Universidad Autónoma de Querétaro

Facultad de Informática

Doctorado en Ciencias de
la Computación

EEG PREPROCESSING SIGNALS USING ARTIFICIAL NEURAL NETWORKS FOR BRAIN-COMPUTER APPLICATIONS

TESIS

That as part of the requirements to obtain the degree of
Doctor in Computer Science

Present:

César Javier Ortiz Echeverri

Main advisor

Juvenal Rodríguez Reséndiz

SINODALES

Juvenal Rodríguez Reséndiz
President

Roberto Augusto Gómez Loenzo
secretary

Efrén Gorrostieta Hurtado
Vocals

Jesús Carlos Pedraza Ortega
substitute

Gerardo Israel Pérez Soto
substitute

Centro Universitario
Querétaro, QRO
México.
January 2020

This thesis is dedicated to my parents María and César, they who gave me life and the wings to pursue my dreams. It is also dedicated to Lili...she always helped me find the light in my darkest moments.....

Acknowledgments

I would like to express my profound gratitude to Dr. Juvenal Rodríguez Reséndiz, for the guidance and encouragement and support throughout all these years. I would also like to thank Dra. Rebeca Romo Vázquez for her valuable contributions to the development of this thesis. I'm very grateful to doctors Roberto Augusto Gómez, Efrén Gorrostieta, Gerardo Israel Pérez, and Jesús Carlos Pedraza for being my committee members and giving me constructive suggestions. Finally, I want to make a very special thanks to a great friend and colleague, Sebastián Salazar Colores, with whom I discussed during all these years, from the most elementary matters to the deepest questions of existence itself. To whom I also owe a great deal of knowledge in the technical matters dealt with in this work.

Abstract

Brain-Computer Interfaces (BCI) are systems that provide an alternative communication between the human brain and a computer, where electroencephalography (EEG) is the the non-invasive and most viable way to obtain the electrophysiological activity. However, the EEG register presents several important challenges for both the instrumentation and the signal processing techniques involved in a BCI implementation. Some of the most relevant drawbacks are due to the low signal-to-noise ratio, the presence of undesirable signals such as ocular, cardiac, and muscular activity, as well as a low spatial resolution due to the distance and high impedance between the sources and the location of the electrodes. Therefore, the raw of EEG signals have very low amplitude, very low signal-to-noise ratio, and considerable noise contamination. Digital signal processing, and machine learning have been included in the preprocessing, feature extraction and classification stages of BCI systems in order to improve the signal-to-noise-ratio and hence, increase their efficiency. The first part of the present dissertation consists of comparing the performance of different preprocessing algorithms to estimate the original sources from the EEG registers; particularly, the preprocessing was made using Blind Source Separation (BSS) algorithms. This kind of spatial filters are based on Second Order Statistic (SOS) or High Order Statistic (HOS) information. Most representative BSS algorithms are (SOBI, SOBIRO, fastICA, and Infomax) were compared using semi-simulated sources, using the Pearson's correlation coefficient and the Wavelet Coherence (WC) as metrics. This analysis was conducted in the electrophysiological bands.

On the other hand, an analysis of different descriptors in time and frequency domain was performed for the extraction of relevant features was made in order to find the most relevant information and thus reduce the dimensionality of input data. In this stage, the used classifier was an Multilayer Perceptron (MLP).

Finally, new method built from the combination of a Blind Source Separation (BSS) to obtain estimated independent components, a 2D representation of these component signals using the Continuous Wavelet Transform (CWT), and a classification stage using a Convolutional Neural Network (CNN) has been proposed. A criterion based on the spectral correlation with a Movement Related Independent Component (MRIC) is used to sort the estimated sources by BSS, thus reducing the spatial variance. The experimental results of 94.21% using a k-fold cross validation are competitive with techniques recently reported in the state-of-the-art.

Contents

List of Figures

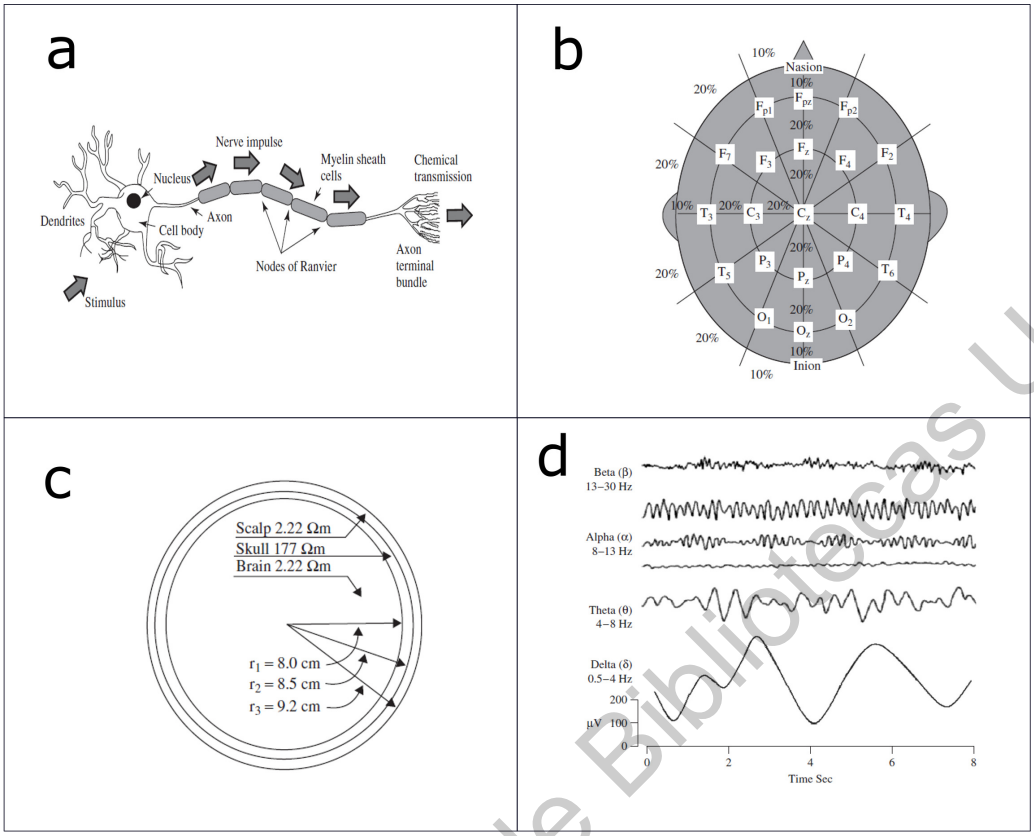
List of Tables

CHAPTER 1

Introduction

The electroencephalography (EEG) is a measurement of currents that flow during synaptic excitations of the dendrites of many pyramidal neurons in the cerebral cortex generated by differences of electrical potentials that create electrical dipoles between the body of a neuron, and apical dendrites, which branch from neurons (Figure 1.1 *a*). The current in the brain is generated mostly by pumping the positive ions of sodium, $Na+$, potassium, $K+$, calcium, $Ca++$, and the negative ion of chlorine, $Cl-$, through the neuron membranes in the direction governed by the membrane potential [2]. The register of EEG is done by placing electrodes on the scalp which capture electrical impulses associated with neuronal activity in the brain cortex, cortex. Generally, these electrodes are placed according to the 10-20 system as shown in Figure 1.1 *b*. EEG is a non-invasive and low-cost technique to register the brain activity and therefore, a good alternative to obtain certain information from the brain cortex. However, the human head is formed by different layers such as the scalp, skull, and many other thin layers in between as is illustrated in Figure 1.1 *c*. Therefore, the signal-to-noise-ratio decreases significantly in the observed signals obtained by electrodes. Furthermore, the observed signals are also a mixture of original sources, for this reasons becomes indispensable the implementation of algorithms for noise reduction and source separation in the signal processing of EEG measurements. Most of the concepts in EEG digital signal processing have their origin in distinct disciplines such as seismological analysis, communications engineering, speech and music information retrieval. In all mentioned applications, the principal topics include time-domain, frequency-domain, and time-frequency domain analysis. These transforms are useful because the EEG activity is in many cases analyzed in the frequency domain in four principal bands that present biological significance depicted in Fig 1.1 *d*.

Delta δ (0.5-4 Hz) observed in infants and sleeping adults, theta θ (4-7.5 Hz) children and sleeping adults, alpha α (8-13 Hz) detected in the occipital brain region when there is no attention, and beta β (14-26 Hz) appears frontally and parietally with low amplitude [2]. In recent years, EEG has become in an essential element in the development of Brain-Computer Interfaces (BCI), designed for mainly for people with severe neuromuscular disorders such as spinal cord injury, amyotrophic lateral sclerosis, stroke and cerebral palsy. A BCI in simple terms, is a direct interface between the human brain and an artificial system that can be designed either for medical diagnosis of brain signals or even use such signals to interact with robotic systems, such as automated wheelchairs or limb prostheses. The principal steps of signal processing present in a BCI are: 1) preprocessing, which try to reduce the noise and artefacts influence in the observed signals; 2) feature extraction, which make a dimensional reduction of the data, taking the relevant information and discarding



(a)

Figure 1.1: a) Biological neuron b)10-20 system c)The three main layers of the brain d)electrophysiological bands in EEG signals

redundant information; and 3) classification, where the intended motion is detected and converted into a control signal for some external device (wheelchair, prosthesis, mouse pointer etc).
According to the nature of the input signal, the BCI systems can be classified into two groups: 1) Endogenous and 2) Exogenous. . Endogenous BCI depend on the ability to control the electrophysiological activity, such as the amplitude of the EEG in a specific frequency band over a specific area of the cerebral cortex; control signals for endogenous BCI systems can be Motor Imagery (MI) based on performing mental tasks that produce changes in the amplitude of sensorimotor rhythms recorded on the somatosensory and motor zone of the cerebral cortex, or Slow Cortical Potentials (SCP) that are slow changes of voltage generated on the cerebral cortex with a variable duration between 0.5 and 10 seconds. On the other hand, exogenous systems depend on electrophysiological activity evoked by external stimuli. Control signals for exogenous BCI systems can be P300 evoked potentials, based on amplitude peaks generated 300 milliseconds after an infrequent auditory or visual stimulus or Stable State Visual Evoked Potentials (SSEVP), which are generated when the person is subjected to visual stimuli. Endogenous BCI is a more natural form of communication between the user and the system to be controlled. However this approach represent a major challenge in terms of signal processing, particularly in the preprocessing stage, due the low signal-to-noise-

ratio and the non-linearity of the sources.
For this reason, endogenous BCI systems require a preprocessing stage to improve the signal-to-noise-ratio prior to the feature extraction and classification stages, by filtering the noise or unwanted signals that are embedded within the EEG, such as the muscle movements called Electromyographic signals (EMG), the activity due to movement of the eyes called Electroculographic signals (EOG), and the heart activity called Electrocardiographic signals (ECG), should also be considered as unwanted signals or artifacts. An appropriate preprocessing tool can enhance the performance of the entire BCI system while using the same feature extraction and classification processes, so preprocessing the EEG signal is often considered the most important BCI stage [3]. Some approaches for preprocessing stage in BCI are based on time and frequency domain filters (band-pass, band-stop, kalman filter, among others). Since the EEG registers are obtained using electrodes placed in different spatial regions on the scalp, the use of space filters has become common in EEG preprocessing. Some of frequently used spatial methods in BCI are: Common Average Reference (CAR); Laplacian method, and Common Spatial Patterns (CSP). These approaches detects the spatial patterns in the EEG to improve in the accuracy of classification, but a minor change in the position of the electrodes may lead to a loss in the gained improvements [3].
On the other hand, statistical methods have been also implemented in preprocessing stage, some of most reported in endogenous BCI systems are based on the Blind Source Separation (BSS) concept. Several works show the use of BSS for artefacts rejection [4, 5, 6, 7, 8, 9, 10]. The BSS is a method that estimates the original source signals of a mixture, as in the case of EEG recording.
The main objective of this work is the study of the BSS algorithms as preprocessing stage in endogeous BCI systems, evaluating their performance in later stages of extraction and classification of features.

