, complexity hjorth param, approximate entropy, total power spectral density, centroid power spectral density, determinism, trapping time, average diagonal line length, compute recurrence rate, spectral edge frequency 25, spectral edge frequency 75, hurst exponent

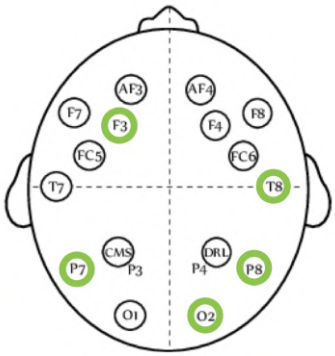


Figure 7.22: Optimal electrodes found for Ada.

7.1.18 Ada evaluation

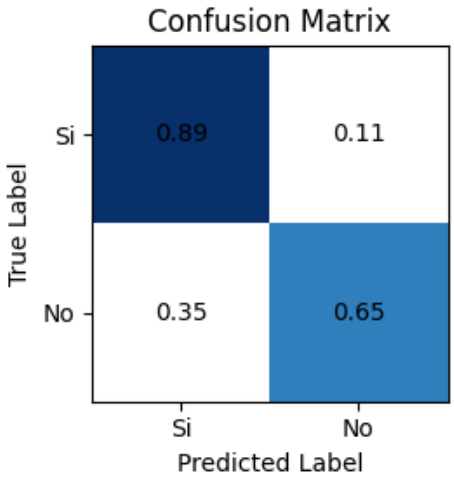


Figure 7.23: Ada confusion matrix.

	precision	recall	f1-score	support
0	0.81	0.89	0.85	28
1	0.79	0.65	0.71	17
accuracy			0.80	45
macro avg	0.80	0.77	0.78	45
weighted avg	0.80	0.80	0.80	45

Figure 7.24: Ada performance metrics.

7.1.19 NB-GA performance

Fig. 7.25 corresponds to the GA combined with the Naïve Bayes classifier model. best

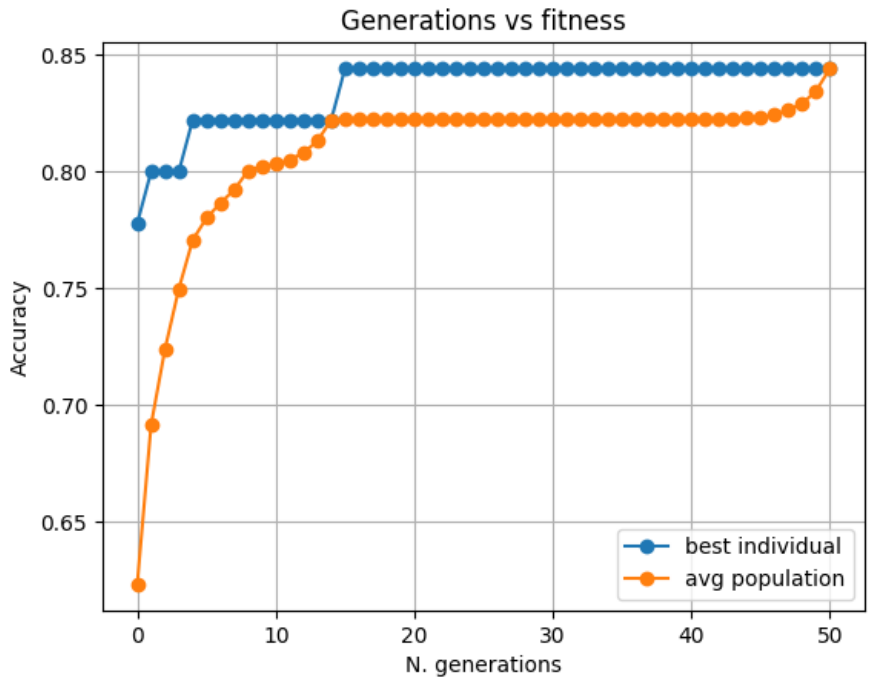


Figure 7.25: NB-GA training performance.

genome =
(1,0,1,1,0,1,1,1,0,0,0,0,0,1,0,0,1,1,1,0,0,0,0,0,1,0,0,0,0,0,0,0,1,1,0,0,1,1,0,1,1,1,
1,1,0,0,0,0,1,1,0,0,0,1,1,0,1,1,0,0,0,1,1,0,0,1,0,1,1,1,0,0,0,0,1,0,0,1,0,1,0,0,0,1,1,
1,0,0,1,1,0,1,0,1,0,1,1,1,1,1,1,1,0,0,0,1,1,1,0,0,0,0,1,1,0,1,0,1,0,1,1,0,0,1,0,0,0,1,0,
1,1,0,0,1,0,0,0,1,1,0,1,1,0,0,1,0,0,1,0,1,0,1,0,1,0,0,0,1,1,1,0,1,0,0,1,1,0,1,1,0,1,1,
0,1,1,1,1,1,0,0,0,0,0,0,1,0,0,1,1,0,0,1,1,0,1,0,1,1,1,0,1,0,0,0,1,0,1,1,0,0,0,1,0,0,0,
0,0,1,1,1,0,1,1,0,1,1,0,0,1)

7.1.20 Decode genome for NB

7 selected electrodes: AF3, F3, FC5, P7, O1, O2, AF4

13 features for raw data: coeff var, median, mode, kurtosis, spectral entropy, higuchi fractal dimension, determinism, trapping time, average diagonal line length, compute recurrence rate, spectral edge frequency 25, spectral edge frequency 50, spectral edge frequency 75

16 features for delta band: mean, std, max, min, third quartile, inter quartile range, activity hjorth param, mobility hjorth param, approximate entropy, higuchi frac-

tal dimension, total power spectral density, centroid power spectral density, compute recurrence rate, spectral edge frequency 75, singular valued decomposition entropy, band amplitude

21 features for theta band: mean, std, mode, max, first quartile, inter quartile range, skewness, detrended fluctuation analysis, activity hjorth param, mobility hjorth param, complexity hjorth param, permutation entropy, total power spectral density, centroid power spectral density, determinism, spectral edge frequency 25, spectral edge frequency 50, hurst exponent, petrosian fractal dimension, relative band power, band amplitude

14 features for alpha band: coeff var, min, third quartile, inter quartile range, detrended fluctuation analysis, permutation entropy, approximate entropy, higuchi fractal dimension, total power spectral density, trapping time, compute recurrence rate, spectral edge frequency 50, hurst exponent, petrosian fractal dimension

20 features for beta band: mean, std, coeff var, mode, first quartile, third quartile, kurtosis, skewness, activity hjorth param, mobility hjorth param, permutation entropy, approximate entropy, spectral entropy, higuchi fractal dimension, total power spectral density, spectral edge frequency 25, hurst exponent, singular valued decomposition entropy, relative band power, band amplitude

17 features for gamma band: std, median, mode, max, first quartile, skewness, activity hjorth param, mobility hjorth param, spectral entropy, diagonal line entropy, average diagonal line length, compute recurrence rate, spectral edge frequency 50, spectral edge frequency 75, singular valued decomposition entropy, petrosian fractal dimension, band amplitude

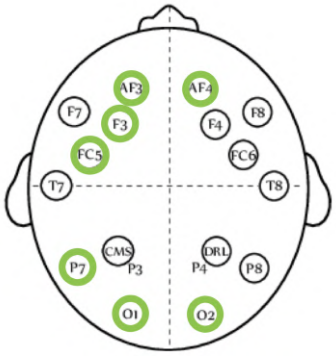


Figure 7.26: Optimal electrodes found for NB.

7.1.21 NB evaluation

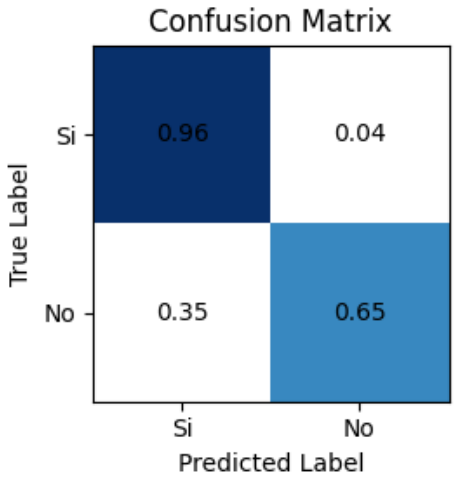


Figure 7.27: NB confusion matrix.

	precision	recall	f1-score	support
0	0.82	0.96	0.89	28
1	0.92	0.65	0.76	17
accuracy			0.84	45
macro avg	0.87	0.81	0.82	45
weighted avg	0.86	0.84	0.84	45

Figure 7.28: NB performance metrics.

7.1.22 LDA-GA performance

Fig. 7.29 corresponds to the GA combined with the Linear Discriminant Analysis classifier model. best genome =

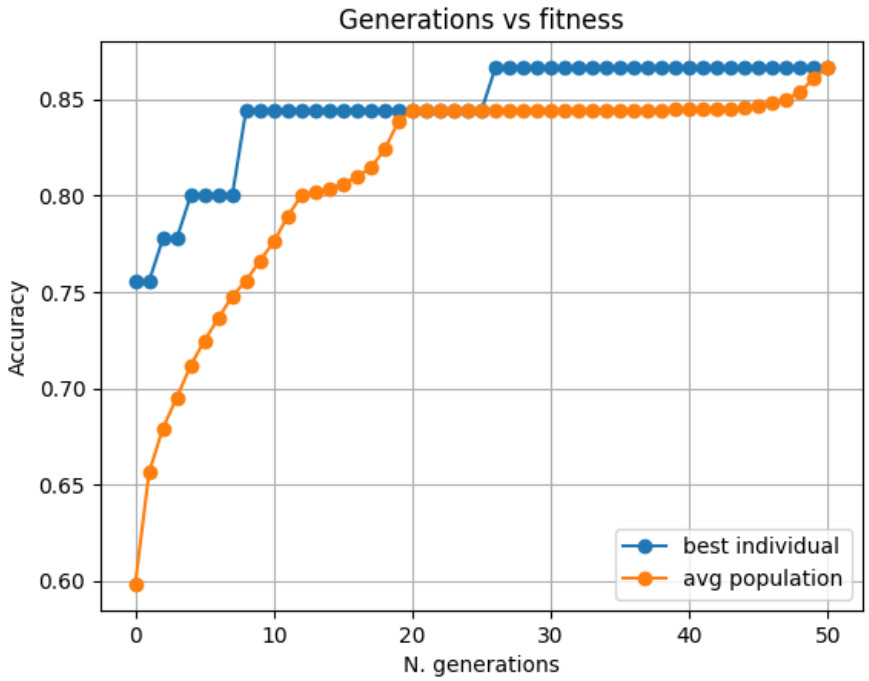


Figure 7.29: LDA-GA training performance.

(0,1,1,0,0,1,1,1,1,1,1,0,1,0,0,0,0,0,1,0,1,1,0,1,1,0,1,1,1,0,0,0,1,1,0,1,1,0,0,1,1,
1,0,0,0,0,0,1,1,1,1,0,1,1,1,1,0,1,1,0,1,0,0,0,0,1,1,0,1,1,1,0,1,0,0,1,0,1,1,1,1,0,0,0,
1,0,1,0,1,1,1,1,1,0,0,0,1,1,1,1,1,1,1,1,1,0,0,1,0,0,0,1,1,1,0,0,1,1,1,1,1,0,1,1,1,1,0,0,
1,1,0,0,1,0,1,0,1,1,1,1,0,1,1,1,0,1,1,0,1,1,0,1,1,0,1,0,1,0,0,0,0,0,1,0,0,0,1,0,1,0,1,0,1,1,0,0,
0,0,0,1,1,1,1,0,0,0,0,1,0,0,0,1,0,1,0,0,0,0,0,1,0,0,1,1,0,0,1,1,0,1,0,0,0,1,0,0,1,1,
0,0,1,0,0,0,0,1,0,0,0,0,0,1,1,0,0,1,0)

7.1.23 Decode genome for LDA

9 selected electrodes: F7, F3, P7, O1, O2, P8, T8, FC6, F8

16 features for raw data: mode, min, first quartile, inter quartile range, kurtosis, detrended fluctuation analysis, activity hjorth param, mobility hjorth param, spectral entropy, higuchi fractal dimension, centroid power spectral density, determinism, average diagonal line length, compute recurrence rate, spectral edge frequency 25, katz fractal dimension

21 features for delta band: mean, std, coeff var, mode, max, min, first quartile, inter quartile range, kurtosis, detrended fluctuation analysis, approximate entropy, spectral entropy, total power spectral density, centroid power spectral density, determinism, diagonal line entropy, spectral edge frequency 25, spectral edge frequency 75, hurst exponent, singular valued decomposition entropy, band amplitude

23 features for theta band: std, median, mode, max, min, first quartile, skewness, detrended fluctuation analysis, activity hjorth param, mobility hjorth param, complexity hjorth param, permutation entropy, approximate entropy, spectral entropy, centroid power spectral density, average diagonal line length, compute recurrence rate, spectral edge frequency 25, hurst exponent, singular valued decomposition entropy, petrosian fractal dimension, katz fractal dimension, band amplitude

20 features for alpha band: mean, std, coeff var, max, min, inter quartile range, skewness, activity hjorth param, mobility hjorth param, complexity hjorth param, permutation entropy, spectral entropy, higuchi fractal dimension, total power spectral density, determinism, trapping time, average diagonal line length, compute recurrence rate, spectral edge frequency 50, katz fractal dimension

13 features for beta band: std, median, max, first quartile, third quartile, mobility hjorth param, complexity hjorth param, permutation entropy, approximate entropy, determinism, compute recurrence rate, spectral edge frequency 50, relative band power

13 features for gamma band: std, coeff var, max, min, third quartile, detrended fluctuation analysis, complexity hjorth param, permutation entropy, higuchi fractal dimension, diagonal line entropy, hurst exponent, singular valued decomposition entropy, relative band power

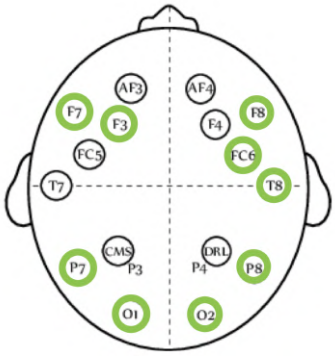


Figure 7.30: Optimal electrodes found for LDA.

7.1.24 LDA evaluation

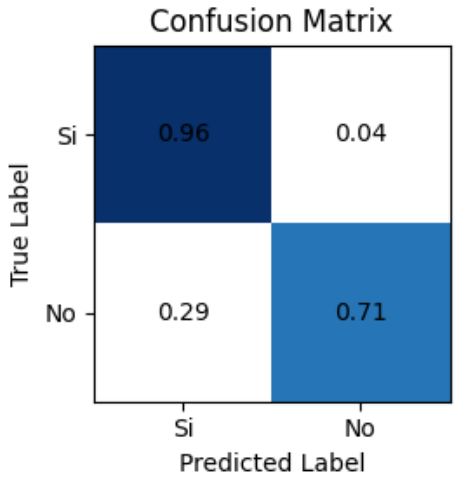


Figure 7.31: LDA confusion matrix.

	precision	recall	f1-score	support
0	0.84	0.96	0.90	28
1	0.92	0.71	0.80	17
accuracy			0.87	45
macro avg	0.88	0.84	0.85	45
weighted avg	0.87	0.87	0.86	45

Figure 7.32: LDA performance metrics.

7.1.25 MLP-GA performance

Fig. 7.33 corresponds to the GA combined with the Multi Layer Perceptron classifier model. best genome =

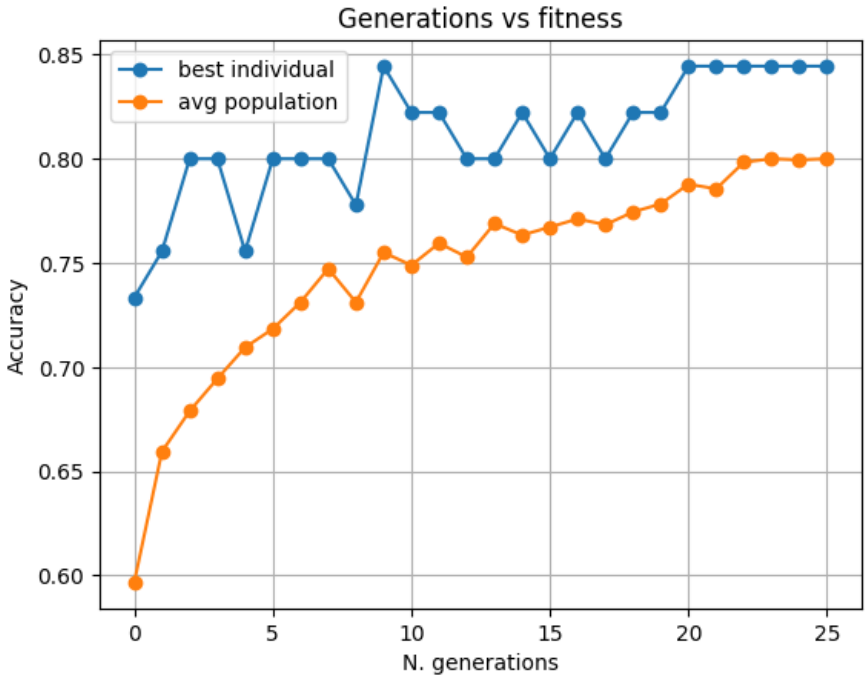


Figure 7.33: MLP-GA training performance.