1.1 Motivation

The BCI systems are fundamental in the development of new bionic prostheses that allow people to control specific movements of certain mechanisms with their thought. At present, there are some prostheses and exoskeletons that have already begun to incorporate this technology. According to *Instituto Nacional de Estadística y Geografía* (INEGI), in México 6, % of the population has some disability. It is estimated that of the country's 31.5 million households, 6.1 million live with at least one person with a disability (19 out of 100). In 78% of them there is one person with a disability, in 18% two persons and in 3% three or more (INEGI 2012). Walking difficulties are the most common type of disability (64%), moving or using arms or hands (33%), bathing, dressing or eating (24%), and speaking or communicating (18%) (INEGI 2016). [11] It is therefore important to promote, in Latin American countries, research focused on the development of technologies that improve the quality of life of people who for different circumstances are in a condition of motor disability, which involves forming interdisciplinary groups ranging from deep understanding in neuronal processes to bio-mechanics of the prostheses to be designed. Particularly this work is particularly focused on the preprocessing, feature extraction, and classification of MI BCI systems, due the great challenge to obtain relevant information from MI EEG in a non-invasive manner [12]. The results presented in this research suggest that a proper selection of BSS algorithm, as well as the parameter adjustment in feature extraction and classification stages, can improve the performance and classification accuracy of endogenous BCI systems.

1.2 Objectives

The general objective and the specific objectives of this project are as follows:

1.2.1 General objective

To study different BCI configurations based on BSS algorithms and artificial neural networks.

1.2.2 Specific Objectives

- To study the statistical characteristics of MI-EEG signals.
- To study the mathematical concepts behind BSS algorithms.
- To design a methodology to measure the performance of BSS algorithms.
- To implement feature extraction based on time and frequency descriptors.
- To implement classification stage based on machine learning (ML) approaches.

1.3 Thesis Structure

The thesis is organized as follows:

- Chapter 2 is about the theoretical concepts of BCI systems, primarily focused on BSS algorithms and nature of the EEG sources.
- Chapter 3 describes about the feature extraction technique to reduce the dimensionality of the data, as well as two proposed methods to classify the MI intentions, based on Multilayer Perceptron (MLP) and Convolutional Neural Networks (CNN) respectively.
- Chapter 4 presents the proposed methodology and the obtained results.
- Chapter 5 presents the conclusions and future works raised from this research.

1.4 Competitiveness

In this section is analyzed the competitiveness of the proposed thesis. According to the formal definition, competitiveness is the ability of a person or organization to develop competitive advantages over its competitors and thus obtain a prominent position in its environment. Usually is based on a competitive advantage, i.e., certain skill, resources, technology or attributes that make it superior to the person who has it. This is a relative concept in which the performance of a person or organization is compared to others. In the particular case of this research focused on the preprocessing algorithms used in BCI systems that are mainly used in biomedical applications as well as in virtual reality, videogames and neurofeedback or also known as electroencephalographic feedback, which is a neurobehavioral treatment aimed at the acquisition of self-control over certain patterns of brain activity and the application of these skills in daily living activities.

There are currently some commercial BCI devices. However, their performance of correct operation can still be improved as well as the commercial cost. One of the biggest drawbacks of

this kind of devices is that its signal processing algorithms are hidden in closed architectures. For this reason, the study of different approaches to preprocessing, feature extraction and classification has a high significance in order to improve the actual commercial devices and therefore, propose the implementation of a embedded BCI that could ever be competitive in the actual marketplace.

1.5 State-of-the-art

The state-of-the-art review in this work is focused on the three main stages of a BCI system; preprocessing stage, where the input channels are treated to reduce the influence of noise and artifacts, and thus, improve the signal-to-noise-ratio; feature extraction, where relevant information is sought, and redundant data is discarded; and classification, where some machine learning approach search to differentiate between classes of movement intentions.

1.5.1 Preprocessing stage

The primary requirement of the preprocessing stage in BCI is to filter out the noise, artifacts or unwanted signals that are embedded within the EEG. In view of the fact that EEG-based MI classification depends solely on the EEG as the signal of interest such as EMG, EOG, ECG. In Table 1.1 are mentioned the most reported pre-processing methods for EEG signals.

Table 1.1: Preprocessing methods for MI EEG

Method	Reference
Referencing Method	[13],[14],[15],[16]
Principal Component Analysis	[17],[18],[19],[20],[21]
Independent Component Analysis	[22],[23],[24],[25],[26],[27]
Kalman filter	[28],[29],[30]
Autoregresive modeling	[31],[32],[33]

1.5.2 Feature extraction

Feature extraction describes the signal to be classified in terms of its distinguishing invariant features. The extracted features must capture salient signal characteristics, which can be used as a basis for the differentiation between task-specific brain states. Some BCIs involves a process of feature selection, where only the most discriminant of features in a proposed feature set are passed to the classifier with the aim of reducing computation time and increasing accuracy. Based on the selected features. In Table 1.2 are mentioned the most used feature extraction methods for EEG signals.

Table 1.2: Feature extraction methods for MI EEG

Method	Reference
Time-domain methods	[34],[35],[36]
Frequency-domain methods	[37],[38],[39],[40]
Time-Frequency methods	[26],[17],[41],[42],[43]
Common Spatial Patterns	[44],[45],[46],[47],[48],[49]

In the particular case of CSP is one of the most common feature extraction methods used in BCI systems. CSP is a spatial filtering method used to transform EEG data into a new space where the variance of one of the classes is maximized while the variance of the other class is minimized. It is a strong technique for MI EEG processing since different frequency bands of the signal contain different information, and CSP enables the extraction of this information from particular frequency bands. However, pure CSP analysis is not adequate for high-performance MI classification because different subjects exhibit activity in different frequency bands and the optimal frequency band is subject-specific. This means that a wide band of frequencies, typically between 4 Hz and 40 Hz, must be used for MI classification, leading to the inclusion of redundant data being processed [50].

1.5.3 Classification

Finally, the BCI systems are composed of a classification stage, in order to obtain a control signal for a particular application. The most common methods are mentioned in Table 1.3.

Table 1.3: Classification methods for MI EEG

Method	Reference
Linear Discriminant Analysis (LDA)	[37],[51],[52],[53]
Support Vector Machine (SVM)	[54],[55],[56],[57],[43],[58]
k-Nearest-Neighbor	[59],[60],[61],[62]
Naïve Bayes	[63],[64],[65]
Multilayer Perceptron (MLP)	

Some recent BCI implementations based on the mentioned three stages are for example, the detection of Event Related Potentials (ERP) in the EEG signals termed as event-related desynchronization/synchronization (ERD/ERS) patterns. Filter bank and Common Spatial Patterns (CSP) are combined to obtain the features, and a Support Vector Machine (SVM) is used as classifier [66]. Similar Filter Bank with CSP, but using a Naïve Bayes Parzen Window (NBPW) Classifier is proposed in [65].A state-to-state, zero-training method to construct spatial filters for extracting EEG changes induced by motor imagery. Independent component analysis (ICA) was separately applied to the multi-channel EEG in the resting and the motor imagery states to obtain motor-related spatial filters. The resultant spatial filters were then applied to single-trial EEG to differentiate left- and right-hand imagery movements. A Fisher Discriminant Analysis was used as classifier [67]. Empirical Mode Decomposition, Discrete Wavelet Transform, and Wavelet Packet Decomposition) were investigated for the decomposition of EEG Signals in BCI system for a classification task. K-nearest neighbor was used as classifier[68] The use of Empirical Mode Decomposition (EMD) in the preprocessing stage, The features namely: the raw moment of the first derivative of instantaneous frequency, area, the spectral moment of power spectral density, and the peak value of PSD were used as feature extraction, and Least-Squeares SVM (LS-SVM) for classification [69]. A multichannel EEG signal is decomposed into multiple sub-bands (SD), and tangent features (TF) are estimated on each sub-band. Mutual Information Analysis (MIA) and Principal Component Analysis (PCA) select and reduce dimensionality of sub-bands that containing features. Classification is accomplished by a SVM [70]. A tunable-Q wavelet transform (TQWT)-based feature extraction method is proposed for the classification of different MI. The TQWT parameters are tuned for the decomposition of EEG signal into sub-bands. Time domain measures of sub-bands are considered as features for MI tasks EEG signals. The TQWT-based features are tested on

(LS-SVM) classifier for the classification of right-hand and right-foot MI tasks [71]. A variant of ERP called spatio-temporal discrepancy feature (STDF) is proposed, which evaluates the difference of the EEG signals from the left and the right sensorimotor area. STDF as a temporal feature is then combined respectively with three kinds of frequency features describing the ERD/ERS phenomenon to further improve the classification performance [72]. A method to channel selection was proposed that can obtain high classification accuracies. The BCI proposed system is based on ICA spatial filter, where a comparison between ICA and CSP filters was made. They report better results using ICA, as long as the adecuate channels are chosen [73] on ICA spatial filter us proposed to look for the best choice of electrodes A Subject-Specific Decision Tree (SSDT) framework with filter geodesic minimum distance to Riemannian mean (FGMDRM) is designed to identify MI tasks and reduce the classification error in the non-separable region of FGMDRM. Feature extraction algorithm combines semisupervised joint mutual information (semi-JMI) with General Discriminate Analysis (GDA) to reduce the dimension of vectors in the Riemannian tangent plane [74]. Some of recent BCI are summarized in table 1.4.

Table 1.4: Works related to BCI based on MI

Author	Year	Method	Classifier
1. Thomas et al.	2009	Filter Bank-Common Spatial Pattern	Support Vector Machine
2. Chin et al.	2009	Filter Bank-Common Spatial Pattern	Naïve Bayes Parzen Window
3. Lu et al.	2010	Regularized Common Spatial Pattern	Fisher Discriminant Analysis
4. Wang et al.	2012	Independent Component Analysis	Fisher's Discriminant Analysis
5. Zhang et al.	2013	Z-score	Linear Discriminant Analysis
6. Park et al.	2013	Multivariate Empirical Mode Decomposition	Support Vector Machine
7. Aghaei et al.	2016	Separable Common Spatio-Spectral Patterns	Naïve Bayes Parzen Window
8. Siuly et al.	2016	Optimum Allocation	Naïve Bayes
9. Kervic et al.	2017	Multiscale Principal Component Analysis	k-Nearest-Neighbor
10.Taran et al.	2018	Tunable-Q wavelet transform	least squares-Support Vector Machine
11.Taran et al.	2018	Empirical Mode Decomposition-Freq. Feat	least squares-Support Vector Machine
12.Islam et al.	2018	Tangent Space Mapping	Support Vector Machine
13.Luo et al.	2018	Spatio-temporal Discrepancy Feature	Support Vector Machine
14.Zhoua et al.	2019	Independent Component Analysis-time Feat	Trial-to-Trial Test
15.Guan et al.	2019	subject-specific decision tree	k-Nearest-Neighbor

1.5.4 Deep learning approaches

Some systems group together the feature extraction, feature selection and classification tasks within a single signal processing block. These systems are based on deep learning and largely use a convolutional neural network (CNN) structure. The architecture most widely used in MI EEG processing is Convolutional Neural Network (CNN), but Recurrent Neural Network (RNN), Staked Autoencoders (SAE) and Deep Belief Network (DBN) have also been used. CNN hold many advantages for MI EEG data processing raw data can be input to the system thus removing the need to prior feature extraction and they inherently exploit the hierarchical nature of certain signals and they perform well using large datasets. However, their disadvantages are also evident, since the large number of hyper parameters which must be learnt during training can increase the training time compared to other methods, they can produce incorrect classification results with great certainty, and the features learnt can be difficult to understand in the context of the original signal [75]. It should be noted that CNN were adopted in EEG signal processing after first being

established as a tool in image processing. Thus, when using CNN for the classification of MI EEG, one of the greatest differences between approaches involves the preprocessing of the input data, which can mainly be divided into two solutions, i.e., either configuring the EEG data as an image or not configuring the EEG data as an image. In approaches which convert the EEG data to an image, a time-frequency domain image is obtained from the data. Deep learning holds much potential in MI EEG classification. Future work could involve a heavier focus on integrating elements of feature selection. For example, the potential of stacked denoising auto encoders, which has been used to locate robust features [76]. In table 1.5 are depicted some recent MI BCI systems implemented using deep learning approaches.