(0,1,0,0,1,1,1,1,1,0,0,1,1,1,0,0,1,1,1,1,0,0,1,1,0,1,1,1,1,0,1,1,0,1,0,0,1,0,0,0,1,0,
1,0,0,1,0,0,1,0,1,0,1,1,0,0,0,1,1,0,0,1,1,0,1,1,0,0,0,0,1,1,1,1,1,0,0,1,1,0,0,0,0,1,1,
1,1,1,1,1,1,0,1,1,1,0,0,1,0,1,1,1,0,1,1,1,0,0,0,1,1,0,0,0,1,1,0,0,0,1,1,1,0,0,0,0,0,
1,1,0,0,1,1,0,1,1,1,1,1,1,0,1,0,0,1,1,0,1,0,1,0,0,0,0,0,1,0,1,0,0,0,1,0,1,1,0,1,0,1,1,
1,0,0,0,1,1,0,0,1,1,0,1,1,1,1,1,1,1,1,0,0,0,0,0,1,0,1,1,1,1,1,1,1,1,1,0,1,0,1,1,0,0,1,
1,0,0,0,0,0,0,1,0,1,0,0,1,0,0,0,0,1)

7.1.26 Decode genome for MLP

Here is the breakdown of the selected electrodes and features for each band for the MLP model:

9 selected electrodes: F7, T7, P7, O1, O2, P8, F4, F8, AF4

17 features for raw data: coeff var, median, mode, max, third quartile, inter quartile range, skewness, detrended fluctuation analysis, activity hjorth param, mobility hjorth param, permutation entropy, approximate entropy, higuchi fractal dimension, determinism, compute recurrence rate, spectral edge frequency 50, singular valued decomposition entropy

18 features for delta band: mean, coeff var, mode, max, inter quartile range, kurtosis, activity hjorth param, mobility hjorth param, permutation entropy, approximate entropy, determinism, trapping time, diagonal line entropy, average diagonal line length, spectral edge frequency 50, spectral edge frequency 75, relative band power, band amplitude

22 features for theta band: mean, std, coeff var, median, mode, max, first quartile, third quartile, inter quartile range, detrended fluctuation analysis, mobility hjorth param, complexity hjorth param, permutation entropy, spectral entropy, higuchi fractal dimension, total power spectral density, diagonal line entropy, average diagonal line length, spectral edge frequency 75, hurst exponent, relative band power, band amplitude

18 features for alpha band: mean, min, first quartile, kurtosis, skewness, activity hjorth param, mobility hjorth param, complexity hjorth param, permutation entropy, approximate entropy, spectral entropy, total power spectral density, trapping time, diagonal line entropy, compute recurrence rate, spectral edge frequency 50, katz fractal dimension, band amplitude

19 features for beta band: median, max, min, third quartile, kurtosis, skewness, detrended fluctuation analysis, permutation entropy, approximate entropy, total power spectral density, centroid power spectral density, trapping time, diagonal line entropy, average diagonal line length, compute recurrence rate, spectral edge frequency 25, spectral edge frequency 50, spectral edge frequency 75, band amplitude

18 features for gamma band: std, coeff var, median, mode, max, min, first quartile, third quartile, inter quartile range, skewness, activity hjorth param, mobility hjorth param, approximate entropy, spectral entropy, average diagonal line length, spectral edge frequency 25, hurst exponent, band amplitude

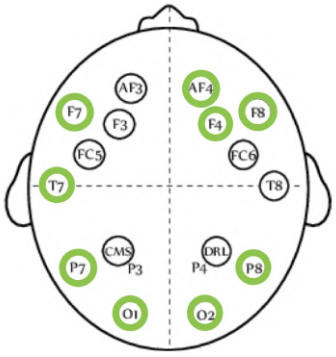


Figure 7.34: Optimal electrodes found for MLP.

7.1.27 MLP evaluation

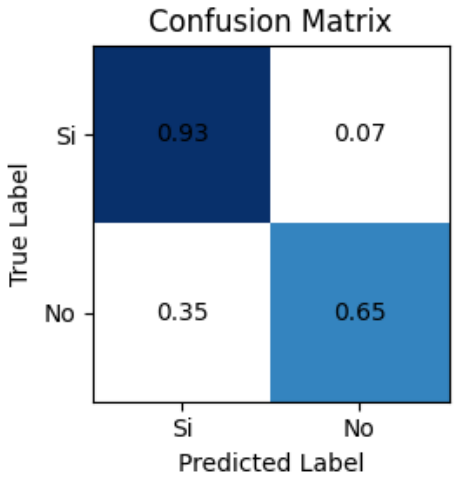


Figure 7.35: MLP confusion matrix.

	precision	recall	f1-score	support
0	0.81	0.93	0.87	28
1	0.85	0.65	0.73	17
accuracy			0.82	45
macro avg	0.83	0.79	0.80	45
weighted avg	0.83	0.82	0.82	45

Figure 7.36: MLP performance metrics.

7.1.28 SVM-GA performance

Fig. 7.37 corresponds to the GA combined with the Suppoert Vector Machine classifier model. best genome =

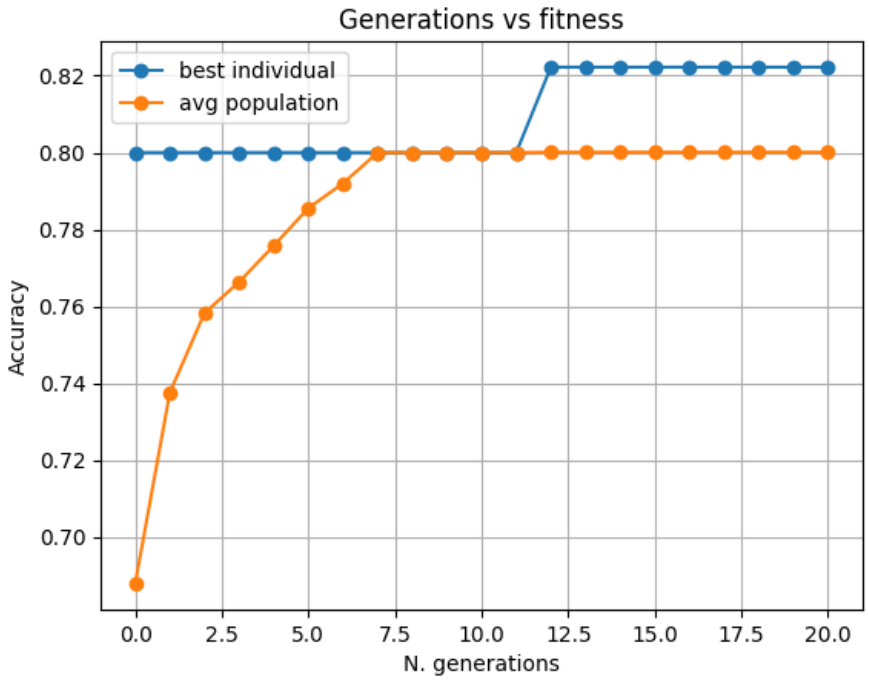


Figure 7.37: SVM-GA training performance.

(0,0,0,0,1,0,1,0,1,0,0,1,0,0,0,1,0,0,1,1,0,1,1,0,1,1,0,1,0,1,0,1,0,0,1,0,0,1,1,1,1,0,
1,1,0,0,1,0,0,0,0,0,1,1,0,0,0,1,0,1,0,1,1,0,0,1,1,1,0,0,1,0,0,0,0,0,0,1,1,1,0,0,1,0,
0,1,0,1,0,0,0,0,0,0,1,1,1,0,0,1,1,1,1,1,0,0,0,1,0,1,1,0,1,1,1,1,1,0,1,1,1,1,0,0,1,1,
1,1,0,1,0,1,1,1,0,1,0,0,1,0,0,1,1,1,0,1,0,1,1,0,0,1,1,0,1,1,0,0,0,0,0,1,0,1,1,0,0,1,
0,1,1,0,0,1,1,1,0,0,1,0,0,0,1,0,1,0,0,1,1,0,1,0,0,0,0,1,1,0,0,0,0,1,0,1,0,1,1,0,0,1,
1,1,0,0,1,0,1,0,0,0,0,0,0,1,0,0,0,0)

7.1.29 Decode genome for SVM

Here is the breakdown of the selected electrodes and features for each band for the SVM model:

5 selected electrodes: F3, T7, O1, P8, FC6

19 features for raw data: std, coeff var, mode, first quartile, kurtosis, skewness, mobility hjorth param, permutation entropy, approximate entropy, spectral entropy, higuchi fractal dimension, total power spectral density, trapping time, diagonal line entropy, compute recurrence rate, spectral edge frequency 25, spectral edge frequency

50, hurst exponent, petrosian fractal dimension

22 features for delta band: std, mode, max, min, third quartile, inter quartile range, skewness, detrended fluctuation analysis, activity hjorth param, mobility hjorth param, complexity hjorth param, approximate entropy, centroid power spectral density, trapping time, average diagonal line length, compute recurrence rate, spectral edge frequency 50, hurst exponent, petrosian fractal dimension, katz fractal dimension, relative band power, band amplitude

18 features for theta band: std, mode, max, min, first quartile, third quartile, detrended fluctuation analysis, permutation entropy, approximate entropy, spectral entropy, higuchi fractal dimension, centroid power spectral density, determinism, diagonal line entropy, hurst exponent, katz fractal dimension, relative band power, band amplitude

13 features for alpha band: mean, mode, max, min, third quartile, inter quartile range, kurtosis, mobility hjorth param, spectral entropy, higuchi fractal dimension, compute recurrence rate, petrosian fractal dimension, relative band power

19 features for beta band: mean, std, coeff var, median, max, min, third quartile, kurtosis, activity hjorth param, mobility hjorth param, approximate entropy, spectral entropy, higuchi fractal dimension, centroid power spectral density, diagonal line entropy, average diagonal line length, compute recurrence rate, spectral edge frequency 50, katz fractal dimension

19 features for gamma band: mean, std, coeff var, median, min, first quartile, third quartile, inter quartile range, kurtosis, detrended fluctuation analysis, activity hjorth param, mobility hjorth param, centroid power spectral density, trapping time, diagonal line entropy, spectral edge frequency 50, petrosian fractal dimension, relative band power, band amplitude

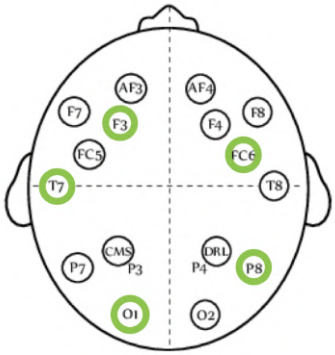


Figure 7.38: Optimal electrodes found for SVM.

7.1.30 SVM evaluation

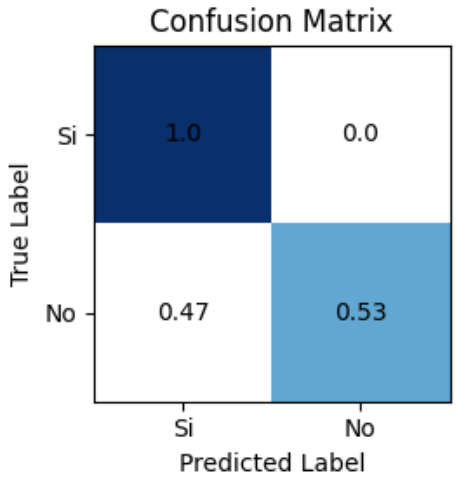


Figure 7.39: SVM confusion matrix.

	precision	recall	f1-score	support
0	0.78	1.00	0.88	28
1	1.00	0.53	0.69	17
accuracy			0.82	45
macro avg	0.89	0.76	0.78	45
weighted avg	0.86	0.82	0.81	45

Figure 7.40: SVM performance metrics.

7.1.31 Binary overall results for /Sí/ and /No/

In Table 7.1 can be observed the overall results for binary classification using /Sí/ and /No/ classes. It is presented information about the number of electrodes employed as well as the number of features, in addition, metrics such as Recall, F1-Score, Precision and Accuracy are shown. This information is presented because they are the simplest words in the set and its convenient to observe it this concise form.

In addition, certain values are highlighted in bold in Table 7.1. This formatting is used to emphasize the following:

- Classifiers using the fewest electrodes: Logistic Regression and Ada Boost, both employing only 5 electrodes
- Classifiers using the fewest electrodes: GBM employing only 99 features
- Classifier with highest recall: kNN
- Classifier with highest f1-score: kNN
- Classifier with highest precision: kNN
- Classifier with highest accuracy: kNN

Table 7.1: Model performance results.

Classifier model	**Electrodes** (Max: 14)	**Features** (Max: 214)	**Recall**	**F1-Score**	**Precision**	**Accuracy**
kNN	8	103	**0.97**	**0.94**	**0.95**	**0.96**
Logistic Regression	**5**	107	0.87	0.85	0.85	0.87
Random Forest	7	106	0.90	0.79	0.81	0.84
Decision Tree	8	101	0.89	0.88	0.88	0.89
GBM	6	**99**	0.88	0.84	0.85	0.87
AdaBoost	**5**	107	0.80	0.77	0.78	0.80
Naive Bayes	7	101	0.87	0.81	0.82	0.84
LDA	9	106	0.88	0.84	0.85	0.87
MLP	9	112	0.83	0.79	0.80	0.82
SVM	**5**	110	0.89	0.76	0.78	0.82

7.2 One-to-one classification for the nine classes

As mentioned in Chapter 6, this approach has strength in binary classification, therefore, the following results corresponds to the one-to-one classification accuracy.

7.2.1 One-to-one kNN classifier

Figures 7.41 and 7.42 show the performance of the GA in conjunction with the kNN classifier. Note that since it is a one-to-one classification, there will be 36 subsets of words. For this reason, there are 36 different graphs. It is also possible to notice that all the graphs start with a relatively low accuracy, but this accuracy increases as the generations of the genetic algorithm progress.

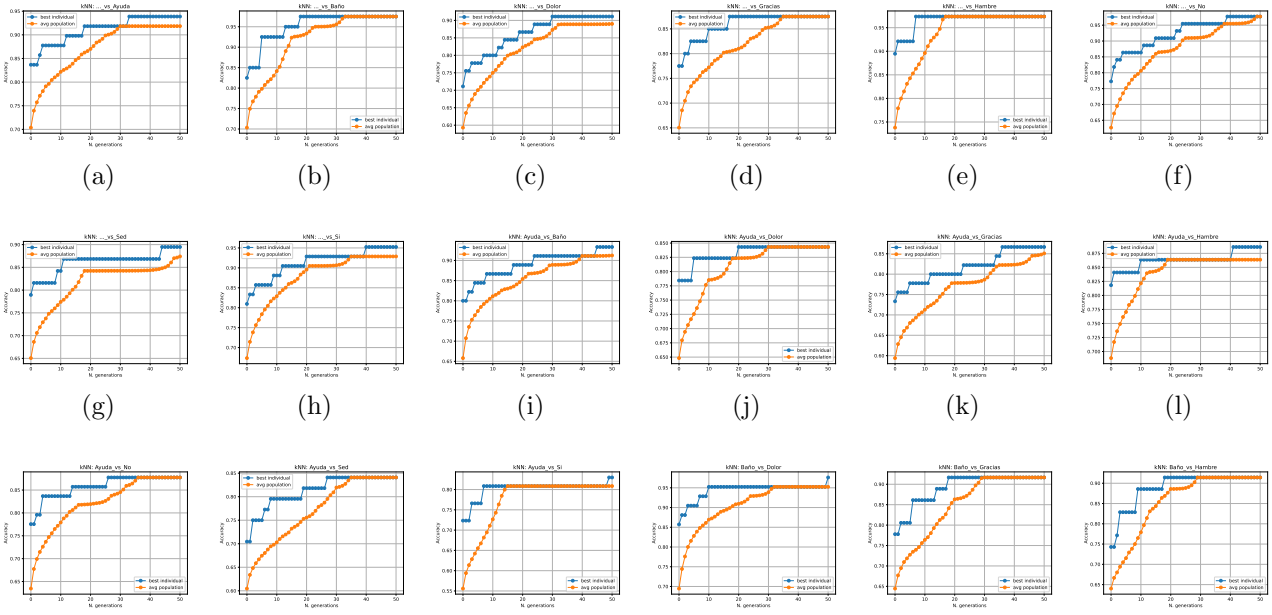


Figure 7.41: GAs performance for the one-to-one subsets using kNN classifier.

(**a**) ... vs Ayuda. (**b**) ... vs Baño. (**c**) ... vs Dolor. (**d**) ... vs Gracias. (**e**) ... vs Hambre. (**f**) ... vs No. (**g**) ... vs Sed. (**h**) ... vs Sí. (**i**) Ayuda vs Baño. (**j**) Ayuda vs Dolor. (**k**) Ayuda vs Gracias. (**l**) Ayuda vs Hambre. (**m**) Ayuda vs No. (**n**) Ayuda vs Sed. (**o**) Ayuda vs Sí. (**p**) Ayuda vs Dolor. (**q**) Ayuda vs Gracias. (**r**) Baño vs Hambre.

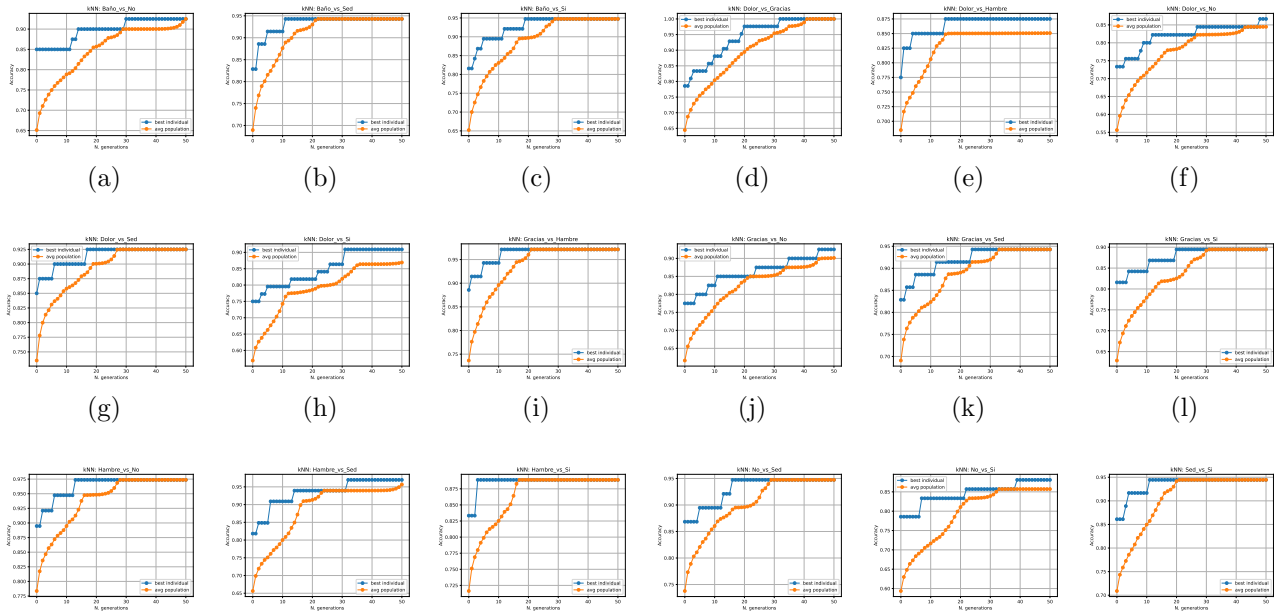


Figure 7.42: GAs performance for the one-to-one subsets using kNN classifier.