Table 1.5: Deep Learning approaches for MI BCI

Author	Year	Input data form	Deep learning approach
1. X. An et al.	2014	Time series	Deep Belief Network
2. G. Xu et al.	2016	time-frequency maps with STFT	Pre-trained Convolutional Neural Network
3. Z. Tayeb et al.	2016	time-frequency maps with STFT	Recurrent and Convolutional Neural Network
4. Z. Yin, J. Zhang	2017	Power Spectral Density features	Stacked Denoising AutoEncoder
5. N. Lu et al.	2017	Frequency-domain representation	restricted Boltzmann machine
6. Y. Tabar et al.	2017	time-frequency maps with STFT	Convolutional Neural Network
7. Z. Tang et al.	2017	Time series	Convolutional Neural Network
8. Z. Jiao et al.	2018	time-frequency maps with STFT	Convolutional Neural Network
9. S.Chaudhary et al.	2018	time-frequency maps with STFT and CWT	Deep Convolutional Neural Network (ALexNet)
10.Ruffini et al.	2018	time-frequency maps with STFT	Recurrent and Convolutional Neural Network
11. Z. Zhang et al.	2019	time-frequency maps with CWT	Convolutional Neural Network
12.M.Dai et al.	2019	time-frequency maps with STFT	Convolutional Neural Network and Variational Autoencoder

CHAPTER 2

Blind Source Separation

Blind Source Separation (BSS) is relatively new concept in signal processing. It was initially proposed by Christian Jutten, Bernard Ans and Jeanny Hérault around 1982 from a problem of movement decoding in neurobiology. BSS is related to the problem about separate the independent sources which can only be observed in a mixed way by an array of sensors (microphones, electrodes, antennas). The term "blind" refers to the lack of information about the mixture parameters. BSS is a technique with many applications, in different fields as bio-medical, telecommunications, acoustics, seismographic, exploration, geophysics, signal analysis, etc. For the particular case of electroencephalography, the main application is the elimination of artifacts. However, if the stated objective is the same, the different authors do not model the problem in the same way and therefore do not use the same method of separation. As a result, there is no consensus in the literature on the most suitable algorithm for this application. BSS covers a large number of methods depending on the characteristics and assumptions of the mixing model and sources. Mixtures could be linear, non-linear and convolutive; in the linear case the observed signals are linear combinations of the sources at the same time. In the convolutive case, the sources may contribute to the mixture with several different delays: The difference between the linear convolutive mixing model and the linear instantaneous one, is that delayed values of the source signals contribute to the output at a given time. this may happen in some applications such as telecommunications where channel models often include multipath propagation. In the nonlinear case, the BSS can be expressed as $\mathbf{s}(t) = \mathscr{A}\mathbf{x}(t)$ where $\mathscr{A}$ is a nonlinear transform. In this thesis, we are mainly interested in the separation of instant linear mixtures considering that the various sources (of brain origin and non-brain) are added by neuronal conduction and recovered by the surface EEG, it can be assumed that each $\mathbf{x}$ is an instantaneous linear mixture from unknown sources.
As an example of the linear case, for two different signals s_1 and s_2 we could have two simultaneous measurements $x_1(t)$ and $x_2(t)$ over time t,

$$\begin{aligned} x_1(t) &= a_{11}s_1 + a_{12}s_2, \\ x_2(t) &= a_{21}s_1 + a_{22}s_2 \end{aligned} \tag{2.1}$$

or expressed in matrix form as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{2.2}$$

where $\mathbf{x}(t)$ is the vector of the mixed signals, $\mathbf{A}$ is the unknown non-singular mixing matrix. Then,

BSS consist of finding a $\mathbf{B}$ that in an ideal case $\mathbf{B} = \mathbf{A}^{-1}$ and then the estimated signals will be

$$\mathbf{s}(t) = \mathbf{B}\mathbf{x}(t) \tag{2.3}$$

However, the exact inverse of $\mathbf{A}$ is, in theory, impossible to obtain. Consequently, the sources can be estimated, with the exception of their order (permutation) and their amplitude. BSS algorithms can be grouped according to these assumptions into three main families:

- A first kind of hypothesis involves uncorrelated and stationary sources by windows. The solution for this type of source is simultaneous decorrelation for all windows. In practice, this decorrelation is achieved by diagonalization of the different correlation matrices. For this solution, the Second Order Statistics (SOS) is sufficient.

- A second kind of very common assumption is the statistical independence of stationary sources. In this case, the solution is the decorrelation of the followed signals a maximization of their statistical independence, estimated by assuming stationary signals. This Independent Component Analysis (ICA) seeks to find estimates of unknown sources by assimilating them to signals with maximum independence. In general, the algorithms have two steps: decorrelation followed by maximization independence through Higher Order Statistics (HOS).

- A last type of hypothesis involves stationary, auto-correlated sources and uncorrelated regardless of the offset. The solution based on this assumption is the simultaneous decorrelation of signals for different offsets and therefore the simultaneous diagonalization of the different auto-correlation matrices. So the Second order statistics (SOS) is also sufficient for this solution.

According to the assumptions made about the sources and under the assumption of a linear mixture instantaneous, a possible approach to address the separation problem is to find a linear transformation of the measured signals such as the sources $\mathbf{s}'$ are as close as possible to the original sources $\mathbf{s}$. The common objective of the different algorithms (HOS and SOS) is to estimate a matrix $\mathbf{B}$ separation, which allows the estimation of sources $\mathbf{s}$ from the $\mathbf{x}$ measurements.

2.1 Whitening

We can consider that the first step towards separation is the whitening or decorrelation. This step consist in the estimation of a a matrix $\mathbf{W}$

$$\mathbf{z} = \mathbf{W}\mathbf{x} \tag{2.4}$$

such that correlation matrix of $\mathbf{z}$

$$\mathbf{R}_{\mathbf{z}} = \mathbb{E}\left(\mathbf{z}\mathbf{z}^T\right) \tag{2.5}$$

are diagonal, i.e., $\mathbf{R}_{\mathbf{z}} = \mathbf{I}$. A first method for this purpose is based on the variance-covariance matrix of the observations

$$\mathbf{R}_{\mathbf{x}} = \mathbb{E}\left(\mathbf{x}\mathbf{x}^T\right) \tag{2.6}$$

for which it is useful the eigenvalue decomposition (EVD) given by

$$\mathbf{R_x} = \mathbf{VDV}^T \tag{2.7}$$

where $\mathbf{V}$ is an orthogonal matrix of eigenvectors and $\mathbf{D}$ is the diagonal matrix of eigenvalues. Therefore

$$\mathbf{R_z} = \mathbf{WVDV}^T\mathbf{W}^T \tag{2.8}$$

To obtain a matrix $\mathbf{R_Z}$ equal to $\mathbf{I}$, a solution for the whitening matrix $\mathbf{W}$ is:

$$\mathbf{W} = \mathbf{D}^{-1/2}\mathbf{V}^T \tag{2.9}$$

where the matrix $\mathbf{D}^{-1/2}$ is computed as $diag(d_1^{-1/2}, ..., d_n^{-1/2})$
An equivalent approach is to perform Singular Values Decomposition (SVD) directly on the observed signal matrix $\mathbf{x}$ (which is equivalent to estimating the covariance matrix from the observations). As by hypothesis, $\mathbf{R_s} = \mathbf{I}$,

$$\mathbf{R_z} = \mathbf{WR_xW}^T = \mathbf{WAR_sA}^T\mathbf{W}^T = \mathbf{WAA}^T\mathbf{W}^T = \mathbf{I} \tag{2.10}$$

that is, if the sources are whitening, the product between $\mathbf{WA}$ is a orthogonal matrix. In Figure 2.1 are shown an example of whitening for two signals, mixed and whitening, and their respective joint distributions.

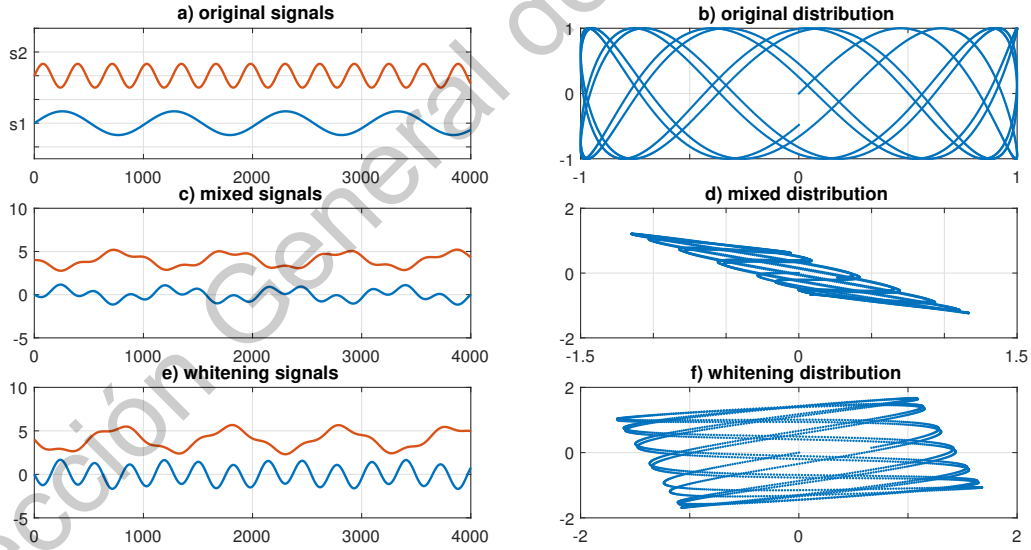


Figure 2.1: whitening example for two signal

It can be noticed in Figure 2.1 b) that the joint distribution of original signals is projected in a rectangular plane, while the joint distribution of mixed signals are projected in a parallelogram. Once whitening the mixture, the joint distribution is more like the original distribution, but rotated with reference to the original plane. This concept must be generalized to r number of signals ($\mathcal{R}^r$ dimentional space) [77].

2.2 Rotation

The estimation of the separation matrix $\mathbf{B}$, in most BSS algorithms, is carried out in two steps. In the first step, a decorrelation of the observations is performed with the estimation of a spatial whitening matrix $\mathbf{W}$. The second step is dedicated to the estimation of an orthogonal transformation $\mathbf{J}$ called rotation matrix which is estimated by minimizing cost functions either by using HOS or SOS. Separation matrix $\mathbf{B}$ is now write as $\mathbf{B} = \mathbf{JW}$ with $\mathbf{W}$ obtained in the previous step. Two main approaches mentioned before (SOS and HOS) are presented in the literature to estimate $\mathbf{J}$

2.2.1 Estimation of rotation matrix with SOS

If the sources are self-correlated (or not stationary), then the second order statistic provides a solution that is satisfactory even to separate Gaussian sources. The starting hypothesis on the sources are based on the decorrelation of the sources instead the statistical independence (case for HOS algorithms).