(**a**) Baño vs No. (**b**) Baño vs Sed. (**c**) Baño vs Sí. (**d**) Dolor vs Gracias. (**e**) Dolor vs Hambre. (**f**) Dolor vs No. (**g**) Dolor vs Sed. (**h**) Dolor vs Sí. (**i**) Gracias vs Hambre. (**j**) Gracias vs No. (**k**) Gracias vs Sed. (**l**) Gracias vs Sí. (**m**) Hambre vs No. (**n**) Hambre vs Sed. (**o**) Hambre vs Sí. (**p**) No vs Sed. (**q**) No vs Sí. (**r**) Sed vs Sí.

Table 7.2 shows the highest accuracy value found by the GA for each of the 36 data subsets in the one-to-one classification for the kNN algorithm.

	...	Sí	No	Baño	Hambre	Sed	Ayuda	Dolor	Gracias
...									
Sí	0.94								
No	0.98	0.93							
Baño	0.91	0.84	0.98						
Hambre	0.88	0.87	0.92	1.00					
Sed	0.90	0.87	0.93	0.98	0.98				
Ayuda	0.91	0.85	0.94	0.98	0.95	0.98			
Dolor	0.91	0.86	0.93	0.97	0.94	0.97	0.98		
Gracias	0.94	0.89	0.94	0.98	0.98	0.99	0.99	0.97	

Table 7.2: One-to-one accuracies for kNN classifier.

7.2.2 One-to-one LogReg classifier

Figures 7.43 and 7.44 show the performance of the GA in conjunction with the LogReg classifier. Note that since it is a one-to-one classification, there will be 36 subsets of words. For this reason, there are 36 different graphs. It is also possible to notice that all the graphs start with a relatively low accuracy, but this accuracy increases as the generations of the genetic algorithm progress.

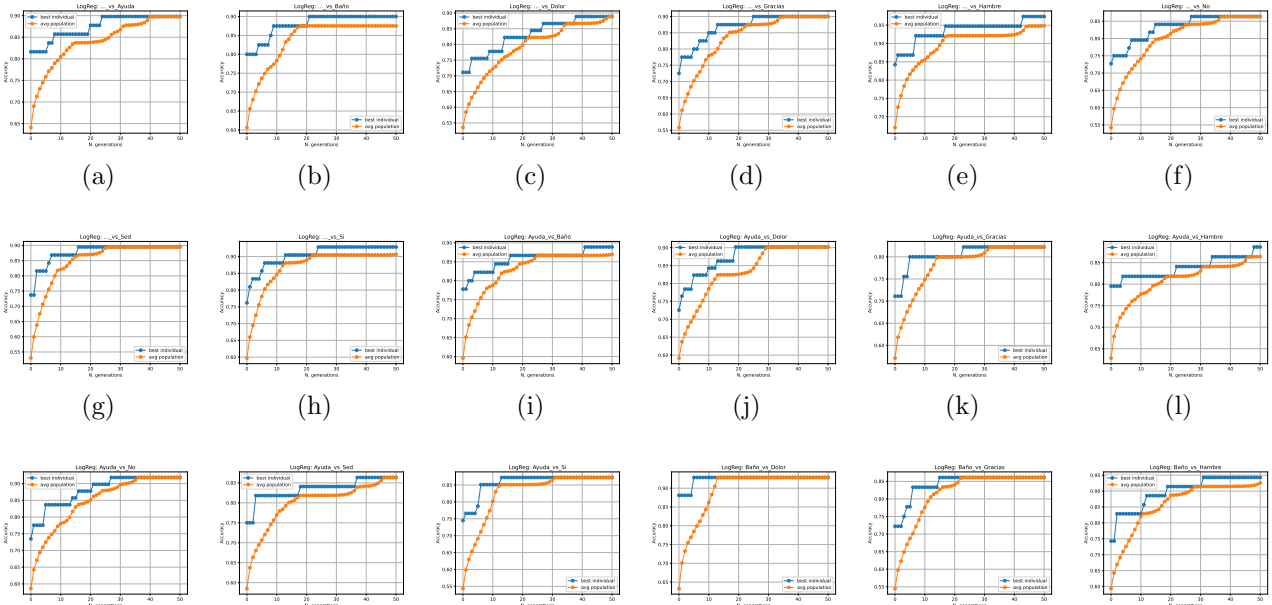


Figure 7.43: GAs performance for the one-to-one subsets using LogReg classifier.

(**a**) ... vs Ayuda. (**b**) ... vs Baño. (**c**) ... vs Dolor. (**d**) ... vs Gracias. (**e**) ... vs Hambre. (**f**) ... vs No. (**g**) ... vs Sed. (**h**) ... vs Sí. (**i**) Ayuda vs Baño. (**j**) Ayuda vs Dolor. (**k**) Ayuda vs Gracias. (**l**) Ayuda vs Hambre. (**m**) Ayuda vs No. (**n**) Ayuda vs Sed. (**o**) Ayuda vs Sí. (**p**) Ayuda vs Dolor. (**q**) Ayuda vs Gracias. (**r**) Baño vs Hambre.

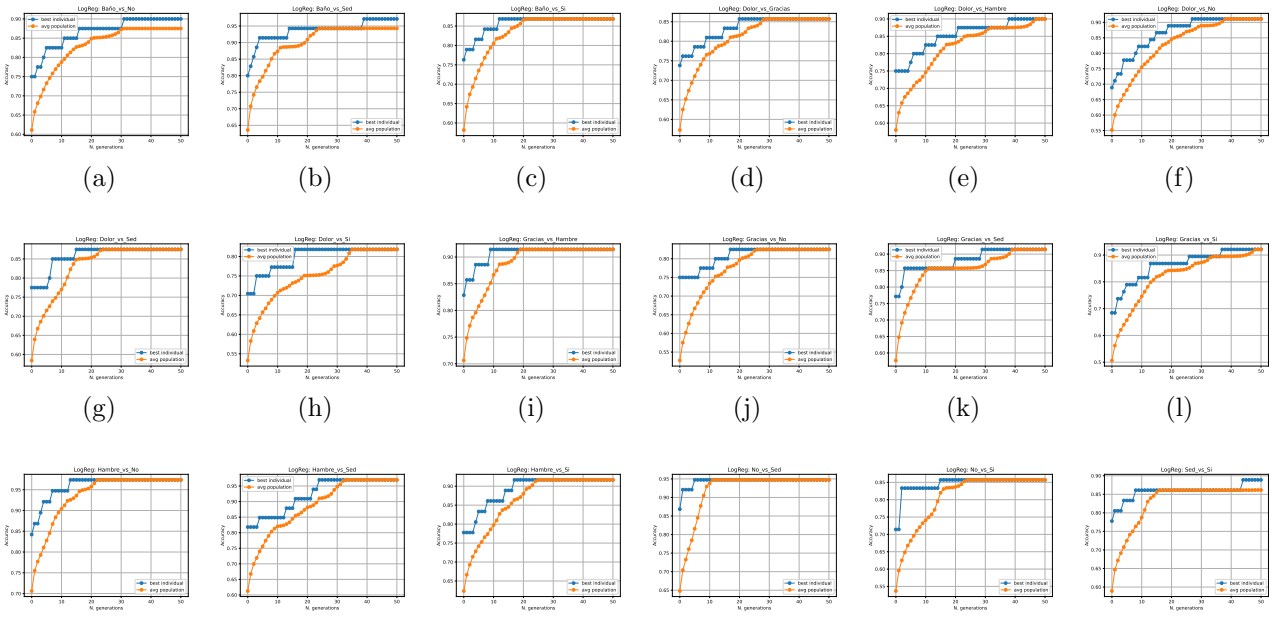


Figure 7.44: GAs performance for the one-to-one subsets using LogReg classifier.

(**a**) Baño vs No. (**b**) Baño vs Sed. (**c**) Baño vs Sí. (**d**) Dolor vs Gracias. (**e**) Dolor vs Hambre. (**f**) Dolor vs No. (**g**) Dolor vs Sed. (**h**) Dolor vs Sí. (**i**) Gracias vs Hambre. (**j**) Gracias vs No. (**k**) Gracias vs Sed. (**l**) Gracias vs Sí. (**m**) Hambre vs No. (**n**) Hambre vs Sed. (**o**) Hambre vs Sí. (**p**) No vs Sed. (**q**) No vs Sí. (**r**) Sed vs Sí.

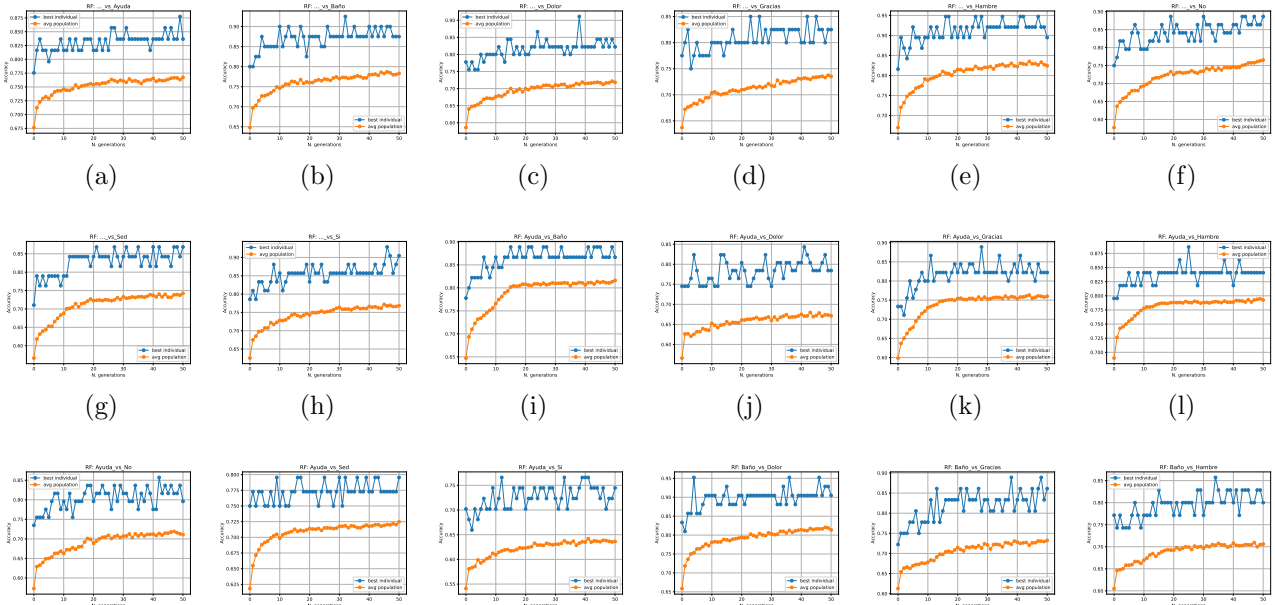
Table 7.3 shows the highest accuracy value found by the GA for each of the 36 data subsets in the one-to-one classification for the LogReg algorithm.

	...	Sí	No	Baño	Hambre	Sed	Ayuda	Dolor	Gracias
...									
Sí	0.9								
No	0.9	0.89							
Baño	0.89	0.9	0.93						
Hambre	0.9	0.82	0.86	0.86					
Sed	0.97	0.89	0.94	0.9	0.91				
Ayuda	0.86	0.92	0.9	0.91	0.82	0.97			
Dolor	0.89	0.86	0.97	0.88	0.91	0.97	0.95		
Gracias	0.93	0.87	0.87	0.82	0.92	0.92	0.86	0.89	

Table 7.3: One-to-one accuracies for LogReg classifier.

7.2.3 One-to-one RF classifier

Figures 7.45 and 7.46 show the performance of the GA in conjunction with the RF classifier. Note that since it is a one-to-one classification, there will be 36 subsets of words. For this reason, there are 36 different graphs. It is also possible to notice that all the graphs start with a relatively low accuracy, but this accuracy increases as the generations of the genetic algorithm progress.



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

(m) (n) (o) (p) (q) (r)

Figure 7.45: GAs performance for the one-to-one subsets using RF classifier.

(**a**) ... vs Ayuda. (**b**) ... vs Baño. (**c**) ... vs Dolor. (**d**) ... vs Gracias. (**e**) ... vs Hambre. (**f**) ... vs No. (**g**) ... vs Sed. (**h**) ... vs Sí. (**i**) Ayuda vs Baño. (**j**) Ayuda vs Dolor. (**k**) Ayuda vs Gracias. (**l**) Ayuda vs Hambre. (**m**) Ayuda vs No. (**n**) Ayuda vs Sed. (**o**) Ayuda vs Sí. (**p**) Ayuda vs Dolor. (**q**) Ayuda vs Gracias. (**r**) Baño vs Hambre.

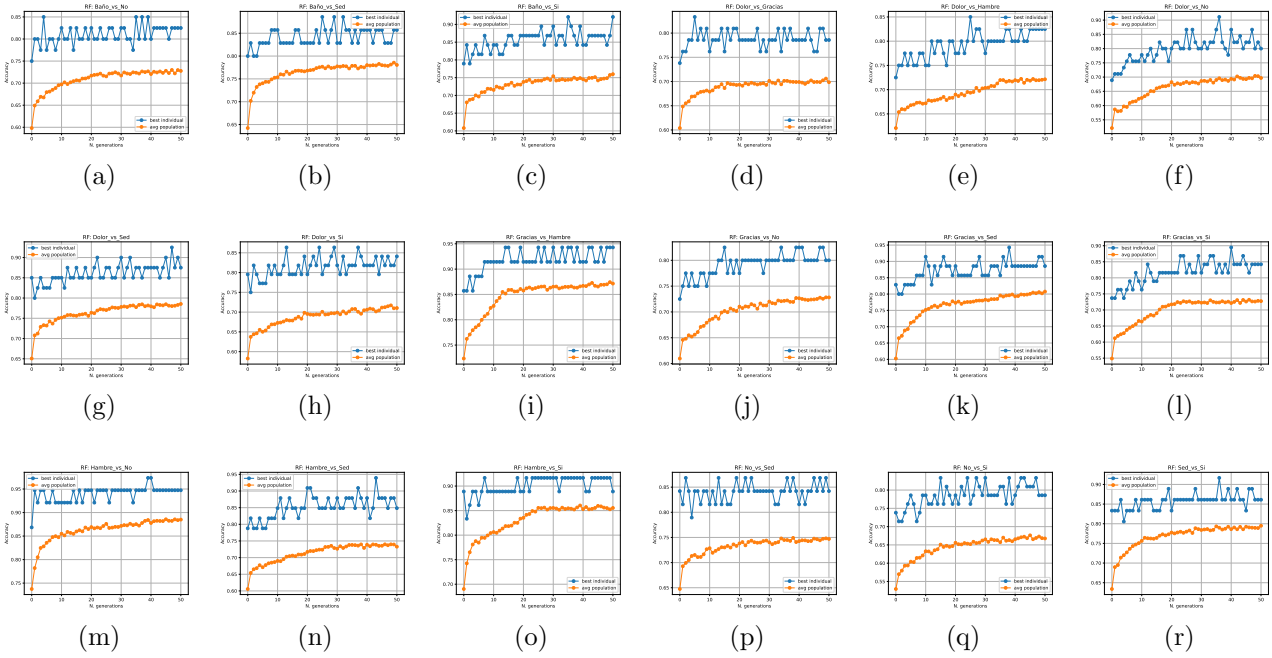


Figure 7.46: GAs performance for the one-to-one subsets using RF classifier.

(**a**) Baño vs No. (**b**) Baño vs Sed. (**c**) Baño vs Sí. (**d**) Dolor vs Gracias. (**e**) Dolor vs Hambre. (**f**) Dolor vs No. (**g**) Dolor vs Sed. (**h**) Dolor vs Sí. (**i**) Gracias vs Hambre. (**j**) Gracias vs No. (**k**) Gracias vs Sed. (**l**) Gracias vs Sí. (**m**) Hambre vs No. (**n**) Hambre vs Sed. (**o**) Hambre vs Sí. (**p**) No vs Sed. (**q**) No vs Sí. (**r**) Sed vs Sí.

Table 7.4 shows the highest accuracy value found by the GA for each of the 36 data subsets in the one-to-one classification for the RF algorithm.

	...	Sí	No	Baño	Hambre	Sed	Ayuda	Dolor	Gracias
...									
Sí	0.88								
No	0.92	0.89							
Baño	0.91	0.84	0.95						
Hambre	0.85	0.89	0.89	0.83					
Sed	0.95	0.89	0.86	0.85	0.94				
Ayuda	0.89	0.86	0.85	0.91	0.82	0.97			
Dolor	0.87	0.80	0.89	0.92	0.94	0.94	0.87		
Gracias	0.93	0.77	0.92	0.86	0.89	0.92	0.83	0.92	

Table 7.4: One-to-one accuracies for RF classifier.

7.2.4 One-to-one DT classifier

Figures 7.47 and 7.48 show the performance of the GA in conjunction with the DT classifier. Note that since it is a one-to-one classification, there will be 36 subsets of words. For this reason, there are 36 different graphs. It is also possible to notice that all the graphs start with a relatively low accuracy, but this accuracy increases as the generations of the genetic algorithm progress.

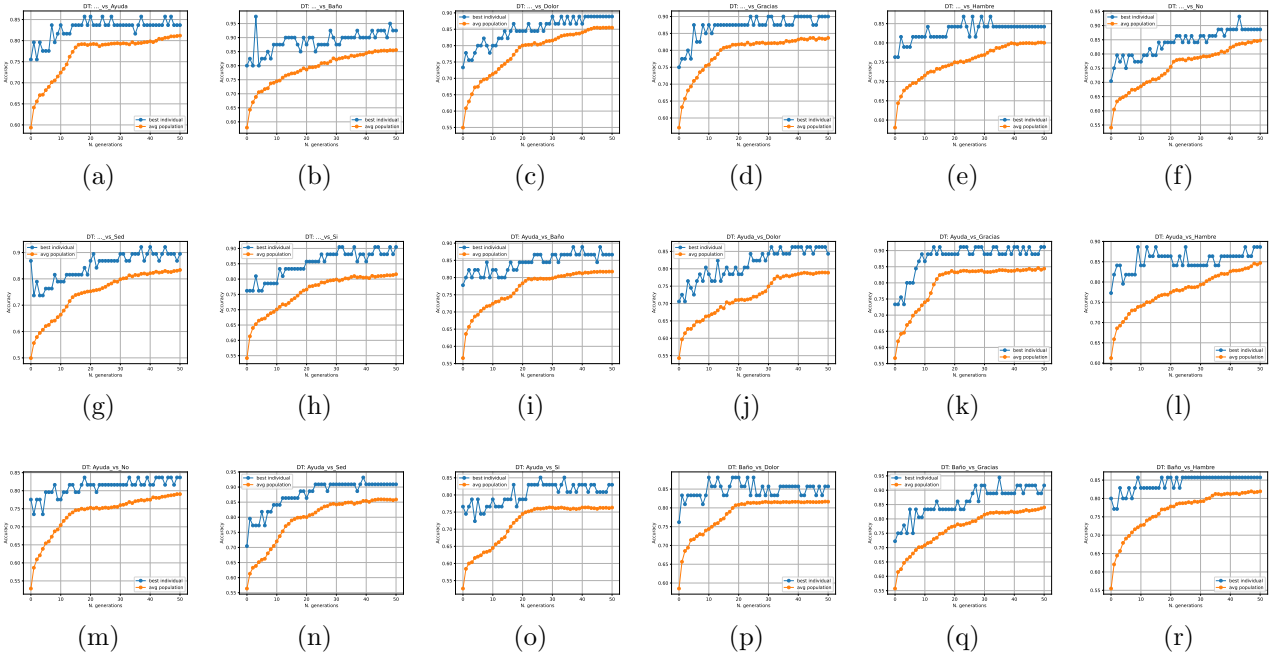


Figure 7.47: GAs performance for the one-to-one subsets using DT classifier.

(**a**) ... vs Ayuda. (**b**) ... vs Baño. (**c**) ... vs Dolor. (**d**) ... vs Gracias. (**e**) ... vs Hambre. (**f**) ... vs No. (**g**) ... vs Sed. (**h**) ... vs Sí. (**i**) Ayuda vs Baño. (**j**) Ayuda vs Dolor. (**k**) Ayuda vs Gracias. (**l**) Ayuda vs Hambre. (**m**) Ayuda vs No. (**n**) Ayuda vs Sed. (**o**) Ayuda vs Sí. (**p**) Ayuda vs Dolor. (**q**) Ayuda vs Gracias. (**r**) Baño vs Hambre.

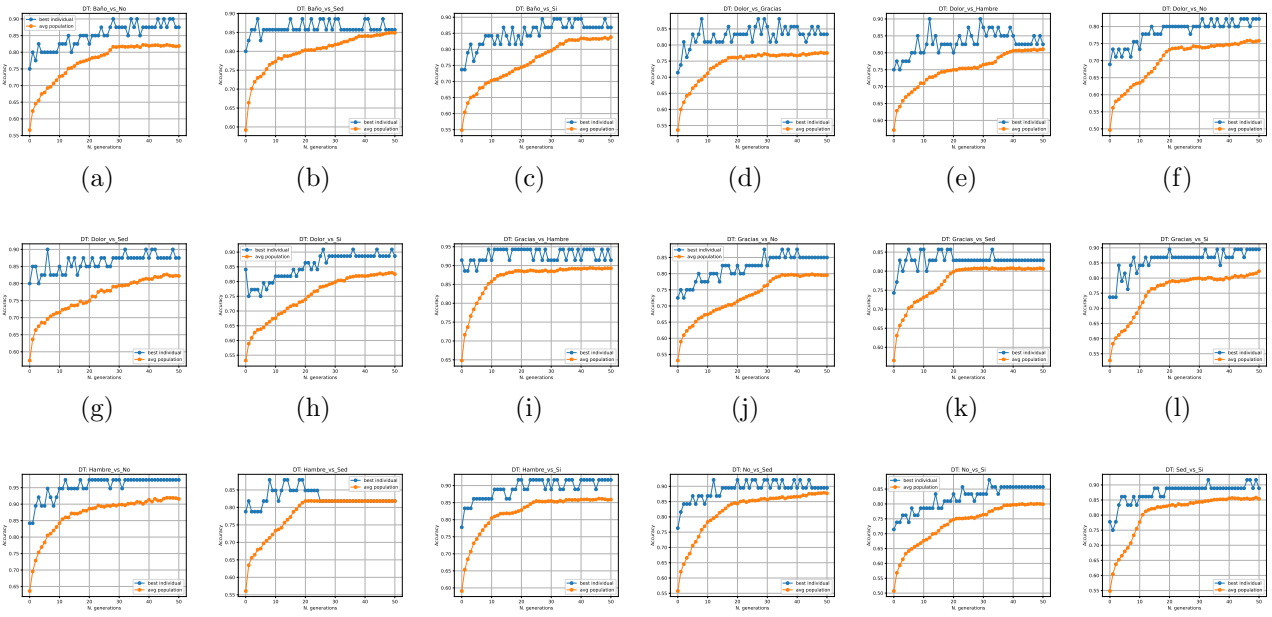


Figure 7.48: GAs performance for the one-to-one subsets using DT classifier.

(**a**) Baño vs No. (**b**) Baño vs Sed. (**c**) Baño vs Sí. (**d**) Dolor vs Gracias. (**e**) Dolor vs Hambre. (**f**) Dolor vs No. (**g**) Dolor vs Sed. (**h**) Dolor vs Sí. (**i**) Gracias vs Hambre. (**j**) Gracias vs No. (**k**) Gracias vs Sed. (**l**) Gracias vs Sí. (**m**) Hambre vs No. (**n**) Hambre vs Sed. (**o**) Hambre vs Sí. (**p**) No vs Sed. (**q**) No vs Sí. (**r**) Sed vs Sí.

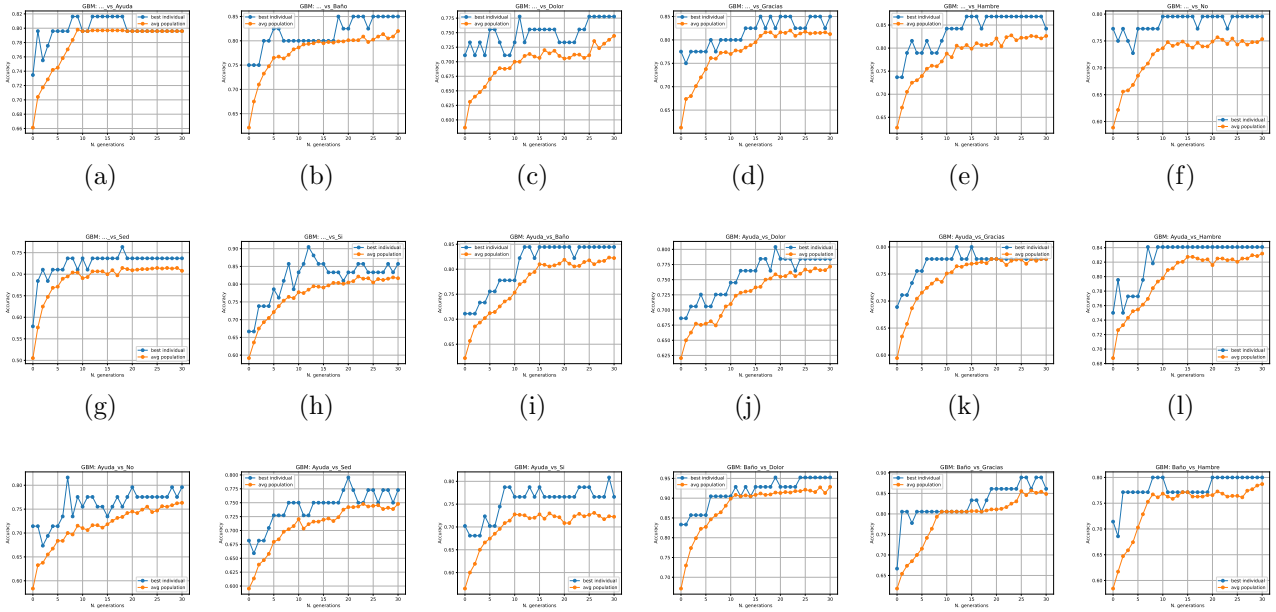
Table 7.5 shows the highest accuracy value found by the GA for each of the 36 data subsets in the one-to-one classification for the DT algorithm.

	...	Sí	No	Baño	Hambre	Sed	Ayuda	Dolor	Gracias
...									
Sí	0.86								
No	0.98	0.89							
Baño	0.89	0.86	0.88						
Hambre	0.90	0.91	0.94	0.88					
Sed	0.87	0.89	0.86	0.90	0.94				
Ayuda	0.93	0.84	0.90	0.82	0.88	0.97			
Dolor	0.92	0.93	0.89	0.90	0.86	0.88	0.92		
Gracias	0.90	0.85	0.89	0.91	0.89	0.92	0.88	0.92	

Table 7.5: One-to-one accuracies for DT classifier.

7.2.5 One-to-one GBM classifier

Figures 7.49 and 7.50 show the performance of the GA in conjunction with the GBM classifier. Note that since it is a one-to-one classification, there will be 36 subsets of words. For this reason, there are 36 different graphs. It is also possible to notice that all the graphs start with a relatively low accuracy, but this accuracy increases as the generations of the genetic algorithm progress.



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

(m) (n) (o) (p) (q) (r)

Figure 7.49: GAs performance for the one-to-one subsets using GBM classifier.

(**a**) ... vs Ayuda. (**b**) ... vs Baño. (**c**) ... vs Dolor. (**d**) ... vs Gracias. (**e**) ... vs Hambre. (**f**) ... vs No. (**g**) ... vs Sed. (**h**) ... vs Sí. (**i**) Ayuda vs Baño. (**j**) Ayuda vs Dolor. (**k**) Ayuda vs Gracias. (**l**) Ayuda vs Hambre. (**m**) Ayuda vs No. (**n**) Ayuda vs Sed. (**o**) Ayuda vs Sí. (**p**) Ayuda vs Dolor. (**q**) Ayuda vs Gracias. (**r**) Baño vs Hambre.

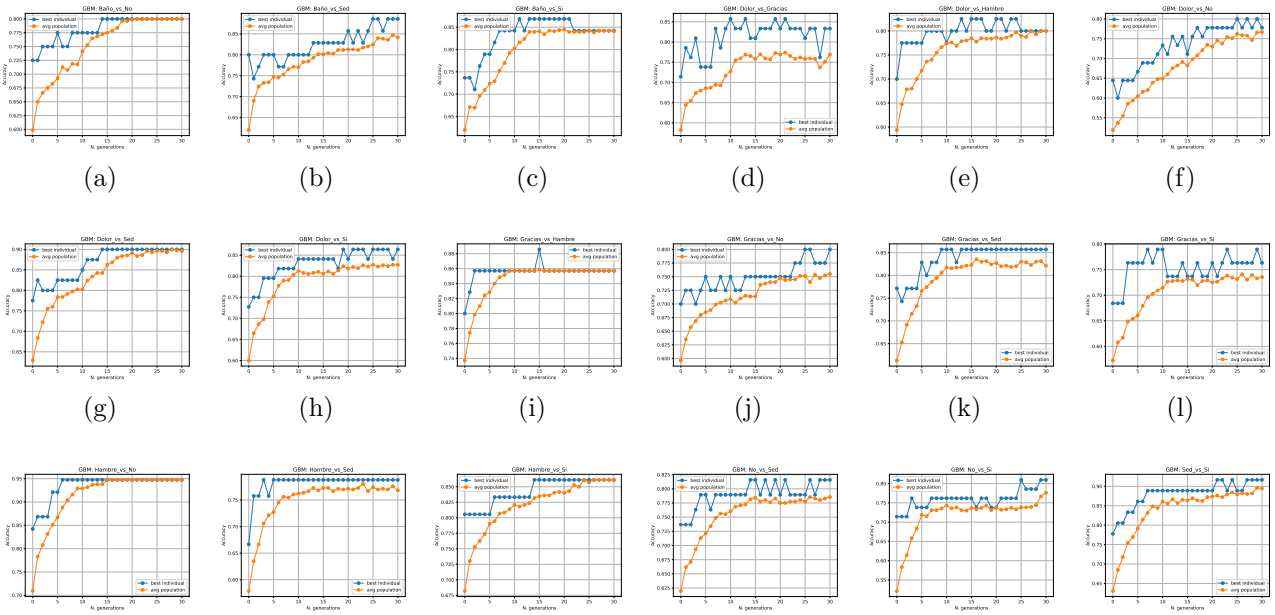


Figure 7.50: GAs performance for the one-to-one subsets using GBM classifier.

(**a**) Baño vs No. (**b**) Baño vs Sed. (**c**) Baño vs Sí. (**d**) Dolor vs Gracias. (**e**) Dolor vs Hambre. (**f**) Dolor vs No. (**g**) Dolor vs Sed. (**h**) Dolor vs Sí. (**i**) Gracias vs Hambre. (**j**) Gracias vs No. (**k**) Gracias vs Sed. (**l**) Gracias vs Sí. (**m**) Hambre vs No. (**n**) Hambre vs Sed. (**o**) Hambre vs Sí. (**p**) No vs Sed. (**q**) No vs Sí. (**r**) Sed vs Sí.

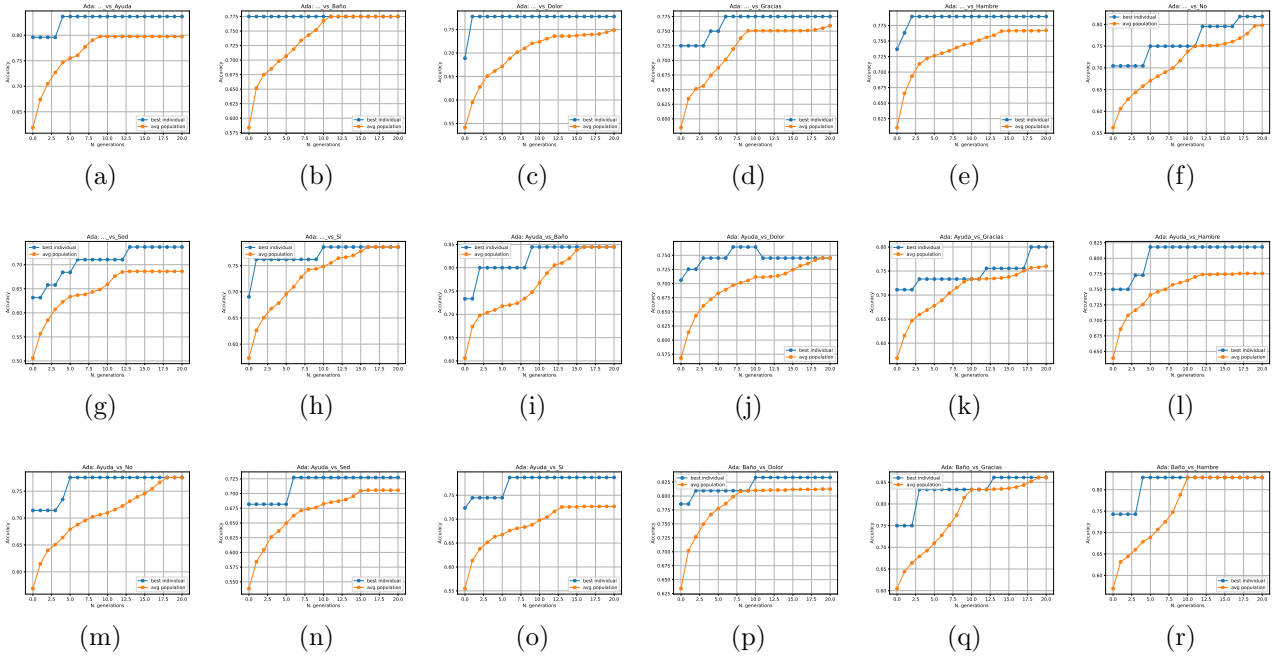
Table 7.6 shows the highest accuracy value found by the GA for each of the 36 data subsets in the one-to-one classification for the GBM algorithm.

	...	Sí	No	Baño	Hambre	Sed	Ayuda	Dolor	Gracias
...									
Sí	0.82								
No	0.85	0.84							
Baño	0.78	0.80	0.95						
Hambre	0.85	0.80	0.89	0.86					
Sed	0.87	0.84	0.80	0.82	0.89				
Ayuda	0.80	0.82	0.80	0.80	0.80	0.95			
Dolor	0.76	0.80	0.89	0.90	0.86	0.79	0.82		
Gracias	0.90	0.81	0.87	0.86	0.79	0.86	0.81	0.92	

Table 7.6: One-to-one accuracies for GBM classifier.

7.2.6 One-to-one AdaBoost classifier

Figures 7.51 and 7.52 show the performance of the GA in conjunction with the AdaBoost classifier. Note that since it is a one-to-one classification, there will be 36 subsets of words. For this reason, there are 36 different graphs. It is also possible to notice that all the graphs start with a relatively low accuracy, but this accuracy increases as the generations of the genetic algorithm progress.



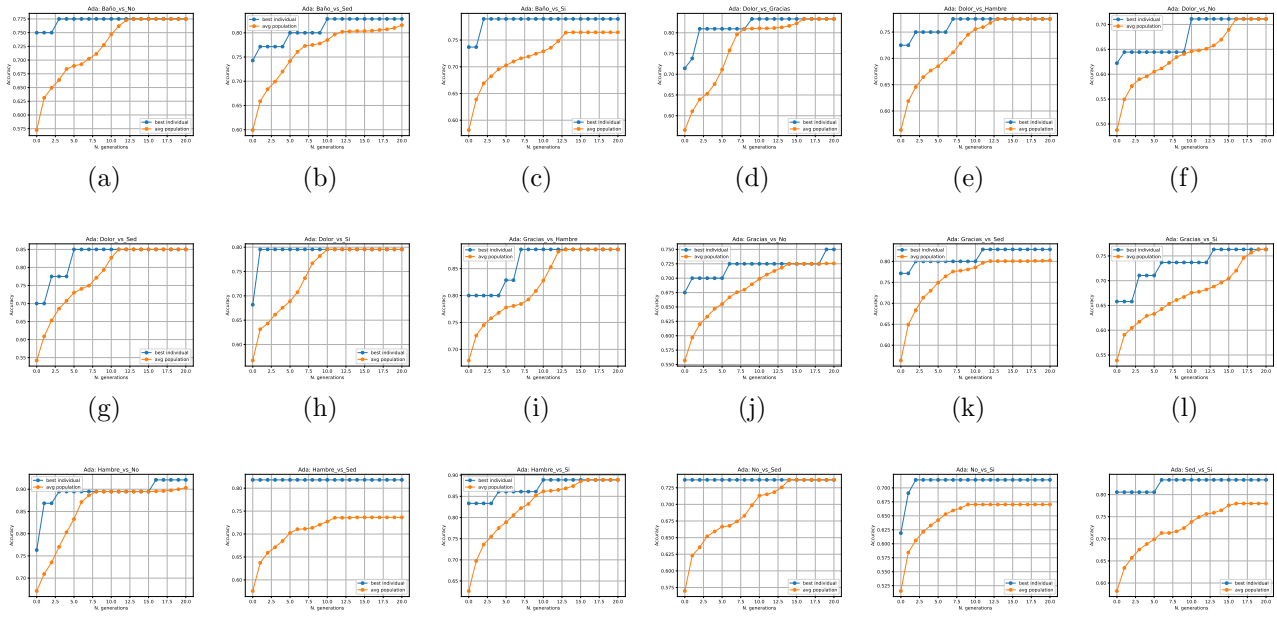
(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

(m) (n) (o) (p) (q) (r)

Figure 7.51: GAs performance for the one-to-one subsets using AdaBoost classifier.