Simultaneous diagonalization One of the possibilities to minimize mutual information consist in minimize simultaneously the cross-cumulants of estimated sources. A simple example is covariance: minimizing the different covariances implies a decorrelation of the sources, therefore the diagonalization of the covariance matrix. This decorrelation of the estimated sources is a necessary condition of independence. The same simultaneous diagonalization method can be applied by considering shifted time covariance matrices instead high order cumulant matrices [78]. All these covariance matrices have as their objective, finally, to encode the signal dependency (or independence) information, and they will be estimated from the measured data. The source separation, under certain decorrelation hypotheses, can therefore be obtained by the simultaneous diagonalization of two or more covariance matrices extracted from the data. If this diagonalization is performed after bleaching, its objective is is to find the orthogonal rotation matrix $\mathbf{J}$. The correlation matrix $\mathbf{R}_x$ is whitened using the whitening matrix $\mathbf{W}$. The whitened sources $\mathbf{z}$ are represented by

$$\mathbf{z} = \mathbf{Wx} = \mathbf{D}^{-1/2}\mathbf{V}^T\mathbf{x} \tag{2.11}$$

with the correlation matrix $\mathbf{R}_\mathbf{z} = \mathbf{I}$, it is possible to define a correlation matrix $\mathbf{R}_{x(\tau)}$ expressed as

$$\mathbf{R}_{x(\tau)} = \mathbb{E}[x(k)x(k+\tau)^T] = \mathbf{A}\mathbb{E}[x(k)x(k+\tau)^T]\mathbf{A}^T = \mathbf{A}\mathbf{R}_{s(\tau)}\mathbf{A}^T \tag{2.12}$$

therefore,

$$\mathbf{R}_{z(\tau)} = \mathbf{W}\mathbf{R}_{x(\tau)}\mathbf{W}^T = \mathbf{D}^{-1/2}\mathbf{V}^T\mathbf{R}_{x(\tau)}\mathbf{V}\mathbf{D}^{-1/2} \tag{2.13}$$

A unit transformation can be found for the diagonalization of the matrix $\mathbf{R}_{z(\tau)}$ for a given τ. The decomposition in eigenvalues is expressed as

$$\mathbf{R}_{z(\tau)} = \mathbf{J}\mathbf{\Sigma}_2\mathbf{J}^T \tag{2.14}$$

where $\mathbf{\Sigma}_2$ is a diagonal matrix and $\mathbf{J}$ is an orthogonal matrix. Assuming $y(k) = \mathbf{J}^T\mathbf{z}_(k)$ is satisfied that

$$\mathbf{R}_y = \mathbf{J}^T\mathbf{R}_z\mathbf{J} = \mathbf{I} \tag{2.15}$$

$$\mathbf{R}_{y(\tau)} = \mathbf{J}^T \mathbf{R}_{z(\tau)} \mathbf{J} = \mathbf{\Sigma}_2 \tag{2.16}$$

These two matrices $\mathbf{R}_z$ and $\mathbf{R}_{z(\tau)}$ are diagonalized simultaneously by the linear tranformation.

$$\mathbf{B} = \mathbf{J}^T \mathbf{W} = \mathbf{J} = \mathbf{J}^T \mathbf{\Sigma}_e^{-1/2} \mathbf{V}^T \tag{2.17}$$

that satisfies the equation 2.10.

Diagonalization of covariance matrices for different offsets Starting from the assumption of stationary, auto-correlated and uncorrelated sources regardless of the offset τ, we have a set of offset correlation matrices $\mathbf{R}_{x(\tau)}$ to diagonalize, using the approximate attached diagonalization (exact attached diagonalization is generally impossible).
Considering $R_x = \left[\mathbf{R}_{x(0)}, \mathbf{R}_{x(1)}, ..., \mathbf{R}_{x(k)}\right]$, the set of correlation matrices with $\mathbf{R}_{x(0)}$ positive. The objective is to determine the separation matrix $\mathbf{B}$ so that the matrices $\mathbf{B}\mathbf{R}_{x(\tau)}\mathbf{B}^T$ are as close as possible to diagonal matrices. The first step is the whitening, which transforms R_x into a new set

$$R_z = \left[\mathbf{R}_{z(\tau)} = \mathbf{W}\mathbf{R}_{x(\tau)}\mathbf{W}^T\right] \tag{2.18}$$

where $\mathbf{R}_{z(0)} = \mathbf{I}$ and $\mathbf{W}$ is the whitening matrix. For an orthogonal $\mathbf{J}$, given $\mathbf{B} = \mathbf{J}^T\mathbf{w}$ performs the exact diagonalization of $\mathbf{R}_{x(0)}$.
Let $O(\mathbf{M}) = \sum_{j \neq l} m^2(l, j)$ be the sum of the squares of the elements outside $\mathbf{M}$ diagonal. The criterion consists in minimizing, compared to the orthogonal matrix $\mathbf{J}$, the quantity,

$$C_\tau\left(\mathbf{J}, \mathbf{R}_{z(0)}\right) = \sum_{\tau=0}^{k} O\left(\mathbf{J}\mathbf{R}_{z(\tau)}\mathbf{J}^T\right) \tag{2.19}$$

The matrix $\mathbf{J}$ is obtained as a product of successive rotations of dimension 2. This method is known as SOBI (Second Order Blind Identification).
In Figure 2.2 is shown an example for the estimation of two signals using an approach based on SOS.

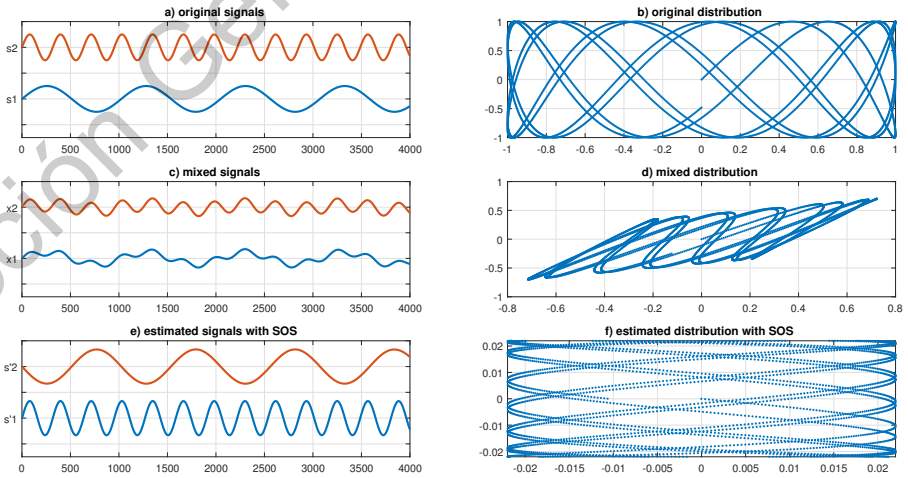


Figure 2.2: SOS estimation example for two signal

2.2.2 Estimation of rotation matrix with HOS

By using HOS, statistical independence is exploited (in the strict sense) of the sources. We obtain criteria for achieving separation without any additional information. This leads to the concept of Independent Component Analysis (ICA). The Kullback-Leibler divergence and mutual information can provide a global insight into the many ICA algorithms developed in recent years. Let $p_1(s_1)$ and $p_2(s_2)$ be two probability densities of two components of $\mathbf{s}$. The Kullback-Leibler divergence is defined by:

$$D(p_1|p_2) = \int_{-\infty}^{\infty} p_1(s_1) \log \left(\frac{p_1(s_1)}{p_2(s_2)} \right) \tag{2.20}$$

The Kullback-Leibler divergence is cancelled when the two components of $\mathbf{s}$ follow the same law: $p_1(s_1) = p_2(s_2)$. The Kullback-Leibler divergence is used to measure the difference between the density and the product of marginal densities. This gap is the measure of independence statistical. Several random variables are referred to as mutually reinforcing components statistically independent if and only if:

$$p(\mathbf{s}) = p_s(s_1, s_2, ..., s_n) = p_1(s_1)p_2(s_2)...p_n(s_n) = \prod_i p_i(s_i) \tag{2.21}$$

where p_s is the distribution of s and $\prod_i p_i(s_i)$ is the product of its marginal distributions. Therefore, by replacing in equation (2.11) $p_1(s_1)$ by $p(\mathbf{s})$ and $p_2(s_2)$ by $\prod_i p_i(s_i)$ and by minimizing the Kullback-Leibler divergence obtained, the independence of s_i is maximized. On the other hand, mutual information can be considered as the distance between joint densities and marginal densities. Thus, the mutual information of $\mathbf{s}$ noted $I\mathbf{s}$, can be expressed as the divergence of Kullback-Leibler between its distribution $p(\mathbf{s})$ and the product of its marginal distributions $p(y_i)$ such that:

$$I(\mathbf{s}) = D \left(p(\prod_i p_i(s_i) \right) \tag{2.22}$$

with $i = 1...q$;
Using the concept of differential entropy, it is possible to reformulate the mutual information $I(\mathbf{s})$ as follows:

$$I(\mathbf{s}) = I(s_1, s_2, ..., s_q) = \sum_i H(s_i) - H(\mathbf{s}) \tag{2.23}$$

where $H(s_i)$ and $H(\mathbf{s})$ are the marginal and joint entropy respectively.
The properties of mutual information suggest an estimate of the structure of separation by minimizing mutual information. The minimization of mutual information is a sufficient theoretical criterion, but which cannot be applied in practice in the form of Equation 2.14. Different estimation methods that lead to the implementation of different independence criteria have been proposed to resolve this problem. The classical algebraic approach of ICA proposed by P. Comon[Comon, 1994], is based on negentropy, which is a measure of the entropy of a random variable with respect to the entropy of a Gaussian distribution variable. By definition negentropy is the Kullback-Liebler divergence between a density of probability $p(\mathbf{s})$ and the Gaussian probability density of $g(\mathbf{s})$ of the same mean and same variance.

The negentropy of $\mathbf{s}$ is defined as

$$J(\mathbf{s}) = H_g(\mathbf{s}) - H(\mathbf{s}) \tag{2.24}$$

where $H_g(\mathbf{s})$ is the entropy of gaussian variable.
According to the deffinition of mutual information, $I(\mathbf{s})$ can be expressed as

$$I(\mathbf{s}) = J(\mathbf{s}) - \sum_i J(y_i) + \frac{1}{2}\log\frac{\prod_i v_{ii}}{\det V} \tag{2.25}$$

where V is the variance-covariance matrix of $\mathbf{s}$, with V_{ii} the diagonal elements. Being since maximizing independence means minimizing mutual information, maximizing of the sum of marginal negentropies is to minimize mutual information after whitening of the observations. This method is also similar to those using the notion of kurtosis. We can then characterize them by the fact that they seek the least Gaussian sources possible. Under these conditions, the separation problem consists in the search for a rotation matrix $\mathbf{J}$ such that $\sum_i H(y_i)$ are minimal. The use of second order moments (or cumulants) is not sufficient to decide if non-Gaussian variables are independent. On the other hand, the use of cumulants (cross-referenced) of all kinds makes it possible to show whether variables are independent or not: if all the crossed cumulants of a set of random variables of all kinds are null, so the random variables are independent. In[Comon, 1994], Comon defines a contrast function using fourth-order cumulants and shows that minimizing mutual information means maximizing the sum of the edges of the cumulants of order four $(Cum_4)^3$ of the estimated sources expressed by

$$I(\mathbf{s}) = \frac{1}{48}\sum_i (Cum_4)^2 \tag{2.26}$$

Other methods of source separation proposed in the literature follow in their principle the same approach and are fundamentally equivalent[Cardoso, 1999]. The main differences are in the method chosen for estimating the mutual information mutual or negentropy[77] as well as the minimization method [79]