(**a**) ... vs Ayuda. (**b**) ... vs Baño. (**c**) ... vs Dolor. (**d**) ... vs Gracias. (**e**) ... vs Hambre. (**f**) ... vs No. (**g**) ... vs Sed. (**h**) ... vs Sí. (**i**) Ayuda vs Baño. (**j**) Ayuda vs Dolor. (**k**) Ayuda vs Gracias. (**l**) Ayuda vs Hambre. (**m**) Ayuda vs No. (**n**) Ayuda vs Sed. (**o**) Ayuda vs Sí. (**p**) Ayuda vs Dolor. (**q**) Ayuda vs Gracias. (**r**) Baño vs Hambre.



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

(m) (n) (o) (p) (q) (r)

Figure 7.52: GAs performance for the one-to-one subsets using AdaBoost classifier.

(**a**) Baño vs No. (**b**) Baño vs Sed. (**c**) Baño vs Sí. (**d**) Dolor vs Gracias. (**e**) Dolor vs Hambre. (**f**) Dolor vs No. (**g**) Dolor vs Sed. (**h**) Dolor vs Sí. (**i**) Gracias vs Hambre. (**j**) Gracias vs No. (**k**) Gracias vs Sed. (**l**) Gracias vs Sí. (**m**) Hambre vs No. (**n**) Hambre vs Sed. (**o**) Hambre vs Sí. (**p**) No vs Sed. (**q**) No vs Sí. (**r**) Sed vs Sí.

Table 7.7 shows the highest accuracy value found by the GA for each of the 36 data subsets in the one-to-one classification for the AdaBoost algorithm.

	...	Sí	No	Baño	Hambre	Sed	Ayuda	Dolor	Gracias
...									
Sí	0.84								
No	0.78	0.84							
Baño	0.78	0.76	0.83						
Hambre	0.78	0.80	0.86	0.83					
Sed	0.79	0.82	0.83	0.78	0.89				
Ayuda	0.82	0.78	0.78	0.71	0.75	0.92			
Dolor	0.74	0.73	0.83	0.85	0.83	0.82	0.74		
Gracias	0.79	0.79	0.79	0.80	0.76	0.89	0.71	0.83	

Table 7.7: One-to-one accuracies for Ada classifier.

7.2.7 One-to-one NB classifier

Figures 7.53 and 7.54 show the performance of the GA in conjunction with the Naïve Bayes classifier. Note that since it is a one-to-one classification, there will be 36 subsets of words. For this reason, there are 36 different graphs. It is also possible to notice that all the graphs start with a relatively low accuracy, but this accuracy increases as the generations of the genetic algorithm progress.

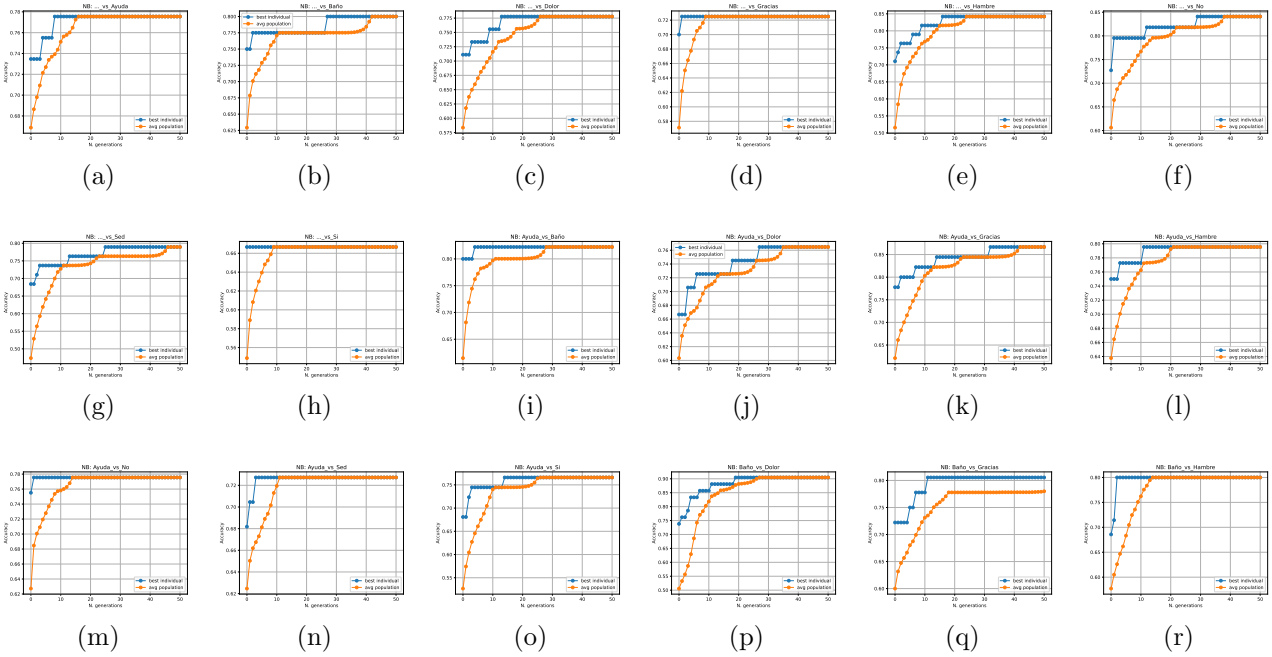


Figure 7.53: GAs performance for the one-to-one subsets using NB classifier.

(**a**) ... vs Ayuda. (**b**) ... vs Baño. (**c**) ... vs Dolor. (**d**) ... vs Gracias. (**e**) ... vs Hambre. (**f**) ... vs No. (**g**) ... vs Sed. (**h**) ... vs Sí. (**i**) Ayuda vs Baño. (**j**) Ayuda vs Dolor. (**k**) Ayuda vs Gracias. (**l**) Ayuda vs Hambre. (**m**) Ayuda vs No. (**n**) Ayuda vs Sed. (**o**) Ayuda vs Sí. (**p**) Ayuda vs Dolor. (**q**) Ayuda vs Gracias. (**r**) Baño vs Hambre.

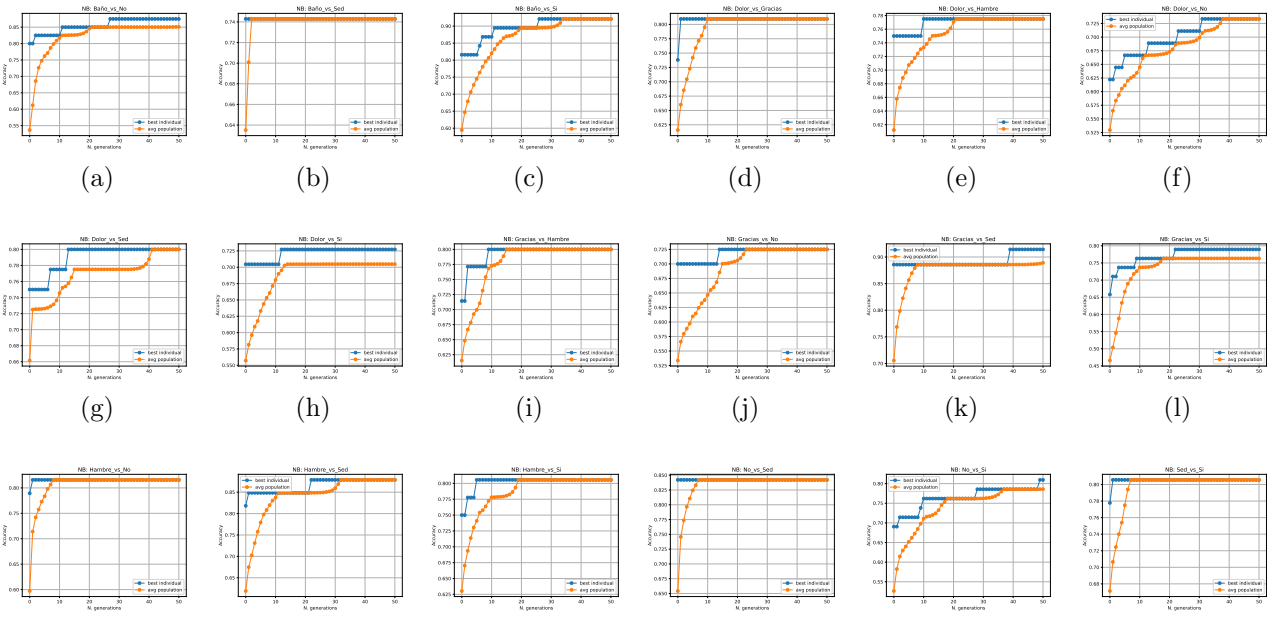


Figure 7.54: GAs performance for the one-to-one subsets using NB classifier.

(**a**) Baño vs No. (**b**) Baño vs Sed. (**c**) Baño vs Sí. (**d**) Dolor vs Gracias. (**e**) Dolor vs Hambre. (**f**) Dolor vs No. (**g**) Dolor vs Sed. (**h**) Dolor vs Sí. (**i**) Gracias vs Hambre. (**j**) Gracias vs No. (**k**) Gracias vs Sed. (**l**) Gracias vs Sí. (**m**) Hambre vs No. (**n**) Hambre vs Sed. (**o**) Hambre vs Sí. (**p**) No vs Sed. (**q**) No vs Sí. (**r**) Sed vs Sí.

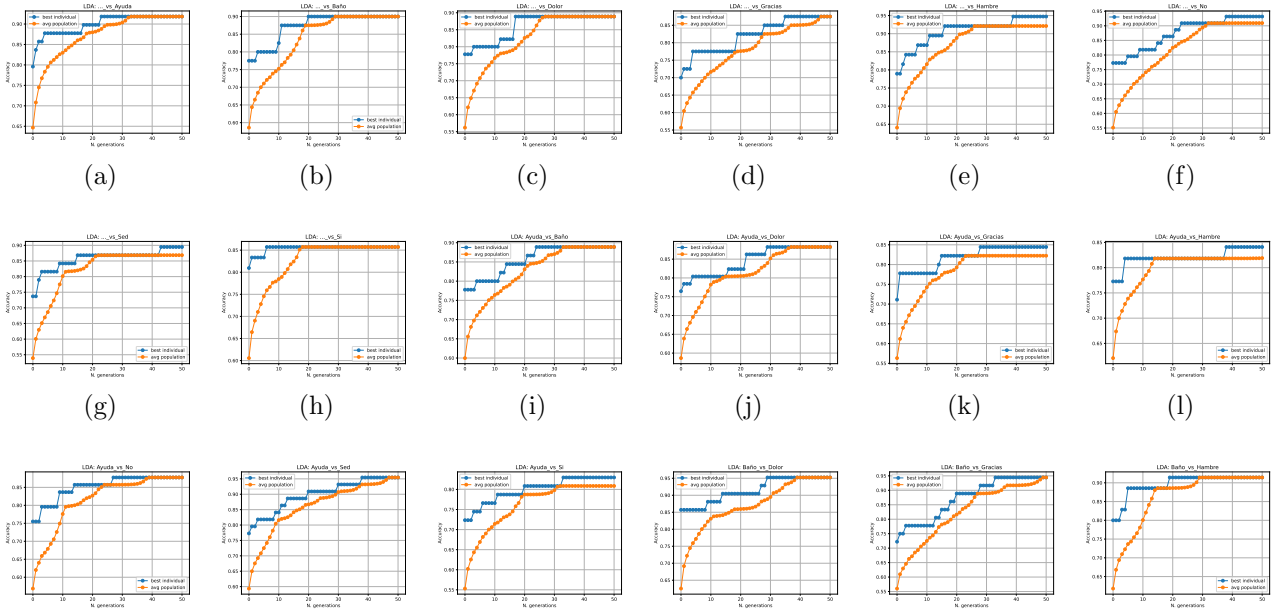
Table 7.8 shows the highest accuracy value found by the GA for each of the 36 data subsets in the one-to-one classification for the Naïve Bayes algorithm.

	...	Sí	No	Baño	Hambre	Sed	Ayuda	Dolor	Gracias
...									
Sí	0.78								
No	0.80	0.82							
Baño	0.78	0.76	0.90						
Hambre	0.72	0.87	0.81	0.81					
Sed	0.84	0.80	0.80	0.78	0.80				
Ayuda	0.84	0.78	0.88	0.73	0.72	0.82			
Dolor	0.79	0.73	0.74	0.80	0.91	0.88	0.84		
Gracias	0.67	0.77	0.92	0.73	0.79	0.81	0.81	0.81	

Table 7.8: One-to-one accuracies for NB classifier.

7.2.8 One-to-one LDA classifier

Figures 7.55 and 7.56 show the performance of the GA in conjunction with the LDA classifier. Note that since it is a one-to-one classification, there will be 36 subsets of words. For this reason, there are 36 different graphs. It is also possible to notice that all the graphs start with a relatively low accuracy, but this accuracy increases as the generations of the genetic algorithm progress.



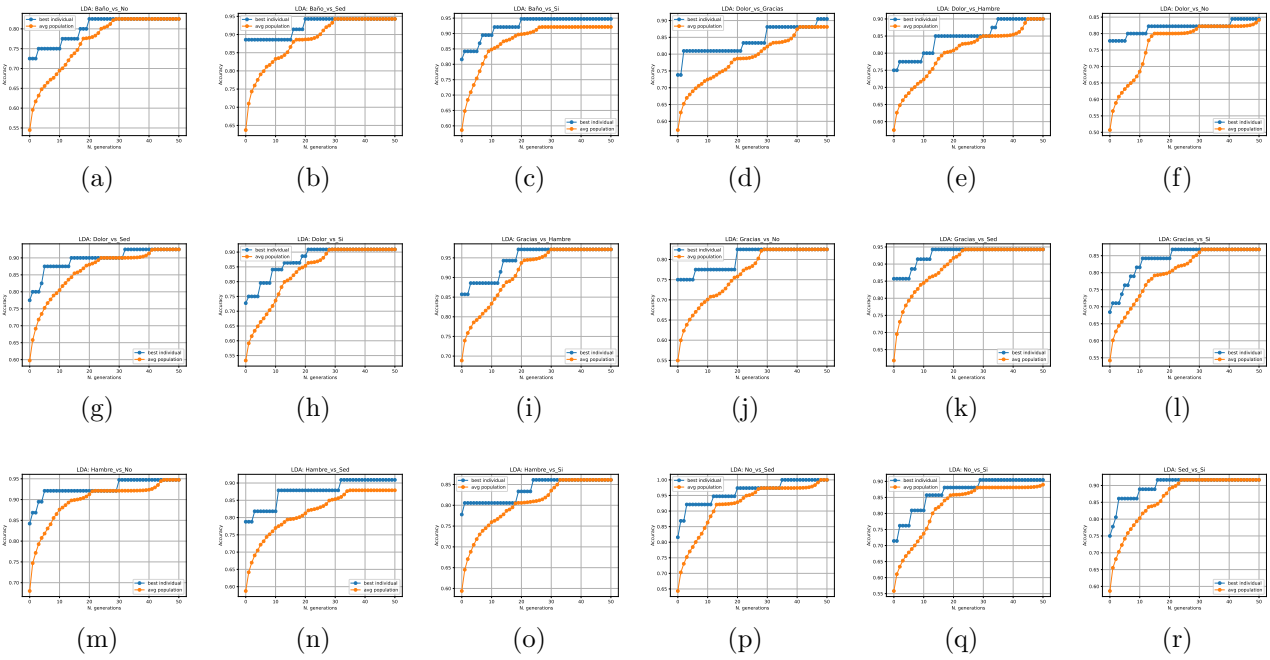
(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

(m) (n) (o) (p) (q) (r)

Figure 7.55: GAs performance for the one-to-one subsets using LDA classifier.

(**a**) ... vs Ayuda. (**b**) ... vs Baño. (**c**) ... vs Dolor. (**d**) ... vs Gracias. (**e**) ... vs Hambre. (**f**) ... vs No. (**g**) ... vs Sed. (**h**) ... vs Sí. (**i**) Ayuda vs Baño. (**j**) Ayuda vs Dolor. (**k**) Ayuda vs Gracias. (**l**) Ayuda vs Hambre. (**m**) Ayuda vs No. (**n**) Ayuda vs Sed. (**o**) Ayuda vs Sí. (**p**) Ayuda vs Dolor. (**q**) Ayuda vs Gracias. (**r**) Baño vs Hambre.



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

(m) (n) (o) (p) (q) (r)

Figure 7.56: GAs performance for the one-to-one subsets using LDA classifier.

(**a**) Baño vs No. (**b**) Baño vs Sed. (**c**) Baño vs Sí. (**d**) Dolor vs Gracias. (**e**) Dolor vs Hambre. (**f**) Dolor vs No. (**g**) Dolor vs Sed. (**h**) Dolor vs Sí. (**i**) Gracias vs Hambre. (**j**) Gracias vs No. (**k**) Gracias vs Sed. (**l**) Gracias vs Sí. (**m**) Hambre vs No. (**n**) Hambre vs Sed. (**o**) Hambre vs Sí. (**p**) No vs Sed. (**q**) No vs Sí. (**r**) Sed vs Sí.

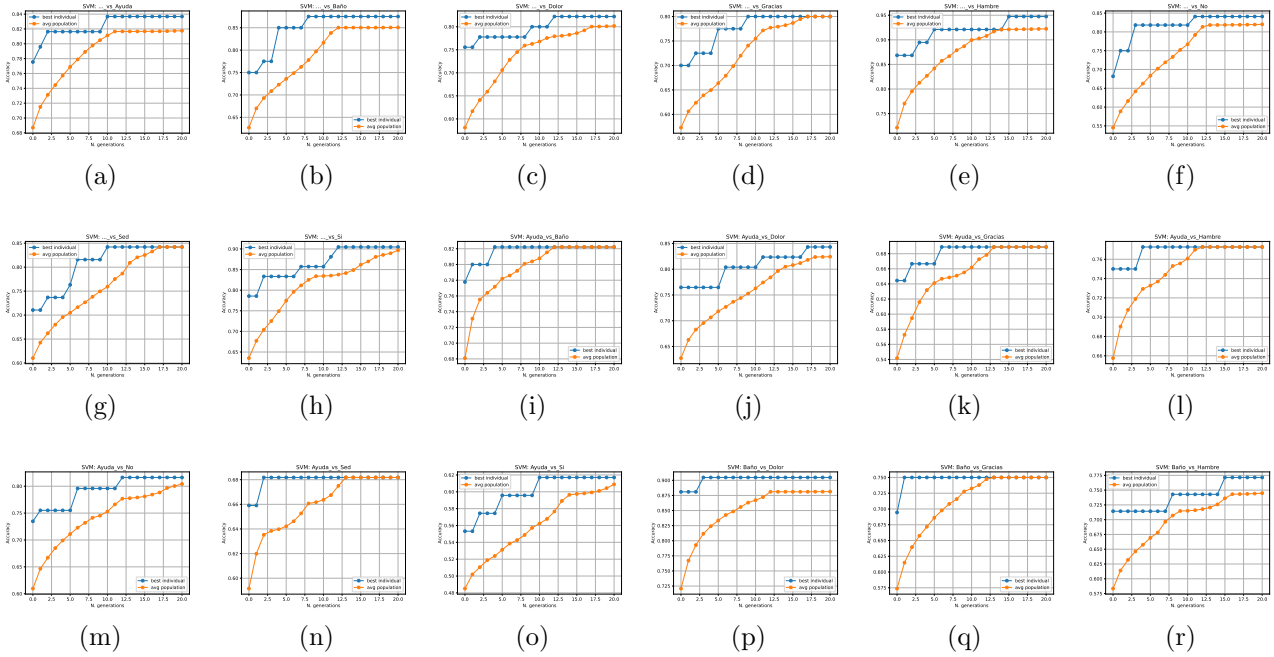
Table 7.9 shows the highest accuracy value found by the GA for each of the 36 data subsets in the one-to-one classification for the LDA algorithm.

	...	Sí	No	Baño	Hambre	Sed	Ayuda	Dolor	Gracias
...									
Sí	0.92								
No	0.90	0.89							
Baño	0.89	0.88	0.95						
Hambre	0.88	0.84	0.94	0.90					
Sed	0.95	0.84	0.91	0.90	0.97				
Ayuda	0.93	0.88	0.82	0.84	0.82	0.95			
Dolor	0.89	0.95	0.94	0.92	0.94	0.91	1.0		
Gracias	0.86	0.83	0.95	0.91	0.87	0.86	0.90	0.92	

Table 7.9: One-to-one accuracies for LDA classifier.

7.2.9 One-to-one SVM classifier

Figures 7.57 and 7.58 show the performance of the GA in conjunction with the SVM classifier. Note that since it is a one-to-one classification, there will be 36 subsets of words. For this reason, there are 36 different graphs. It is also possible to notice that all the graphs start with a relatively low accuracy, but this accuracy increases as the generations of the genetic algorithm progress.



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

(m) (n) (o) (p) (q) (r)

Figure 7.57: GAs performance for the one-to-one subsets using SVM classifier.

(**a**) ... vs Ayuda. (**b**) ... vs Baño. (**c**) ... vs Dolor. (**d**) ... vs Gracias. (**e**) ... vs Hambre. (**f**) ... vs No. (**g**) ... vs Sed. (**h**) ... vs Sí. (**i**) Ayuda vs Baño. (**j**) Ayuda vs Dolor. (**k**) Ayuda vs Gracias. (**l**) Ayuda vs Hambre. (**m**) Ayuda vs No. (**n**) Ayuda vs Sed. (**o**) Ayuda vs Sí. (**p**) Ayuda vs Dolor. (**q**) Ayuda vs Gracias. (**r**) Baño vs Hambre.

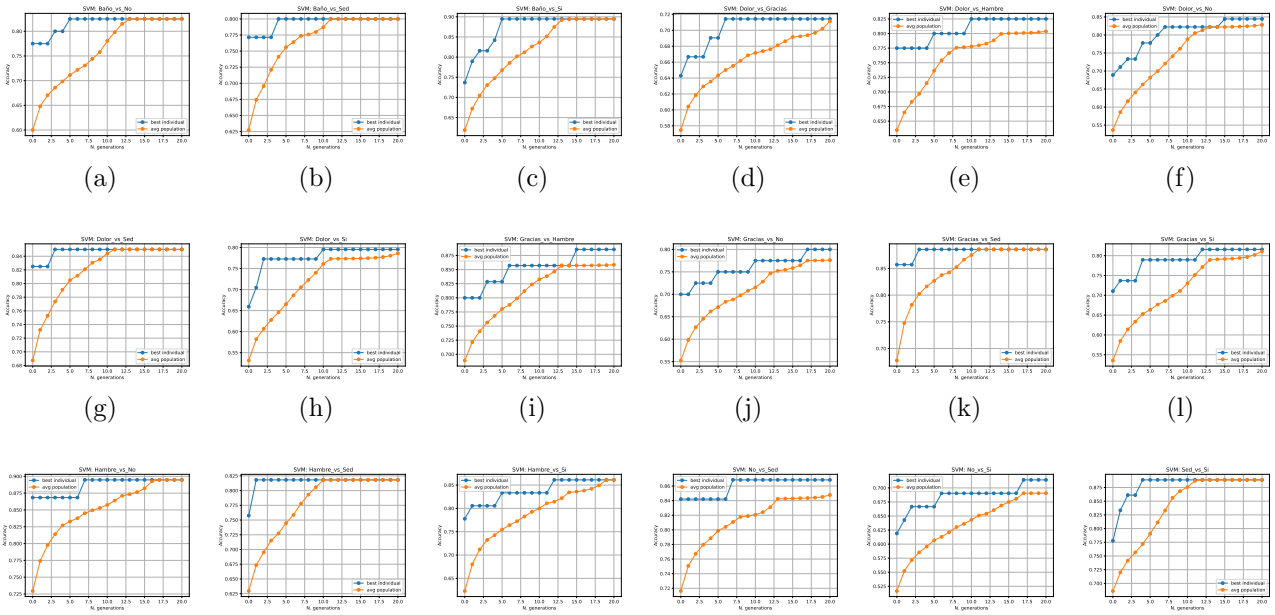


Figure 7.58: GAs performance for the one-to-one subsets using SVM classifier.

(**a**) Baño vs No. (**b**) Baño vs Sed. (**c**) Baño vs Sí. (**d**) Dolor vs Gracias. (**e**) Dolor vs Hambre. (**f**) Dolor vs No. (**g**) Dolor vs Sed. (**h**) Dolor vs Sí. (**i**) Gracias vs Hambre. (**j**) Gracias vs No. (**k**) Gracias vs Sed. (**l**) Gracias vs Sí. (**m**) Hambre vs No. (**n**) Hambre vs Sed. (**o**) Hambre vs Sí. (**p**) No vs Sed. (**q**) No vs Sí. (**r**) Sed vs Sí.

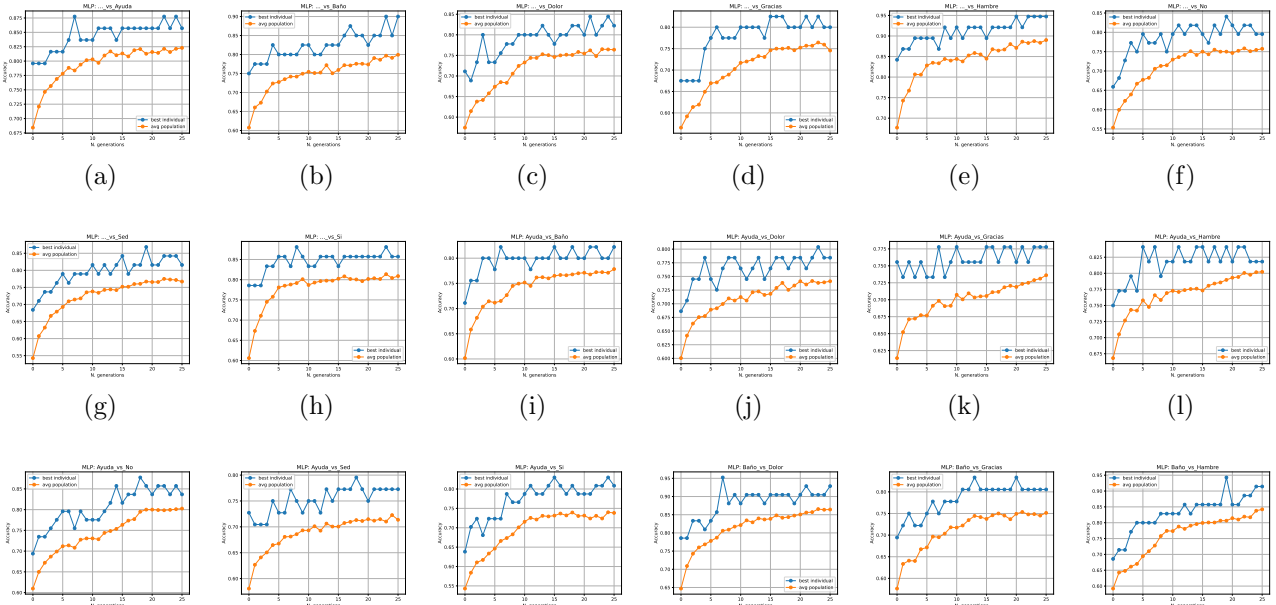
Table 7.10 shows the highest accuracy value found by the GA for each of the 36 data subsets in the one-to-one classification for the SVM algorithm.

	...	Sí	No	Baño	Hambre	Sed	Ayuda	Dolor	Gracias
...									
Sí	0.84								
No	0.88	0.82							
Baño	0.82	0.84	0.90						
Hambre	0.80	0.69	0.75	0.71					
Sed	0.95	0.77	0.77	0.82	0.89				
Ayuda	0.84	0.82	0.82	0.84	0.80	0.89			
Dolor	0.84	0.68	0.80	0.85	0.89	0.82	0.87		
Gracias	0.90	0.62	0.89	0.80	0.82	0.86	0.71	0.89	

Table 7.10: One-to-one accuracies for SVM classifier.

7.2.10 One-to-one MLP classifier

Figures 7.59 and 7.60 show the performance of the GA in conjunction with the MLP classifier. Note that since it is a one-to-one classification, there will be 36 subsets of words. For this reason, there are 36 different graphs. It is also possible to notice that all the graphs start with a relatively low accuracy, but this accuracy increases as the generations of the genetic algorithm progress.



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

(m) (n) (o) (p) (q) (r)

Figure 7.59: GAs performance for the one-to-one subsets using MLP classifier.

(**a**) ... vs Ayuda. (**b**) ... vs Baño. (**c**) ... vs Dolor. (**d**) ... vs Gracias. (**e**) ... vs Hambre. (**f**) ... vs No. (**g**) ... vs Sed. (**h**) ... vs Sí. (**i**) Ayuda vs Baño. (**j**) Ayuda vs Dolor. (**k**) Ayuda vs Gracias. (**l**) Ayuda vs Hambre. (**m**) Ayuda vs No. (**n**) Ayuda vs Sed. (**o**) Ayuda vs Sí. (**p**) Ayuda vs Dolor. (**q**) Ayuda vs Gracias. (**r**) Baño vs Hambre.

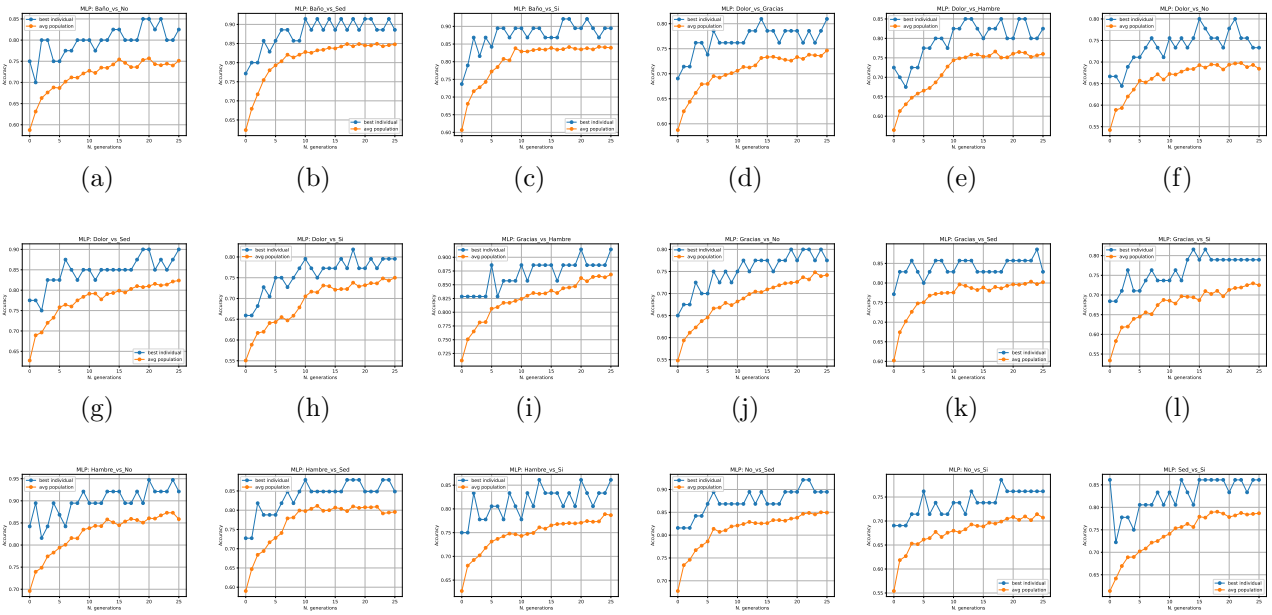


Figure 7.60: GAs performance for the one-to-one subsets using MLP classifier.

(**a**) Baño vs No. (**b**) Baño vs Sed. (**c**) Baño vs Sí. (**d**) Dolor vs Gracias. (**e**) Dolor vs Hambre. (**f**) Dolor vs No. (**g**) Dolor vs Sed. (**h**) Dolor vs Sí. (**i**) Gracias vs Hambre. (**j**) Gracias vs No. (**k**) Gracias vs Sed. (**l**) Gracias vs Sí. (**m**) Hambre vs No. (**n**) Hambre vs Sed. (**o**) Hambre vs Sí. (**p**) No vs Sed. (**q**) No vs Sí. (**r**) Sed vs Sí.

Table 7.11 shows the highest accuracy value found by the GA for each of the 36 data subsets in the one-to-one classification for the MLP algorithm.

	...	Sí	No	Baño	Hambre	Sed	Ayuda	Dolor	Gracias
...									
Sí	0.88								
No	0.90	0.82							
Baño	0.84	0.80	0.95						
Hambre	0.82	0.78	0.83	0.81					
Sed	0.95	0.84	0.94	0.85	0.91				
Ayuda	0.84	0.88	0.85	0.80	0.80	0.95			
Dolor	0.87	0.80	0.91	0.90	0.89	0.88	0.92		
Gracias	0.88	0.83	0.92	0.82	0.82	0.86	0.79	0.86	

Table 7.11: One-to-one accuracies for MLP classifier.

7.3 Evaluation of Average Accuracy for One-to-One Classification

In addition, Table 7.12 shows in summary the average of each of the previous one-to-one accuracy tables. Note that this table has its rows sorted by accuracy in decreasing order, from this we can see that kNN model was the best, while NB model was the one that obtained lower accuracy results.

Model	Accuracy
kNN	0.9194
LDA	0.9014
LogReg	0.8986
DT	0.8958
RF	0.8864
MLP	0.8608
GBM	0.8408
SVM	0.8194
Ada	0.8019
NB	0.8011

Table 7.12: Overall accuracy for one-to-one classification.

7.4 Processing Time for GA-Optimized Feature Sets

In the Figure 7.61 is presented a box plot indicating the elapsed time for each instance for each of ten clasfiers.

The Figure 7.61 shows a box plot indicating the elapsed time to generate a feature map for each of the instances based on the combination of electrodes and features produced by the GA for each of the ten classifier models.

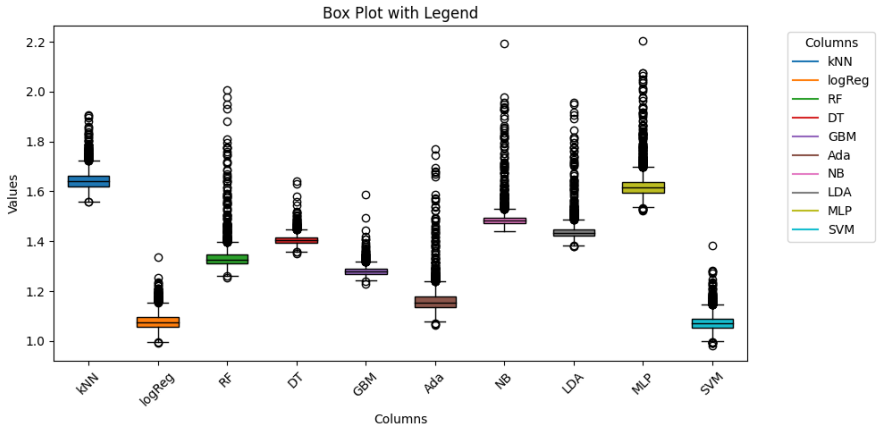


Figure 7.61: Box plot for the fature extraction time for each classifier.

8 Discussions

The proposed method exhibits strong efficacy in recognising and differentiating the basic Spanish words /Sí/ (Yes) and /No/ (No). This research work demonstrated superior accuracy compared to existing methods in the current state-of-the-art literature through rigorous testing. The proposed method efficiently addresses fundamental terms and surpasses prior techniques regarding precision. This accomplishment demonstrates the method's strength and effectiveness, underscoring its possible uses necessitating precise recognition of fundamental responses in Spanish, thus representing a substantial contribution to the field. In forthcoming research, we aim to utilise the complete EEGIS dataset, which includes complex vocabulary, and evaluate additional ML and DL classifier models that exceed the performance of this study.

The proposed method effectively addresses core words and demonstrates superior accuracy compared to prior techniques; however, it is crucial to recognize the limitations of this study.

First, the limited vocabulary size constitutes a considerable constraint. The proposed model exhibits positive outcomes within the designated vocabulary, nevertheless, it has not yet expanded to capture the complete intricacies of spoken language. Improving vocabulary requires further data acquisition and model enhancement, which we consider a crucial direction for forthcoming research.

in second place, the duration of inference time poses a barrier to real-time application. The processing procedure, encompassing feature extraction and classification, requires around several second each instances, even when executed on research computer. Thi situation introduces a latency period that can affect the smoothness of real-time interaction. Enhancing the algorithm's computational efficiency through hardware acceleration (i.e. Graphics Processing Units (GPUs) or specialised edge computing devices) and investigation lightweight model architectures may alleviate this issue in subsequent versions.