2.3 Nature of sources

To understand the rotation matrix in HOS-based algorithms, it is convenient to take a preview analysis of nature of sources and their Probability Density Function (PDF). Considering a first example of six sinusoidal sources with different and independent frequencies, as shown in Figure 2.3a, the PDF of these original sources is shown in Figure 2.3b
in the figure above it can be seen that the recovered and original waveforms are practically identical. However, the algorithm cannot recover the original position of the channel due the blind approach used. The PDF of original, mixed and recovered signals are depicted in Figure 2.3b, 2.3d, and 2.3f respectively. In the figure it can be noticed that the PDF of original signals are non-Gaussian and show fine differences in their amplitudes, while the PDF of mixed signals tend to have a Gaussian distribution. If this condition is satisfied, BSS algorithms based on HOS statistic will find a satisfactory solution, as in this case [80]. Considering a second example of six sinusoidal sources but now with different but harmonic frequencies, as shown in Figure 2.3a, the PDF of these original sources is shown in Figure 2.3b

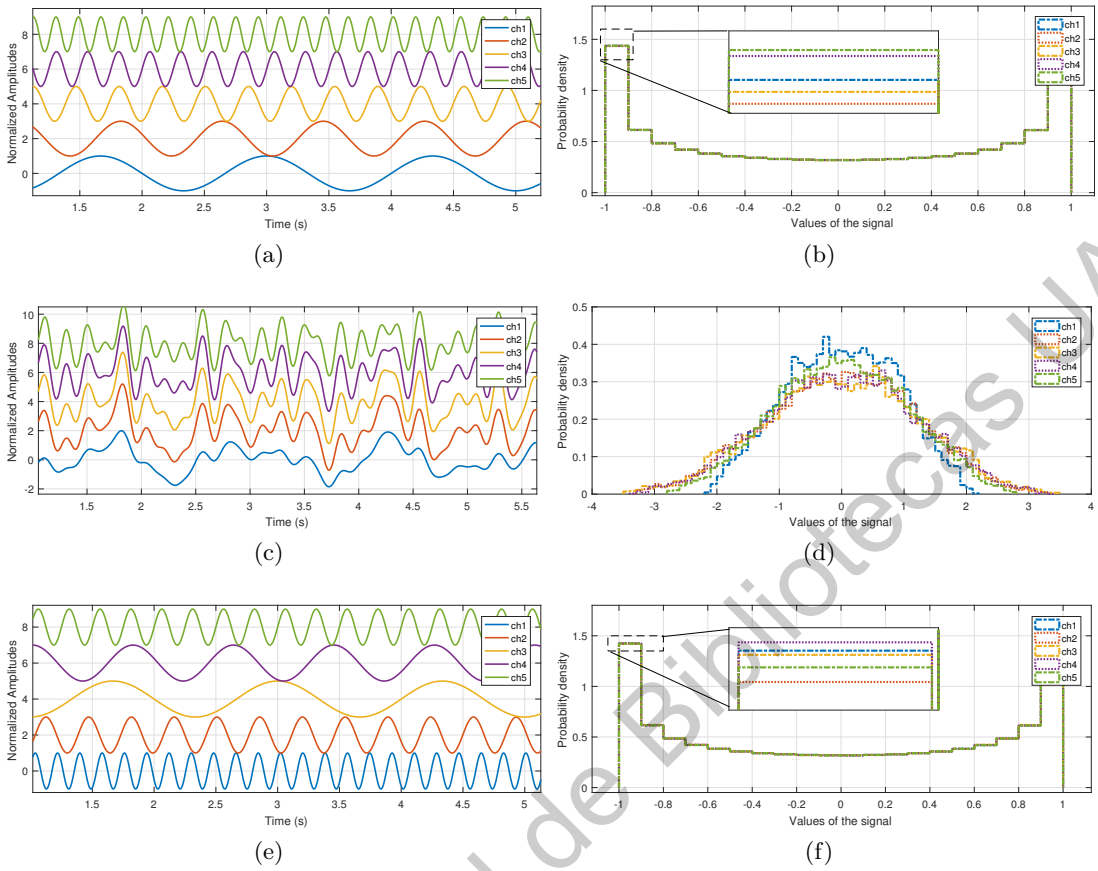


Figure 2.3: Sinusoidal sources with independent frequencies a) original sources (b) PDF of original sources (c)mixed sources(d) PDF of mixed sources (e) recovered sources (f) PDF of recovered sources

Unlike the first case where the frequencies are independent of each other, here there are not differences in the PDF of each source. On the other hand, the PDF of mixed signals in 2.4b are not a Gaussian distribution which is also an important condition for performance of HOS algorithms.

Several BSS algorithms were proposed in last years, some of most used for biomedical applications are are depicted in Table 2.1

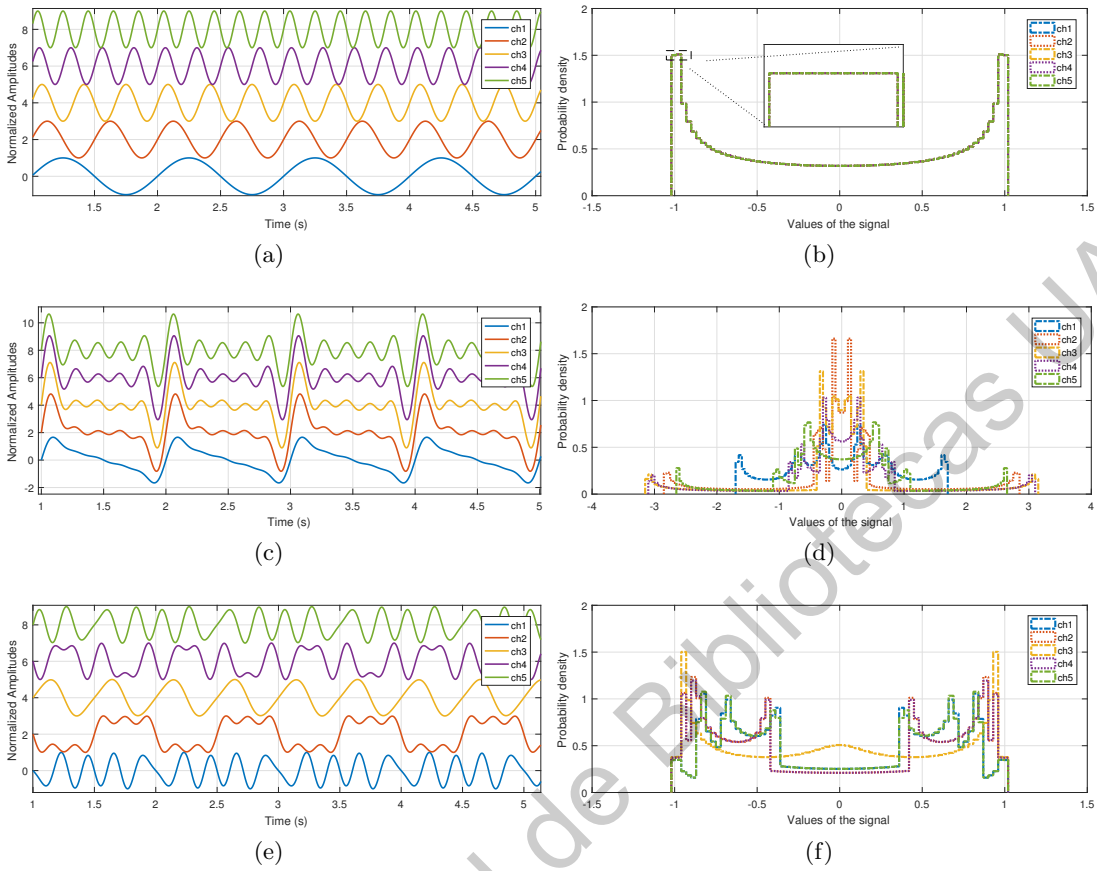


Figure 2.4: Sinusoidal sources with harmonic frequencies (a) original sources (b) PDF of original sources (c)mixed sources(d) PDF of mixed sources (e) recovered sources (f) PDF of recovered sources

Table 2.1: Principal BSS reported algorithms

BSS of EEG correlated with ECoG			
Algorithm	statistical order	whitening	Allowed Gaussian sources
SOBI	2	Yes	All
SOBI-RO	2	$Yes(robust)$	All
TFBSS	2	Yes	All
SOBIUM	2	No	All
COM2	2	Yes	One
JADE	4	Yes	One
STOTD	4	Yes	One
FastICA	4	Yes	One
infomax	4	Yes	One
PICA	> 4	Yes	One

2.4 Simulated signals for the evaluation of BSS algorithms

Despite several comparative studies in literature, there are no consensus about the best BSS algorithm for EEG data, this due to the lack of *a priori* knowledge of original sources. In [8] was proposed a new method based on canonical correlation for quantifying the performance of BSS separation algorithms in EEG signals, comparing with simultaneous measurements of electrocorticography (ECoG) that being obtained directly on the cerebral cortex, and therefore are closer to the original sources. In this chapter is described the proposed methodology for the study and comparisons of BSS algorithms. In order to analyze the behaviour of SOS and HOS, four representative BSS algorithms were selected mentioned below

- SOBI: this SOS algorithm uses the temporal coherence of the source signals, finding the $\mathbf{W}$ minimizing the sum-squared cross-correlations between one component at time t and another component at time $t+\tau$, across a set of time delays. A joint diagonalization of an arbitrary set of time delayed covariance matrices allows the separation of Gaussian sources [81].
- SOBI-RO: this version of SOBI algorithm is adapted for noisy signals by introducing a robust orthogonalization step to reduce influence of white noise.
- fastICA: this algorithm is based on the measure of non-Gaussianity analyzing the fourth-order cumulant known as *kurtosis*, or the entropy, under the assumption that the sources are temporally independent and identically distributed and non-Gaussian, while the observed signals are more nearly to a Gaussian distribution [77].
- RUNICA: this was based on maximizing the output entropy (or information flow) of a neuronal network with non-linear outputs.

The first metric used to compare the sources is the well known Pearson correlation describes by equation

$$Corr_{s,s'} = \frac{cov(s,s')}{\sigma_s \sigma'_s} \tag{2.27}$$

where $cov(s,s')$ is the covariance σ_s and σ'_s the normalize standard deviation of $\mathbf{s}$ and $\mathbf{s}'$ respectively. The second metric used is to analyze the time-frequency performance of the BSS algorithms is the wavelet coherence (WC), which is a method to evaluates the frequential similarity between two temporal signals, in this work the original $\mathbf{s}$ against the estimated signals $\mathbf{s}'$. The WC equation is described by

$$WC_{s,s'}(t,a) = \frac{|W_{s,s'}(t,a)|}{\sqrt{W_s(t,a)W'_s(t,a)}} \tag{2.28}$$

where a is the scale factor, $|W_{s,s'}(t,a)|$ is the wavelet cross spectrum between $\mathbf{s}$ and $\mathbf{s}'$, while $W_s(t,a)$ and $W'_s(t,a)$ are the wavelet auto-spectrum of $\mathbf{s}$ and $\mathbf{s}'$, respectively [82]. Each comparison generates a time-scale map, For the proposed method, we extract four sub-matrix associated with the δ, θ, α, and β bands. Then, the mean of the mean is calculated for each band obtaining a scalar between 0 and 1 where 0 indicates null similitude and 1 absolute similitude in the particular band. The equation for $WCM_{\delta,\theta,\alpha,\beta}$ is

$$WCM_{\delta,\theta,\alpha,\beta} = mean(mean(WC_{\delta,\theta,\alpha,\beta})) \tag{2.29}$$

In Figure 2.5 is shown the proposed method for the generation of WCM. Vectors $\mathbf{s}'(t)$ in the output of BSS algorithms are compared with $\mathbf{s}(t)$ using WC, generating WC_{SOBI}, WC_{SOBIRO}, $WC_{fastICA}$, and $WC_{infomax}$ time-scale maps. A function based on the correlation between s and s' sort the estimates s' signals according to their original signs and positions in order to compare correctly the original and estimated sources. From each WC map, are only extracted and averaged in time and scale the vectors associated with electrophysiological δ, θ, α, and β bands. This process is made for $s'_{1..6}$ generating an array of size 4x1x6 for each BSS algorithm. The process is carried out for n trials (100 in our case).

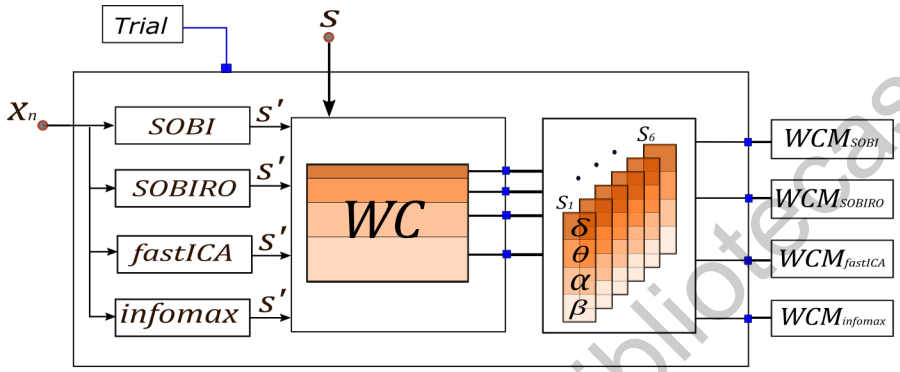


Figure 2.5: Frequency analysis description. The WC is calculated between original and recovered sources for each trial; electrophysiological bands are extracted generating the WCM Method for each BSS algorithm.

BSS algorithms were developed to estimate original sources from the information provides only for the mixed observations. However, to perform a measurement of the performance is essential to know the original sources, i.e. the *ground-truth*, wherefore, in this work two kind of known sources were used for this purpose. In a first case of evaluation we use simulated signals, and subsequently, simultaneous electroencephalogram (EEG) and electrocorticogram (ECoG). In the next sections are describes the two main cases of study. Six semi-simulated sources are used to generate the set of mixed signals composed of: four intracerebral (depth) encephalogram (SEEG) sources placed inside the brain which have fewer artefacts components, extracted from different brain locations, one ocular artefact and one ECG artefact (taken from[7]). In Figure 2.6 are depicted the simulated signals used for the analysis, in a) are the six original sources, in b) the noiseless mixture and in c) a mixture with white noise. For the last case, SNR of 20, 10, and 5 dB were used. For each case, the mixtures were generated 100 times.

According to the proposed methodology, temporal analysis based on correlation and a time-frequency based on WCM were made using mixtures of SNR=20, 10, and 5 dB. In Figure 2.7 are shown the correlations (Corr) between $\mathbf{s}(t)$ and $\mathbf{s}'(t)$ and their respective standard deviation σ. For the case of a noiseless mixture, it can be observed that InfoMax, SOBI, and SOBIRO presented highest Corr for all sources, while fastICA showed the lowest values. Nevertheless, for noise mixtures, the performance of SOBI decreased significantly in comparison with the other algorithms, SOBIRO and InfoMax presented better estimations than SOBI but failed for artefact s_6 with Corr values below 0.4. In contrast with the noiseless case, fastICA achieved a more stable behavior for noise mixtures, this was more evident for the artefact s_6, where the other algorithms presented values of Corr below 0.4, while fastICA achieved values above 0.5 even in the noisiest mixture.

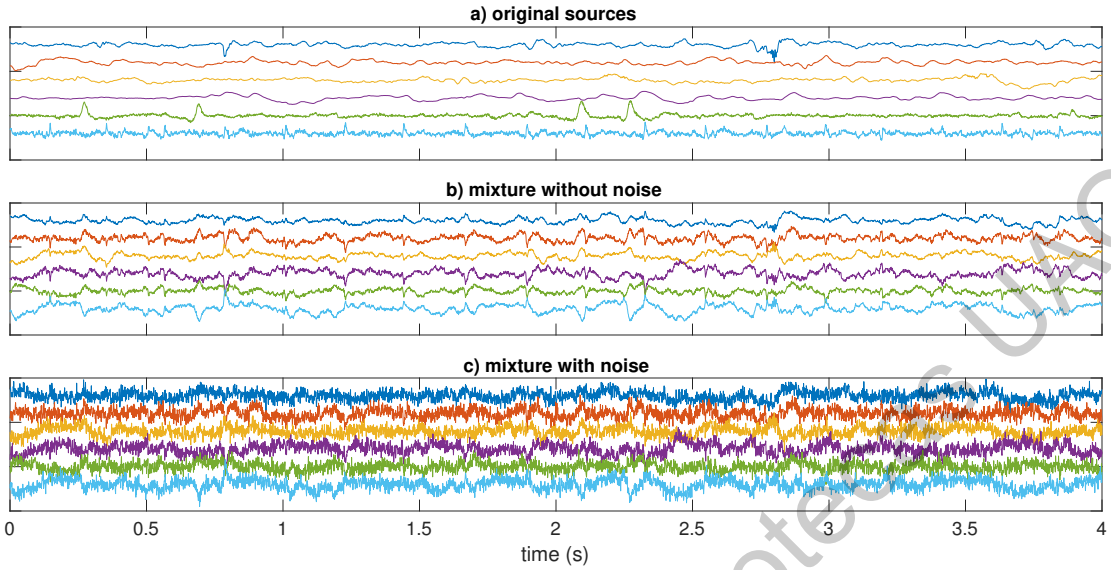


Figure 2.6: a) Original simulated signals, b) mixture without noise, and c) mixture with noise.

Once again the two cases (noiseless and noise mixtures) were evaluated, now applying the proposed WCM to analyze the behavior of evaluated BSS algorithms in electrophysiological bands. In the figures 2.8, 2.9, 2.10, and 2.11 are shown the WCM values for δ, θ, α, and β respectively, for the noiseless, SNR=20, 10, and 5 mixtures.

- δ band: For noiseless mixture InfoMax showed the best performance in this electrophysiological band as well as the best estimation of artefact s_5 with a WCM value of 0.7, while the other algorithms presented values below 0.4. SOBIRO presented similar WCM values except for artefact s_5. SOBI and fastICA presented WCM below 0.8 in s_2. For noise mixtures, SOBIRO followed by InfoMax achieved better performance except for artefacts s_5 and s_6, where all algorithms failed in the identification, while the behavior of fastICA has fewer variations compared to the noiseless case. SOBI presented the lowest values of WCM for sources s_1 to s_4 and close values for s_5 and s_6 to the other algorithms.

- θ band: SOBIRO presented the best WCM values for noiseless and noise mixtures in this band, achieving a WCM value above 0.8 for s_1 and a minimum value of 0.66 for s_3 in the case of SNR=5. InfoMax presented good performance in this band, except for artefact s_6 for noise mixtures. fastICA showed WCM values closest to each other for the different SNR values but with a performance below SOBIRO and InfoMax. Finally, SOBI presented the lowest performance for all values of SNR in noise mixture with all values of WCM below 0.7.

- α band: In the case of noiseless and noise mixture, SOBIRO and InfoMax presented the best performance in this band for sources s_1 to s_5. For artefact s_6 SOBIRO presented highest values of WCM in comparison with the other algorithms. fastICA showed WCM values closest to each other for the different SNR values but with a performance below SOBIRO and InfoMax. SOBI presented WCM values below 0.65 in all sources for all cases of noise mixtures.

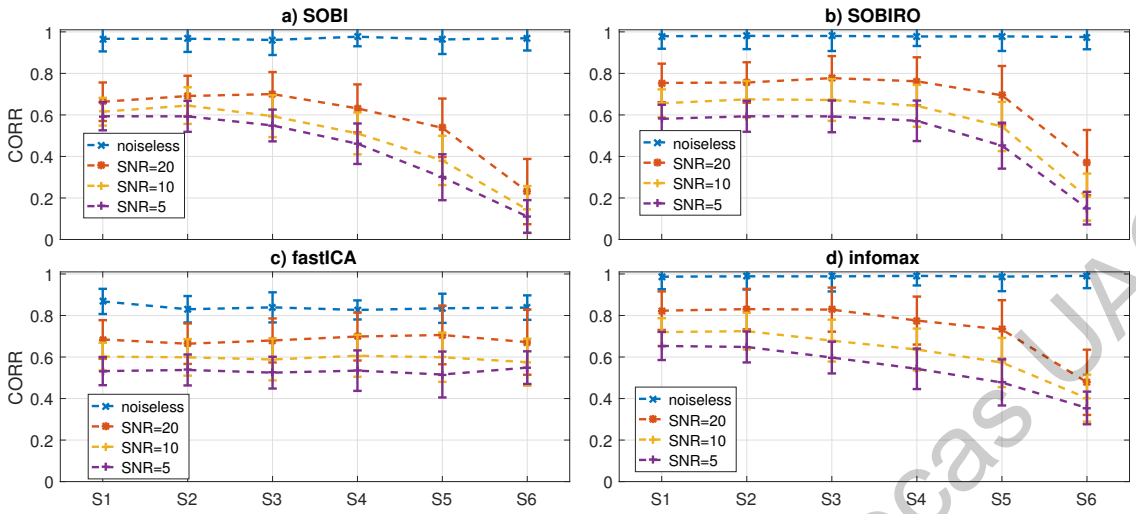


Figure 2.7: Correlation values of the simulated signals with different SNR values for a)SOBI b)SOBIRO c)fastICA d)infomax

- β band: For the noiseless case, InfoMax presented the highest values of WCM in all sources. For the noise mixture, all algorithms present a a low value of WCM in s_4 and high values in artefact s_6. Comparing the rest of sources, SOBIRO presented slightly better performance in s_2, and s_3, fastICA in s_1, and InfoMax in s_5.

Considering that EEG signals are more close to the noise mixture case, the obtained results suggest that SOBIRO is the most suitable BSS, mainly for applications based on MI due to the significance of α and β bands in this kind of applications. However, it can be noticed that a strong presence of noise is a great problem which is not entirely resolved in actual BCI implementations. the results obtained in this chapter show that for source separation in the absence of noise, both SOS and HOS algorithms perform well in the frequency bands studied. However, when mixtures have high noise values, all of the algorithms widely used in EEG signals present a significant decrease in their estimations, even the specific algorithms designed to deal with the noise such as SOBIRO. Most applications perform signal filtering, either using band-pass filters or adaptive filters. However this approach also eliminates information and affects the result of source separation algorithms [7]. HOS algorithms are assuming that the sources are statistically independent with non-Gaussian distribution. SOS-based methods are assuming that autocorrelation characteristics of sources are distinct. The WCM proposed in this chapter can help to obtain detailed information in the performance of BSS algorithms in electrophysiological bands δ, θ, α and β, associated with the EEG biopotentials without the need to apply band-pass filters which alter the waveforms provided by the BSS estimations. These preliminary results suggest that InfoMax have advantages in the absence or low noise case in all studied frequency bands, while SOBIRO and fastICA perform better for noise mixtures, and therefore more suitable for real applications such as EEG preprocessing.