In addition, the headset configuration imposes practical limitations. Signal quality must quality may be compromised as a result of thee tendency of saline solution-based electrodes to dissipate over time. Furthermore, skin irritation at the contact points may result from prologued use. In order to resolve these concerns, it may be necessary to investigate alternative electrode materials, optimise headset ergonomics, o incorporate automated hydration mechanisms to guarantee signal stability over extended periods.

Notwithstanding these limitations, this work lays the basis for forthcoming research. The optimisation of computational efficiency, signal acquisition robustness, and user convenience would be more substantial when this method is implemented in a real-time application. As we proceed towards practical implementations in a real-world settings,

these challenges will be the primary focus of a future work.

In addition, due to the high dimensionality and noise of the axis signals, the classification of these in imagined speech presents various difficulties. In this work, a set of machine learning models were investigated in conjunction with feature extraction combined with genetic algorithms to maximize the classification performance of the models. In this work, the goal is to find not only the most efficient classifiers but also the combination of electrodes and features that best differentiates imagined words.

Within the results we found that the KNN algorithm outperformed all other models with an accuracy of 91.94%. This performance implies that the structure of the selected features and the electrode configuration was optimal to achieve such accuracy. This is also because, as mentioned above, this model was the only one that did not implement the default parameters, but also within the genetic algorithm a heuristic was implemented to search for the value of k.

On the contrary, the Naïve Bayes model obtained the worst performance of the ten classifiers with 80.11%, this probably due to its strong independence assumptions that are usually unrealistic in the context of EEG data where correlations between channels prevail. Especially with non-independent features, this underlines the need for model selection in the classification of neurophysiological data.

Now, the GA's behavior in various models produces another significant finding: some attitude curves show a smooth progression, while others show a fluctuating progression. The evolutionary optimization process showed different speeds of convergence; these variations are due to the stochastic and deterministic nature of some models. These oscillations indicate the interaction between model stability and the GA's exploratory capacity. From this perspective, the GA not only maximized feature selection but also acted as an indirect tool to assess the consistency and robustness of the models.

Ultimately, the findings demonstrate the promise of using EEG signals for imagined speech decoding by combining evolutionary feature selection with classification models. They also highlight the interaction between the structure of such a neurological task, data complexity, and classifier characteristics. Particularly in clinical or assistive technology settings, these results may guide future advances in BCI systems designed for imagined speech communication.

9 Conclusions

The results obtained in this research project for the classification of imagined speech exceed those reported in the state of the art.

A clear distinction will be observed between the classification of simple and complex words; however, the accuracy for the classification of simple words obtained in this work is higher than that reported in previous research.

Furthermore, the use of a genetic algorithm allowed the identification of the optimal electrodes and the best combination of features, which maximized the accuracy of the models used in this work.

In conclusion, this study presented an effective approach for classifying EEG signals associated with "Imagined Speech," or the mental articulation of specific words. By targeting the electrical activity of the brain as captured through EEG, we sought to overcome the inherent challenges of this task, such as the subtle nature of imagined word patterns, signal variability across individuals, and low signal-to-noise ratio. Our proposed method introduces a novel feature extraction strategy that substantially enhances classification accuracy.

At the heart of this approach is a genetic algorithm that optimizes the selection of features from the EEG signals, exploring a broad feature space to identify the most predictive attributes. The genetic algorithm's capacity for balancing extensive exploration with precise feature refinement made it particularly suited for the complexities of EEG signal classification, allowing us to achieve results that outperform current state-of-the-art methods.

This work's findings underscore the potential of genetic algorithms in advancing EEG-based imagined speech recognition, making it a valuable step forward for brain-computer interfaces and, more broadly, for applications in neural communication and assistive technology.

Through this method, we demonstrate an effective pathway for reliable imagined speech classification, laying the groundwork for future research and practical applications in the domain of non-invasive neurotechnology.

10 Appendix

10.1 Patient consent statement

Carta de consentimiento informado para pacientes

Modelo de inteligencia artificial para decodificar el habla imaginada

Marco teórico

La electroencefalografía (EEG) es una técnica utilizada para medir la actividad eléctrica del cerebro. Se basa en la colocación de electrodos en el cuero cabelludo para registrar los cambios en el potencial eléctrico generado por las células cerebrales. Estos cambios reflejan la actividad neuronal y pueden proporcionar información sobre diferentes estados mentales, como el sueño, la vigilia, la atención y la epilepsia, entre otros. La EEG es ampliamente utilizada en medicina clínica y en investigaciones neurológicas debido a su no invasividad y su capacidad para capturar la actividad cerebral en tiempo real.

Objetivo

El objetivo principal de esta investigación es analizar y evaluar la efectividad de los modelos de IA en la clasificación de señales de EEG asociadas al pensamiento de un conjunto específico de palabras, es decir, un algoritmo que analiza señales de EEG que indique en qué palabra (de un conjunto específico) está pensando un individuo.

Supervisor y responsable del proyecto

El supervisor del proyecto es el Dr. András Takács, Profesor de la Universidad Autónoma de Querétaro, Facultad de Ingeniería del Centro Universitario, edificio Biotecnológico, tercer piso.
El responsable del proyecto es el Ing. Fis. Edgar Lara Arellano, estudiante de la Maestría en Ciencias en Inteligencia Artificial, en la Universidad Autónoma de Querétaro, Facultad de Ingeniería del Campus Aeropuerto, edificio de Usos Múltiples, primer piso, cubículo 1 de posgrados.

Descripción del proyecto de investigación

El participante deberá acudir:

- Desayunado
- Con el cabello limpio y seco, lavado solo con shampoo, sin cremas, enjuagues, gel o spray
- Sin desvelo
- Sin tos, gripe, fiebre o malestar estomacal

Durante la sesión de EEG, se colocarán electrodos en su cuero cabelludo para registrar la actividad eléctrica de su cerebro. Estos electrodos están conectados a una máquina que registra las señales cerebrales mientras usted se encuentra en reposo.

Durante la sesión, se le pedirá que se relaje y mantenga una posición cómoda y tranquila. El objetivo de esta parte del estudio es registrar la actividad cerebral en un estado de relajación para establecer una línea base.
Posteriormente, se le mostrarán palabras en una pantalla y se le pedirá que piense en estas palabras. No es necesario que diga las palabras en voz alta, solo debe concentrarse en ellas y pensar en su significado. Esto nos ayudará a estudiar cómo su cerebro procesa y reacciona a diferentes estímulos visuales y cognitivos.

Fase de evaluación inicial

Nombre		Apellido paterno		Apellido materno	
Fecha nacimiento		Edad		Ocupación	
Problemas de sueño	Si ☐ No ☐	Especifique:		Convulsiones	Si ☐ No ☐
Tabaco	Si ☐ No ☐	Alcohol	Si ☐ No ☐		
Drogas	Si ☐ No ☐	Número de comidas al día			
Horas de sueño					
Medicamentos y dosis	Si ☐ No ☐	Especifique:			

Fase de pruebas clínicas y neurofisiológicas
Se realizarán maniobras de activación (apertura y cierre palpebral, estimulación luminosa intermitente, hiperventilación pulmonar). Estas pruebas se realizan para la evaluación de su actividad eléctrica cerebral.

Costos
Este proyecto de investigación NO implica ningún costo para los participantes activos.

Beneficios
Su participación en este proyecto de investigación, brindará beneficios tanto a nivel individual como a nivel científico, un impacto significativo en el avance del conocimiento científico y en el desarrollo de herramientas tecnológicas innovadoras.

Confidencialidad
Los datos personales de identificación son confidenciales y serán encriptados, se le asignará un código que sólo conocerá el investigador responsable del proyecto. Los datos obtenidos como resultado de esta investigación podrían ser utilizados para publicaciones científicas, protegiendo su identidad.

El participante debe ir acompañado de un mayor de 18 años a la sesión para que de esta manera usted pueda participar en la investigación.

El protocolo del que deriva este consentimiento informado, así como el propio consentimiento ha sido evaluado por el comité de Bioética de la Dirección de Investigación y Posgrado de la Universidad Autónoma de Queŕetaro (DIP-UAQ).

Riesgo o molestia
Durante la adquisición de las señales electroencefalográficas, usted puede experimentar molestias leves asociadas con la colocación de electrodos sobre el cuero cabelludo. Se utilizarán electrodos de superficie de acuerdo a los procedimientos estándar para la realización de este estudio.

Eventos adversos
La realización de un EEG en condiciones normales en pacientes sanos se considera un procedimiento seguro y no conlleva riesgos significativos para la salud durante o después del mismo.

Por lo tanto, usted puede tener la tranquilidad de que el procedimiento de EEG realizado bajo condiciones normales no representa un riesgo significativo para su salud física o mental

Participación voluntaria
La participación en este estudio es completamente voluntaria y usted tiene el derecho de retirarse en cualquier momento sin consecuencias negativas.

Es importante destacar que, en virtud de la naturaleza de la investigación y para garantizar la integridad y validez de los resultados, los datos recopilados se conservarán independientemente de la decisión del participante de retirarse del estudio. Esto significa que, aunque un participante decida no continuar con su participación, los datos recopilados hasta ese momento seguirán formando parte del conjunto de datos analizados en el estudio.

Consentimiento informado
YO _____,
HE LEÍDO Y COMPRENDIDO LA INFORMACIÓN PROPORCIONADA SOBRE EL ESTUDIO TITULADO "MODELO DE INTELIGENCIA ARTIFICIAL PARA DECODIFICAR EL HABLA IMAGINADA".

He tenido la oportunidad de discutir cualquier pregunta o preocupación que pueda tener con el equipo de investigación, y todas mis preguntas han sido respondidas satisfactoriamente.

A continuación, DOY MI CONSENTIMIENTO VOLUNTARIO para participar en este estudio y estoy de acuerdo con las siguientes declaraciones:

1. Entiendo que mi participación en este estudio es completamente voluntaria y que tengo el derecho de retirarme en cualquier momento sin consecuencias negativas.
2. Estoy al tanto de los objetivos y los procedimientos del estudio, así como de los posibles riesgos y molestias asociados con mi participación.
3. Entiendo que los datos recopilados durante el estudio se conservarán independientemente de mi decisión de retirarme, y que se mantendrá la confidencialidad de mi información personal.
4. Acepto que mi información pueda ser utilizada para análisis y publicación científica, siempre y cuando se mantenga mi anonimato y confidencialidad.

ACEPTO VOLUNTARIAMENTE y firmo este consentimiento, indicando que estoy de acuerdo con participar en este estudio y que he sido informado adecuadamente sobre todos los aspectos relevantes del mismo.

Querétaro, Qro. _____ _____ _____
Día Mes Año

_____

Nombre y firma del participante

_____ _____

Nombre y firma del testigo

Nombre y firma del investigador responsable del proyecto

10.2 Social Compensation Letters

FACULTAD DE INGENIERÍA
DIVISIÓN DE INVESTIGACIÓN Y POSGRADO

Constancia de actividades de retribución social

Santiago de Querétaro a 10 de enero de 2025

Dra. Liza Elena Aceves López

Coordinadora de Programas para la Formación y Consolidación de la Comunidad

P r e s e n t e.

En cumplimiento a los compromisos establecidos en el numeral 8 "*LOS DERECHOS Y OBLIGACIONES DEL BECARIO, DE LA COORDINACIÓN ACADÉMICA DE PROGRAMA DEL POSGRADO POSTULANTE Y DEL CONAHCYT, CON MOTIVO DE LA ASIGNACIÓN DE LA BECA.*" de la **Convocatoria Becas Nacional (Tradicional) 2023 - 1**, el **C. Edgar Lara Arellano** con número de **CVU 1270170** beneficiado con una beca para obtener el grado de **Maestría** en el programa **005351 – Maestría en Ciencias en Inteligencia Artificial** que se imparte en la **Universidad Autónoma de Querétaro**, realizó las actividades de retribución social que se enlistan en el documento anexo a este documento.

Las actividades de retribución social se realizaron durante el periodo de tiempo que el becario fue alumno regular de esta institución.

Asimismo, hago constar que, conforme a lo establecido en la Ley General de Archivos, la coordinación del posgrado organiza y conserva evidencia documental de dichas actividades en caso de que el Conahcyt o cualquier otra instancia la requiera.

Sin más por el momento, le envío un cordial saludo.

Dr. Saúl Tovar Arriaga

Coordinador del programa Maestría en Ciencias en Inteligencia Artificial

Constancia de actividades de retribución social

Actividad 1. Elaboración de una plantilla de LaTex en Overleaf para la escritura de tesis de maestría de la Facultad de Ingeniería, Universidad Autónoma de Querétaro.

Descripción de la actividad: Desarrollo de una plantilla para la escritura de tesis de maestría de la Facultad de Ingeniería, Universidad Autónoma de Querétaro. Basada en los requerimientos planteados en: https://dip.uaq.mx/docs/posgrado/Guia-tesis-CNyE.pdf La cual se desarrolló utilizando el sistema tipográfico LaTeX Esta plantilla se encuentra en:

- Repositorio público de Zenodo. El cual ya cuenta con un DOI para su citación.https://zenodo.org/records/13315535
- Página oficial de Overleaf. Donde los estudiantes de posgrado de la Facultad de Ingeniería pueden tomar esta plantilla y editarla usando el lenguaje de LaTex directamente en la plataforma de Overleaf. https://www.overleaf.com/latex/templates/uaq-fi-ia-tesis-template/gxttgptdhjtz

Fecha de inicio: 03/06/2024
Fecha de término: 15/08/2024

Institución en la que se realizó la actividad: Universidad Autónoma de Querétaro
Nombre del responsable de supervisar la actividad: Dr. Saul Tovar Arriaga
Datos de contacto del responsable de la actividad: saul.tovar@uaq.mx

Descripción del impacto social de la actividad: el desarrollo de una plantilla en esta plataforma empleando los estándares indicados por la facultad y que se encuentre ya directamente en la plataforma de Overleaf supone una ventaja competitiva y un avance al momento de formalizar la escritura de tesis de los alumnos de posgrado de la Facultad estandarizando y facilitando un formato que está a la mano de cualquier investigador.

Edgar Lara Arellano

1270170

Dr. Saúl Tovar Arriaga

Coordinador del programa Maestría en Ciencias en Inteligencia Artificial

Constancia de actividades de retribución social

Actividad 2. Desarrollo de un paquete de funciones en el lenguaje de programación Python para la extracción de características de series de tiempo y señales de electroencefalograma.

Descripción de la actividad: por medio de técnicas estadísticas, así como de procesamiento digital de señales, se desarrolló una librería en el lenguaje de programación Python, la cual permite la caracterización de series de tiempo, así como de señales de electroencefalograma por medio de la generación de vectores descriptores o bien de funciones independientes según el usuario lo requiera. Esta librería se encuentra pública en PyPi - Python Package Index en https://pypi.org/project/signalytica/

La cual puede ser descargada, instalada y utilizada por cualquier usuario que así lo deseé.

Fecha de inicio: 02/12/2023
Fecha de termino: 04/09/2024

Institución en la que se realizó la actividad: Universidad Autónoma de Querétaro
Nombre del responsable de supervisar la actividad: Dr. Saul Tovar Arriaga
Datos de contacto del responsable de la actividad: saul.tovar@uaq.mx

Descripción del impacto social de la actividad: Supone un reto el análisis de series de tiempo y en particular de señales de electroencefalograma, por esta razón, la herramienta desarrollada permite su utilización tanto para fines de aplicación industriales, académicos y de investigación. De manera que el desarrollo de esta contribuye en la generación e implementación de conocimiento científico.

Edgar Lara Arellano

1270170

Dr. Saúl Tovar Arriaga

Coordinador del programa Maestría en Ciencias en Inteligencia Artificial

10.3 Colloquium acceptance letter

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

FACULTAD DE INGENIERÍA

DIPFI POSGRADO INGENIERÍA

COLOQUIO 18° POSGRADO FI

Estimados Autores del Artículo con título

"Feature generation with genetic algorithms for imagined speech EEG signal classification "

enviado para su
participación en el XVIII Coloquio de Posgrado de la Facultad de Ingeniería este ha sido:

ACEPTADO

Para su participación en el Coloquio a desarrollarse del 19 al 23 de Noviembre del 2024 en la Ciudad de Querétaro, Querétaro

Le solicitamos
atentamente generar y enviar el archivo final del trabajo con toda la información requerida.

Noviembre de 2024
Facultad de ingeniería

Msc. Luis Angel Iturralde Carrera
PRESIDENTE DEL XVIII COLOQUIO DE POSGRADO
Facultad de Ingeniería

10.4 Scientific article published in an indexed journal

eng

MDPI

Article

Feature Generation with Genetic Algorithms for Imagined Speech Electroencephalogram Signal Classification

Edgar Lara-Arellano, Andras Takacs *, Saul Tovar-Arriaga and Juvenal Rodríguez-Reséndiz

Facultad de Ingeniería, Universidad Autónoma de Querétaro, Querétaro 76010, Mexico; edgarlaraarellano@gmail.com (E.L.-A.); saul.tovar@uaq.mx (S.T.-A.); juvenal@uaq.edu.mx (J.R.-R.)
* Correspondence: andras.takacs@uaq.mx

Abstract: This work presents a method for classifying EEG (Electroencephalogram) signals generated when a person concentrates on specific words, defined as "Imagined Speech". Imagined speech is essential to enhance problem-solving, memory, and language development. In addition, imagined speech is beneficial because of its applications in therapy fields like managing anxiety or improving communication skills. EEG measures the electrical activity of the brain. EEG signal classification is difficult as the machine learning (ML) algorithm has to learn how to categorize the signal linked to the imagined word. This work proposes a novel method to generate a specific feature vector to achieve classification with superior accuracy results to those found in the state of the art. The method leverages a genetic algorithm to create an optimal feature combination for the classification task and machine learning model. This algorithm can efficiently explore ample feature space and identify the most relevant features for the task. The proposed method achieved an accuracy of 96% using eight electrodes for EEG signal recordings.

Keywords: imagined speech; EEG; feature extraction; genetic algorithm; classification; optimization algorithms; signal processing; biomedical

Academic Editor: Gian Carlo Cardarilli

Received: 29 January 2025
Revised: 27 March 2025
Accepted: 8 April 2025
Published: 10 April 2025

Citation: Lara-Arellano, E.; Takacs, A.; Tovar-Arriaga, S.; Rodríguez-Reséndiz, J. Feature Generation with Genetic Algorithms for Imagined Speech Electroencephalogram Signal Classification. *Eng* **2025**, *6*, 75. https://doi.org/10.3390/eng6040075

1. Introduction

Imagined speech, also known as silent speech or inner speech, is thinking in the form of sound: "hearing" within one's head silently, without the intentional movement of any limbs such as the lips, tongue, or hands [1].