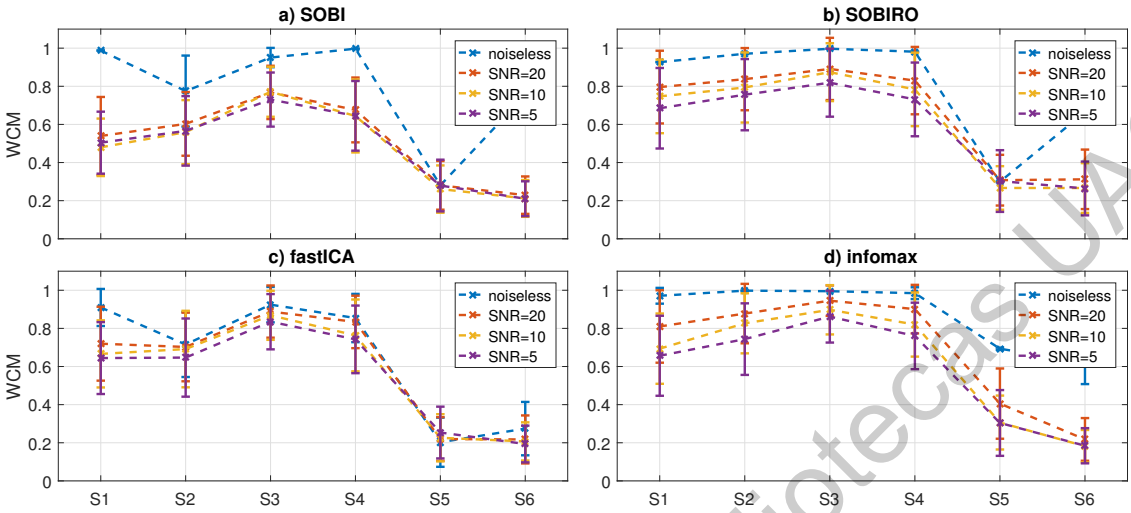


Figure 2.8: WCM for δ band in a)SOBI, b)SOBIRO, c) fastICA, and d) InfoMax

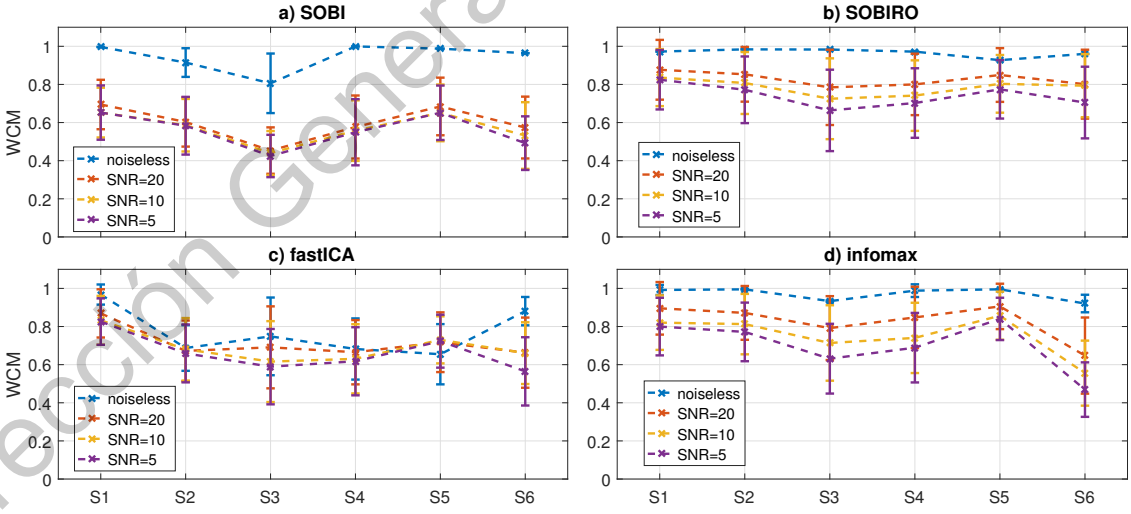


Figure 2.9: WCM for θ band in a)SOBI, b)SOBIRO, c) fastICA, and d) InfoMax

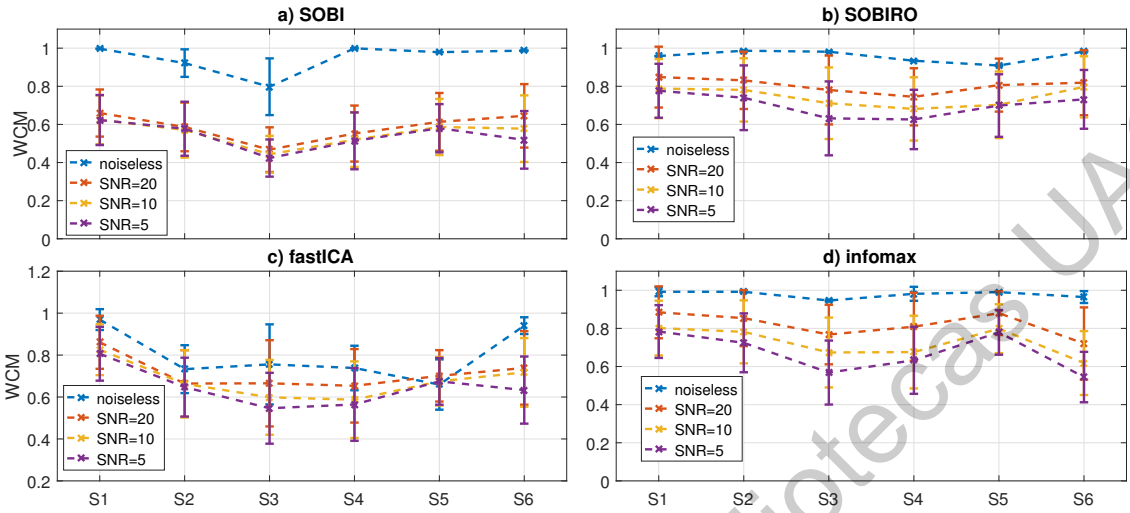


Figure 2.10: WCM for α band in a)SOBI, b)SOBIRO, c) fastICA, and d) InfoMax

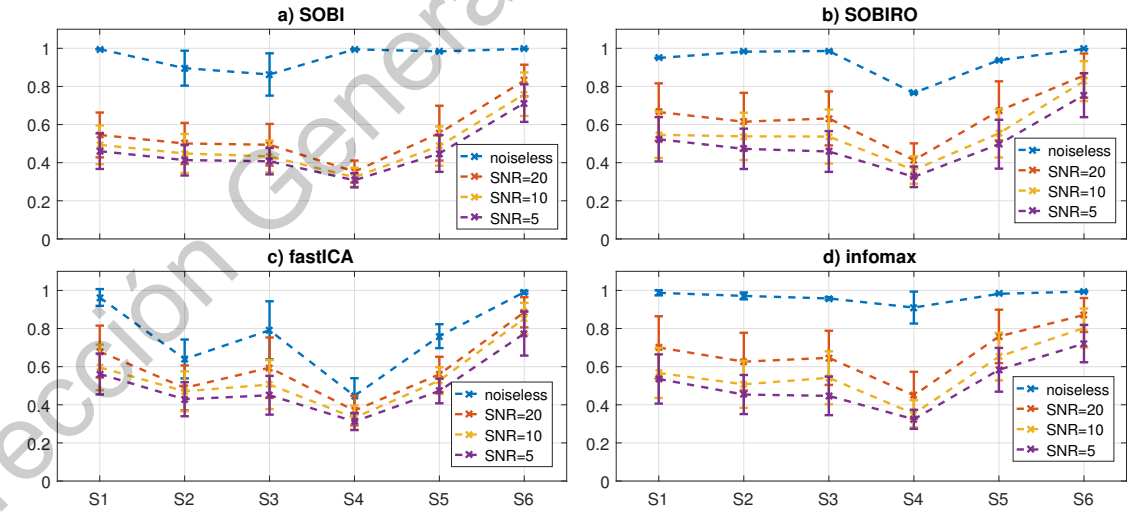


Figure 2.11: WCM for β band in a)SOBI, b)SOBIRO, c) fastICA, and d) InfoMax

CHAPTER 3

Brain-Computer Interface Implementation

3.1 Descriptors

Feature extraction is an important stage in signal processing analysis and is an essential processing step in pattern recognition and machine learning tasks. The goal is to extract a set of features from the signals of interest that represent desired properties of the original data, pondering the differences and minimizing the redundancies at the same time by reducing the dimensionality of the original data. Commonly, the original data, (such as EEG channels) are voluminous and therefore, it is hard to process directly in any analysis task. In order to achieve this goal, it is important to have a good knowledge of the application domain, so that we can decide which are the best features. In the particular case of EEG signals, feature extraction mainly focuses on finding a suitable representation of the data that is lower in dimension relative to the raw EEG, but at the same time preserves the information that is required to discriminate different classes of data.
This stage is crucial in BCI systems, because if the representation is too high dimensional, too correlated, or too noisy, the machine learning algorithm cannot find discriminative information. On the other hand, if the EEG raw is reduced more than necessary, it may result in loss of valuable information. In this chapter,are presented some essential features, describing how to extract these features on a short-term basis and how to compute feature statistics for the case of mid-term audio segments.

3.1.1 Time domain features

Time-domain features are extracted directly from the samples of the signals. Typical examples are the short-term energy and short-term zero-crossing rate. Such features offer a simple way to analyze audio signals.

Energy Considering a fragment of EEG channel as $x_i(n)$, with a length of L, the short-term energy is computed according to the equation

$$E(i) = \frac{1}{L}\sum_{i}^{L} |x_i(n)|^2 \tag{3.1}$$

Short-term energy is expected to exhibit high variation over successive EEG frames, i.e. the energy envelope is expected to alternate rapidly between high and low energy states.

Zero-Crossing rate The Zero-Crossing Rate (ZCR) of an EEG frame is the rate of sign-changes of the signal during the frame. In other words, it is the number of times the signal changes value, from positive to negative and vice versa, divided by the length of the frame. The ZCR is defined according to

$$Z(i) = \frac{1}{2L} \sum_{i}^{L} |sgn[x_i(n)] - sgn[x_i(n-1)]|^2 \tag{3.2}$$

where sgn($\cdot$) is the sign function, (1 for $x_i(n) \geq 0$; 0 for $x_i(n) < 0$).

ZCR can be interpreted as a measure of the noisiness of a signal. For example, it usually exhibits higher values in the case of noisy signals. It is also known to reflect, in a primitive way, the spectral characteristics of a signal.

Entropy of energy The short-term entropy of energy can be interpreted as a measure of abrupt changes in the energy level of a signal. In order to compute it, each short-term frame is divided in K sub-frames $EsubFrame_j$ of fixed duration. Then, for each sub-frame j ,the energy is computed and divided by the total energy, $EshortFrame_i$, of the short-term frame. The division operation is a standard procedure and serves as the means to treat the resulting sequence of sub-frame energy values, e_j, $j = 1, ..., K$, as a sequence of probabilities, as in

$$e_j = \frac{EsubFrame_j}{EshortFrame_i} \tag{3.3}$$

where

$$EshortFrame_i = \sum_{k=1}^{k} EsubFrame_k \tag{3.4}$$

At a final step, the entropy,$H(i)$ of the sequence e_j is computed according to

$$H(i) = -\sum_{j=1}^{k} e_j log_2(e_j) \tag{3.5}$$

The resulting value is lower if abrupt changes in the energy envelope of the frame exist. This is because, if a sub-frame yields a high energy value, then one of the resulting probabilities will be high, which in turn reduces the entropy of sequence e_j .Therefore,this feature can be used for the detection of significant energy changes.

3.1.2 Frequency domain features

Fourier transform can provide information in the frequency domain that in the time domain is difficult to find. Based on the Fast Fourier Transform algorithm (FFT), the Discrete Fourier Transform (DFT) can be easily computed. Once obtained the frequency representation of a time series, it is possible to obtain some descriptors with spectral information.