Processing imagined speech is important because it helps us understand how the brain generates and controls inner thoughts, supports communication skills, and aids in developing brain–computer interfaces (BCI) [2].

Imagined speech improves cognitive processes such as memory by activating neural mechanisms associated with internal rehearsal and working memory. Studies suggest that subvocal articulation, whether expressed or imagined, facilitates encoding and retrieval by strengthening the neural circuits involved in language processing. This phenomenon is especially relevant in memory processes, where inner speech helps to organize and retain information [3].

Advances in this field may lead to more effective therapeutic strategies for individuals with speech disorders, as interventions can be designed to target specific neural pathways associated with particular aspects of speech. A speech or language disorder is characterized by difficulties forming or creating sounds crucial for effective communication. These disorders can manifest in various ways, the most common being articulatory, phonological, voice, and resonance disorders. Articulatory disorders involve difficulties physically producing certain sounds, often due to poor tongue, lips, or jaw positioning [4].

References

[1] E. Lara-Arellano, A. Takacs, S. Tovar-Arriaga, and J. Rodríguez-Reséndiz, "Feature generation with genetic algorithms for imagined speech electroencephalogram signal classification," *Eng*, vol. 6, no. 4, 2025. [Online]. Available: https://www.mdpi.com/2673-4117/6/4/75

[2] K. Brigham and B. V. K. V. Kumar, "Imagined speech classification with eeg signals for silent communication: A preliminary investigation into synthetic telepathy," in *2010 4th International Conference on Bioinformatics and Biomedical Engineering*, June 2010, pp. 1–4.

[3] S.-H. Lee, J.-H. Park, and D.-S. Kim, "Imagined speech and visual imagery as intuitive paradigms for brain-computer interfaces," *arXiv preprint arXiv:2411.09400*, 2024.

[4] L. Jäncke, N. Langer, and J. Hänggi, "Diminished whole-brain but enhanced peri-sylvian connectivity in absolute pitch musicians," *Journal of Cognitive Neuroscience*, vol. 24, no. 6, pp. 1447–1461, 06 2012. [Online]. Available: https://doi.org/10.1162/jocn_a_00227

[5] S. Geva, P. S. Jones, J. T. Crinion, C. J. Price, J.-C. Baron, and E. A. Warburton, "The neural correlates of inner speech defined by voxel-based lesion–symptom mapping," *Brain*, vol. 134, no. 10, pp. 3071–3082, 09 2011. [Online]. Available: https://doi.org/10.1093/brain/awr232

[6] A. W. Kummer, "Perceptual assessment of resonance and velopharyngeal function," in *Seminars in Speech and Language*, vol. 32, no. 02. © Thieme Medical Publishers, 2011, pp. 159–167.

[7] J. Broomfield, "The nature of referred subtypes of primary speech disability," *Child Language Teaching and Therapy*, vol. 20, no. 2, pp. 135–151, 2004.

[8] I. DeWitt, "Phoneme and word recognition in the auditory ventral stream," *Proceedings of the National Academy of Sciences*, vol. 109, no. 8, pp. E505–E514, 2012.

[9] C. Cooney, R. Folli, and D. Coyle, "Optimizing layers improves cnn generalization and transfer learning for imagined speech decoding from eeg," *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 10 2019.

[10] B. Min, J. Kim, H. J. Park, and B. Lee, "Vowel imagery decoding toward silent speech bci using extreme learning machine with electroencephalogram," *BioMed Research International*, vol. 2016, 2016.

[11] T. Morooka, K. Ishizuka, and N. Kobayashi, "Electroencephalographic analysis of auditory imagination to realize silent speech bci," in *2018 IEEE 7th Global Conference on Consumer Electronics, GCCE 2018.* Institute of Electrical and Electronics Engineers Inc., 12 2018, pp. 73–74.

[12] A. Balaji, A. Haldar, K. Patil, S. Ruthvik, and V. Baths, "Eeg-based classification of bilingual unspoken speech using ann," 2017.

[13] A. R. Sereshkeh, R. Trott, A. Bricout, and T. Chau, "Eeg classification of covert speech using regularized neural networks," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, pp. 2292–2300, 12 2017.

[14] G. Hernández, "Predicción de eventos epilépticos mediante técnicas de aprendizaje profundo usando señales eeg," *Universidad Autónoma de Querétaro. Facultad de Ingeniería*, 9 2023.

[15] J. T. Panachakel, A. G. Ramakrishnan, and A. G. Ramakrishnan, "Decoding imagined speech using wavelet features and deep neural networks," 3 2020. [Online]. Available: http://arxiv.org/abs/2003.10433http://dx.doi.org/10.1109/INDICON47234.2019.9028925

[16] S. Fernandez-Fraga, M. Aceves-Fernandez, J. Pedraza-Ortega, and S. Tovar-Arriaga, "Feature extraction of eeg signal upon bci systems based on steady-state visual evoked potentials using the ant colony optimization algorithm," *Discrete Dynamics in Nature and Society*, vol. 2018, no. 1, p. 2143873, 2018.

[17] Y. Li, L. Wu, T. Wang, N. Gao, and Q. Wang, "Eeg signal processing based on genetic algorithm for extracting mixed features," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, 6 2019.

[18] H. Ocak, "Optimal classification of epileptic seizures in eeg using wavelet analysis and genetic algorithm," *Signal Processing*, vol. 88, pp. 1858–1867, 7 2008.

[19] A. Albasri, F. Abdali-Mohammadi, and A. Fathi, "Eeg electrode selection for person identification thru a genetic-algorithm method," *Journal of Medical Systems*, vol. 43, 9 2019.

[20] A. Takacs, M. Toledano-Ayala, A. Dominguez-Gonzalez, A. Pastrana-Palma, D. T. Velazquez, J. M. Ramos, and E. A. Rivas-Araiza, "Descriptor generation and optimization for a specific outdoor environment," *IEEE Access*, vol. 8, pp. 52 550–52 565, 2020.

[21] K. Koizumi, K. Ueda, and M. Nakao, "Development of a cognitive brain-machine interface based on a visual imagery method," *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1062–1065, 2018.

[22] J. S. García-Salinas, L. Villaseñor-Pineda, C. A. Reyes-García, and A. A. Torres-García, "Transfer learning in imagined speech eeg-based bcis," *Biomedical Signal Processing and Control*, vol. 50, pp. 151–157, 4 2019.

[23] S. H. Lee, M. Lee, J. H. Jeong, and S. W. Lee, "Towards an eeg-based intuitive bci communication system using imagined speech and visual imagery," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, vol. 2019-October. Institute of Electrical and Electronics Engineers Inc., 10 2019, pp. 4409–4414.

[24] P. Saha, S. Fels, and M. Abdul-Mageed, "Deep learning the eeg manifold for phonological categorization from active thoughts," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2762–2766, 2019.

[25] J. S. García-Salinas, L. Villaseñor-Pineda, C. A. Reyes-García, and A. A. Torres-García, "Transfer learning in imagined speech eeg-based bcis," *Biomedical Signal Processing and Control*, vol. 50, pp. 151–157, 4 2019.

[26] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2015-August. Institute of Electrical and Electronics Engineers Inc., 8 2015, pp. 992–996.

[27] A. R. Sereshkeh, R. Trott, A. Bricout, and T. Chau, "Eeg classification of covert speech using regularized neural networks," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, pp. 2292–2300, 12 2017.

[28] J. T. Panachakel, A. G. Ramakrishnan, and T. V. Ananthapadmanabha.

[29] A. Jahangiri, J. M. Chau, D. R. Achanccaray, and F. Sepulveda, "Covert speech vs. motor imagery: A comparative study of class separability in identical environments," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2018-July. Institute of Electrical and Electronics Engineers Inc., 10 2018, pp. 2020–2023.

[30] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Inferring imagined speech using eeg signals: a new approach using riemannian manifold features," *Journal of Neural Engineering*, 2 2018.

[31] R. de la Peña Carlos Agosto, "Análisis de ondículas para señales de eeg en el habla imaginada," 9 2023.

[32] R. Zarei, J. He, S. Siuly, and Y. Zhang, "A pca aided cross-covariance scheme for discriminative feature extraction from eeg signals," *Computer Methods and Programs in Biomedicine*, vol. 146, pp. 47–57, 7 2017.

[33] M. A. Rahman, M. F. Hossain, M. Hossain, and R. Ahmmed, "Employing pca and t-statistical approach for feature extraction and classification of emotion from multichannel eeg signal," *Egyptian Informatics Journal*, vol. 21, pp. 23–35, 3 2020.

[34] H. Wang, L. Zhang, and L. Yao, "Application of genetic algorithm based support vector machine in selection of new eeg rhythms for drowsiness detection," *Expert Systems with Applications*, vol. 171, 6 2021.

[35] R. Chaurasiya, "Statistical wavelet features, pca, and svm based approach for eeg signals classification," *nternational Journal of Electrical, Computer, Electronics and Communication Engineering*, vol. 9, pp. 164–168, 1 2015.

[36] O. Guy-Evans, "Wernicke's area: Location and function," September 8 2023, accessed: 2023-10-27. [Online]. Available: https://www.simplypsychology.org/wernickes-area.html

[37] TMSi, "What is the 10-20 system for eeg?" 2022, accessed: 2023-10-17. [Online]. Available: https://info.tmsi.com/blog/the-10-20-system-for-eeg

[38] N. K. Al-Qazzaz, S. H. Ali, S. A. Ahmad, M. S. Islam, J. Escudero, and M. A. Kadir, "Role of eeg as biomarker in the early detection and classification of dementia," *Scientific Reports*, vol. 7, no. 1, pp. 1–17, 2017.

[39] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, "Low-resolution electromagnetic tomography: a new method for localizing electrical activity in the brain," *International Journal of Psychophysiology*, vol. 18, no. 1, pp. 49–65, 1994.

[40] NeuroHealth, "The science of brainwaves – the language of the brain," 2024, accessed: 2023-10-07. [Online]. Available: https://nhahealth.com/brainwaves-the-language/

[41] P. Pandey, R. Tripathi, and K. Miyapuram, "Classifying oscillatory brain activity associated with indian rasas using network metrics," *Brain Informatics*, vol. 9, 12 2022.

[42] F. Lopes da Silva, E. Niedermeyer *et al.*, "Electroencephalography, basic principles, clinical applications and related fields," 2005.

[43] D. Nhu, M. Janmohamed, P. Perucca, A. Gilligan, P. Kwan, T. O'brien, C. W. Tan, and L. Kuhlmann, "Graph convolutional network for generalized epileptiform abnormality detection on eeg," 12 2021.

[44] B. J. Fisch, *Spehlmann's EEG Primer: Basic Principles of Digital and Analog EEG*. Elsevier Health Sciences, 1999.

[45] X. Pardell, "Electroencefalógrafo," 2024, accessed: 2023-10-19. [Online]. Available: https://www.pardell.es/electroencefalografo.html

[46] M. R. Nuwer, "Assessment of digital eeg, quantitative eeg, and eeg brain mapping: report of the american academy of neurology and the american clinical neurophysiology society," *Neurology*, vol. 49, no. 1, pp. 277–292, 1997.

[47] R. Maskeliunas, R. Damaševičius, I. Martišius, and M. Vasiljevas, "Consumer-grade eeg devices: Are they usable for control tasks?" *PeerJ*, vol. 4, 04 2016.

[48] A. C. N. Society, "Guideline 5: Guidelines for standard electrode position nomenclature," *Journal of Clinical Neurophysiology*, vol. 23, no. 2, pp. 107–110, 2006.

[49] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "Artifact reduction strategies for eeg," *Journal of Clinical Neurophysiology*, vol. 116, no. 1, pp. 2107–2123, 2005.

[50] E. Niedermeyer and F. L. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 2004.

[51] N. V. Thakor and S. Tong, "Multiresolution decomposition of eeg for noise and artifact removal," *Journal of Clinical Neurophysiology*, vol. 10, no. 3, pp. 397–403, 1993.

[52] J. W. Miller and M. F. Weiner, "Removing interference from the eeg: Methods and consequences," *Journal of Clinical Neurophysiology*, vol. 24, no. 3, pp. 160–165, 2007.

[53] G. Hickok and D. Poeppel, "The cortical organization of speech processing," *Nature Reviews Neuroscience*, vol. 8, no. 5, pp. 393–402, 2007.

[54] L. Marstaller, H. Burianová, and D. C. Reutens, "Dynamic causal modeling of verbal working memory and the n-back task," *NeuroImage*, vol. 132, pp. 34–42, 2016.

[55] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time signal processing (2nd ed.)*. Prentice Hall, 1998.

[56] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 4th ed. Upper Saddle River, NJ, USA: Pearson, 2007.

[57] S. Haykin and B. V. Veen, *Signals and Systems*, 2nd ed. Hoboken, NJ, USA: Wiley, 2005.

[58] I. I. Goncharova, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Emg contamination of eeg: Spectral and topographical characteristics," *Clinical Neurophysiology*, vol. 114, no. 9, pp. 1580–1593, 2003.

[59] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*, 4th ed. New York, NY, USA: McGraw-Hill, 2011.

[60] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw-Hill, 2002.

[61] S. P. Mukherjee, P. K. Banerjee, and B. P. Misra, "Measures of central tendency: The mean," *Journal of Pharmacology & Pharmacotherapeutics*, vol. 2, no. 2, pp. 140–142, 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3127352/

[62] E. H. Livingston, "The mean and standard deviation: what does it all mean?" *Journal of Surgical Research*, vol. 119, no. 2, pp. 117–123, 2004.

[63] H. Abdi, "Coefficient of variation," *Encyclopedia of research design*, vol. 1, no. 5, pp. 169–171, 2010.

[64] T. N. Edelbaum, "Theory of maxima and minima," in *Mathematics in Science and Engineering*. Elsevier, 1962, vol. 5, pp. 1–32.

[65] D. G. Altman and J. M. Bland, "Statistics notes: quartiles, quintiles, centiles, and other quantiles," *Bmj*, vol. 309, no. 6960, pp. 996–996, 1994.

[66] K. Hu, P. C. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley, "Effect of trends on detrended fluctuation analysis," *Physical Review E*, vol. 64, no. 1, p. 011114, 2001.

[67] B. Hjorth, "Eeg analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology Elsevier*, p. 306, 1 1970.

[68] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.

[69] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[70] J. Gibson, "What is the interpretation of spectral entropy?" in *Proceedings of 1994 IEEE International Symposium on Information Theory*, 1994, pp. 440–.

[71] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica*, vol. 31, pp. 277–283, 1 1988.

[72] "An overview of power spectral density (psd) calculations," *ResearchGate*.

[73] M. Massar, M. Fickus, E. Bryan *et al.*, "Fast computation of spectral centroids," *Advances in Computational Mathematics*, vol. 35, pp. 83–97, 2011. [Online]. Available: https://doi.org/10.1007/s10444-010-9167-y

[74] C. Webber and J. Zbilut, "Recurrence quantification analysis of nonlinear dynamical systems," in *Nonlinear Dynamics and Time Series*. Springer, 2005. [Online]. Available: https://www.nsf.gov/pubs/2005/nsf05057/nmbs/chap2.pdf

[75] "Spectral edge frequency of the electroencephalogram to monitor depth of anaesthesia," *British Journal of Anaesthesia*. [Online]. Available: https://www.bjanaesthesia.org/article/S0007-0912%2817%2943581-1/pdf

[76] B. B. Mandelbrot and J. R. Wallis, "Robustness of the rescaled range r/s in the measurement of noncyclic long run statistical dependence," *Water Resources Research*, vol. 5, no. 5, pp. 967–988, 1969.

[77] H. Liu, Z. Li, and J. Zhang, "Singular value decomposition entropy and its application to time-series analysis," *Remote Sensing*, vol. 14, no. 23, p. 5983, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/23/5983

[78] A. Petrosian, "Kolmogorov complexity of finite sequences and recognition of different preictal eeg patterns," in *Proceedings of the IEEE Symposium on Computer-Based Medical Systems*. IEEE, 7 1995, pp. 212–217.

[79] M. J. Katz, "Fractals and the analysis of waveforms," *Cornput. Bid. Med*, vol. 18, pp. 145–156, 9 1988.

[80] F. A. Torrente, J. M. Cortés, P. Nunez, and E. Gaitán, "Relative band power estimation of brain waves using eeg signals: An overview and a new method," *International Journal of Environmental Research and Public Health*, vol. 20, no. 2, p. 1447, 2023. [Online]. Available: https://www.mdpi.com/1660-4601/20/2/1447

[81] A. Baccigalupi and A. Liccardo, "The huang hilbert transform for evaluating the instantaneous frequency evolution of transient signals in non-linear systems," *Measurement*, vol. 86, pp. 1–13, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224116000956

[82] T. M. Mitchell, *Machine learning*. McGraw-hill New York, 1997, vol. 1, no. 9.

[83] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[84] P. Pawara, E. Okafor, M. Groefsema, S. He, L. R. Schomaker, and M. A. Wiering, "One-vs-one classification for deep neural networks," *Pattern Recognition*, vol. 108, p. 107528, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320320303319

[85] K. Pearson, "On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[86] E. A. Team, "Confusion matrix – evidently ai," 2024, accessed: 2024-10-15. [Online]. Available: https://www.evidentlyai.com/classification-metrics/confusion-matrix

[87] D. E. Goldberg and R. Lingle, "Allele frequencies of finite populations under migration selection balance," *Theoretical Population Biology*, vol. 27, no. 1, pp. 1–20, 1985.

[88] J. F. Miller and R. E. Smith, "Genetic algorithms, selection schemes, and the varying effects of noise," *Evolutionary Computation*, vol. 3, no. 2, pp. 129–157, 1995.

[89] J. H. Holland, "Adaptation in natural and artificial systems," in *University of Michigan Press*, 1975.

[90] D. E. Goldberg, "Genetic algorithms in search, optimization, and machine learning," *Addison-Wesley Reading, MA*, 1989.

[91] A. Eiben, *Introduction to Evolutionary Computing*. Springer, 2003.

[92] E. Lara, J. Rodríguez, and A. Takacs, "Eegis - electroencephalogram imagined speech dataset," 12 2024.

[93] J. Broomfield, "The nature of referred subtypes of primary speech disabilitytitulo de articulo," *Child Language Teaching and Therapy*, vol. 2, no. 20, pp. 135–151, 2004.

[94] I. DeWitt, "Phoneme and word recognition in the auditory ventral stream." *Proceedings of the National Academy of Sciences*, vol. 8, no. 109, pp. E505–E514, 2012.

[95] E. Alpaydin, *Introduction to Machine Learning*. MIT Press, 2014.

[96] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "scikit-learn: Machine learning in python," https://scikit-learn.org, 2011, accessed: 2024-05-05.