Spectral centroid and Spread The spectral centroid and the spectral spread are two simple measures of spectral position and shape. The spectral centroid is the center of 'gravity' of the spectrum. Considering $X_i(k), k = 1, ..., W_{FL}$, as the magnitude of the DFT coefficients of the ith signal frame, the value of spectral centroid,C_i, of the ith audio frame is defined as

$$C_i = \frac{\sum_1^{W_{fl}} kX_i(k)}{\sum_1^{W_{fl}} X_i(k)} \tag{3.6}$$

Spectral spread is the second central moment of the spectrum. In order to compute it,one has to take the deviation of the spectrum from the spectral centroid, according to

$$S_i = \sqrt{\frac{\sum_1^{W_{fl}} (k - C_i)^2 X_i(k)}{\sum_1^{W_{fl}} X_i(k)}} \tag{3.7}$$

The spectral spread measures how the spectrum is distributed around its centroid. Unsurprisingly, in the case of stationary wave, there are no changes in the spectral centroid and spread, while for a variable frequency signal, these value of these descriptors change for each frame.

Spectral entropy Spectral entropy is computed in a similar manner to the entropy of energy, although, this time, the computation takes place in the frequency domain. More specifically, we first divide the spectrum of the short-term frame into L sub-bands. The energy E_f of the fth sub-band, $f = 0, ..., L - 1$, is then normalized by the total spectral energy. The entropy of the normalized spectral energy n_f is finally computed according to the equation

$$H(i) = -\sum_{f=0}^{L-1} n_f log_2(n_f) \tag{3.8}$$

Spectral flux Spectral flux measures the spectral change between two successive frames and is computed as the squared difference between the normalized magnitudes of the spectra of the two successive short-term windows

$$Fl_{(i,i-1)} = -\sum_{f=1}^{Wf_L} (EN_i(k) - EN_{i-1}(k))^2 \tag{3.9}$$

where $EN_i(k)$ is the kth normalized DFT coefficient at the ith frame.

Spectral rolloff This feature is defined as the frequency below which a certain percentage of the magnitude distribution of the spectrum is concentrated. Therefore, if the mth DFT coefficient corresponds to the spectral rolloff of the ith frame, then it satisfies the equation

$$\sum_{k=1}^{m} X_i(k) = C \sum_{k=1}^{Wf_L} X_i(k) \tag{3.10}$$

where C is the adopted percentage.

Mel-Frequency Cepstrum Coefficients MFCC contain filter banks that models the ability of human ear to resolve frequencies nonlinearly across the audio spectrum. Therefore, it is one of the most used descriptors in speech and audio recognition[cita]. In recent years, MFCC is being applied too in EEG tasks classification, achieving high accuracy values

MFCC are a kind of cepstral representation of the signal, where the frequency bands are distributed according to the mel-scale, instead of the linearly spaced approach. The DFT is computed and the resulting spectrum is given as input to a mel-scale filter bank that consists of L filters [83].

If O_k, $k = 1, ..., L$ is the power at the output of the kth filter, the resulting MFCC are given by

$$c_m = \sum_{k=1}^{L} (log(O_k)) cos\left[m\left(k - \frac{1}{2}\right)\frac{\pi}{L}\right] \tag{3.11}$$

$$mel(f) = 2595 log_{10}(1 + \frac{f}{700}) \tag{3.12}$$

3.2 Classification stage

There are various classifiers used in the BCI community, grouped in linear and nonlinear classifiers. Linear classifiers were the most commonly used in BCI until recently, due to their fewer parameters that need to be tuned in comparison to nonlinear classifiers. A nonlinear classifier it's on his part, based on nonlinear functions to separate the classes. The Motor Imagery classification it's a linearly non-separable problem, and therefore, either a good feature extraction stage must be achieved, or a non-linear classification stage must be implemented. In current BCI systems, the most popular classifiers are the support vector machines (SVM), Artificial neural networks (ANN), and nonlinear Bayesian classifiers. The ANN-based approach has been chosen for the present work.

Neural Networks was initially inspired by biological connections and synapses process in the brain. In the synapse, the dendrite from multiple cells connects to the axon of one cell, for the axon to be activated and send an action potential. McCullough and Pitts proposed the first model in 1943. In 1958 Rosenblatt introduced the model of the perceptron that was an early mathematical abstraction of biological neurons, and together with a dedicated optimization scheme, the perceptron learning algorithm it was the first instance of supervised learning for the artificial neural model.

3.2.1 Feedforward Neural Network

One of most popular ANN is known as feedforward neural network, that is a fully connected network in which all the input neurons are connected with all neurons in the first hidden layer, the outputs of these neurons are connected in turn with all the neurons in the next layer. In Figure 3.1 are shown an example of feed-forward neural network with three inputs x_1, x_2, x_3, three neurons in the hidden layer, and two neurons in the output layer. The basic equation of a single neuron in a perceptron is given by

$$y_N = f(\mathbf{W}\mathbf{X} + b) \tag{3.13}$$

where y_N is the output of the neuron, $\mathbf{X}$ is the vector of inputs, $\mathbf{W}$ is the weights parameters, b is the bias, and f is the activation function. Then, the output of each neuron is a linear combination of all their inputs, with an adjust of threshold given by the bias value, and finally the resultant

output is modified by the activation function that commonly is a nonlinear operator such as the $tanh(y_N)$.

The learning task is then the adjust of all weight matrices and bias vectors. In the case of supervised learning task, a set of examples labelled by a target vector are entered in the ANN. The output of the ANN is compared with the target vector to obtain the error, then, the error is propagated in the opposite direction for each epoch of the training.

One of most common performance function for feedforward networks is mean square error (MSE), the average squared error between the network outputs a and the target outputs defined as

$$MSE = \frac{1}{N}\sum_{i} = 1^{N}(e_i)^2 = \frac{1}{N}\sum_{i} = 1^{N}(t_i - a_i)^2 \tag{3.14}$$

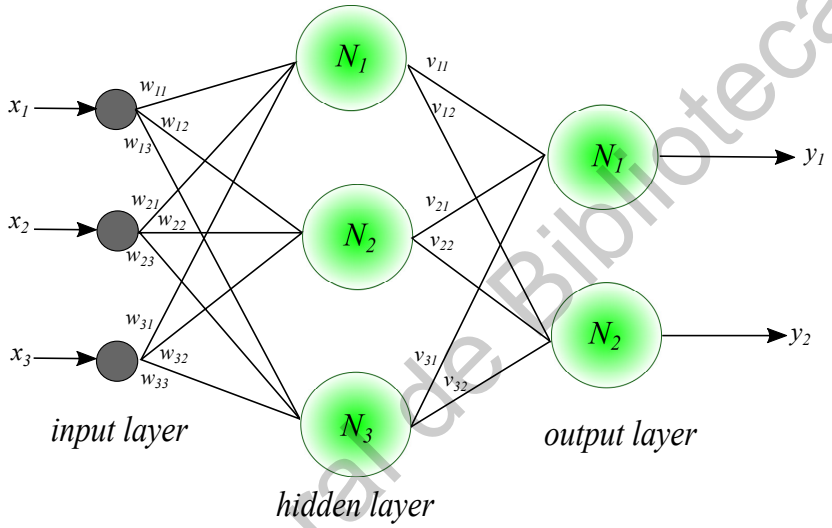


Figure 3.1: Multilayer Neural Network example

3.2.2 Convolutional Neural Network

Inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex, Convolutional Neural Network (CNN) are designed to recognize visual patterns directly from pixel images with minimal preprocessing [84]. CNN architecture is comprised of a sequence of convolutions and sub-sampling layers in which the content or values of the convolutional kernels is learned via an optimization algorithm. With a fixed number of filters in each layer, each individual filter is convoluted transversely with the width and height of the input figure in the forward transmits. The output of this layer is a two-dimensional feature map of that filter to detect the pattern. This is followed by a Rectified Linear Unit (ReLU) where the non-linearity is increased in the network using a rectified function. The governing equation of the convolution operation is given as

$$h_i^l = f(a) = f\left(\sum_{n=1}^{M} W_{ni}^l \cdot h_n^{l-1}, +b_i^l\right) \tag{3.15}$$

with the ReLU function is defined as

$$\mathrm{ReLU}(a) = \ln(1 - e^{a}), \tag{3.16}$$

where h_i^l is the i-th output of layer l, W_{ni}^l is the convolutional kernel that is operated on the n-th map of the $l-1$-th layer used for the i-th output of the l-th layer, and b_i^l is the bias term; $f(a)$ is an activation function imposed on the output of the convolution. The optimization algorithm focuses on optimizing the convolutional kernel W. The output of each convolutional operation or any operation is denoted as a feature map. After convolution, the sub-sampling layer reduces the dimension of the feature maps by representing a neighborhood of values in the feature map by a single value.
CNN are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics [85]. This particular feature makes CNN appropriate to deal with the problem described above with the loss of order in each trial processed by BSS algorithms. The parameters of each convolutional layer are:

- Filters: The number of output filters in the convolution.
- kernel size: The height and width of the 2D convolution window.
- Strides: The strides of the convolution along the height and width.

Another important operation applied usually in each convolutional layer is the Max pooling, that is a sample-based discretization process. The objective is to down-sample an input representation, reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned. This is done to in part to help over-fitting by providing an abstracted form of the representation. As well, it reduces the computational cost by reducing the number of parameters to learn and provides basic translation invariance to the internal representation [86].

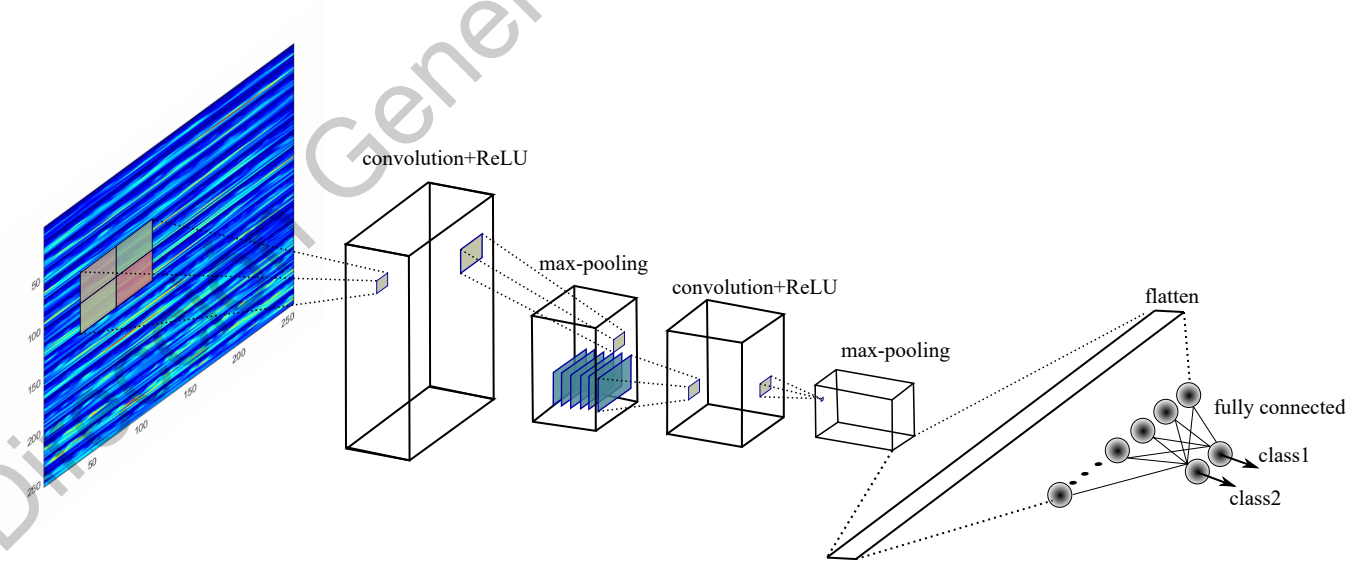


Figure 3.2: Convolutional Neural Networks proposed

CHAPTER 4

Methodology and results

In this chapter, two BSS algorithms, SOBIRO and fastICA were selected in order to compare their performance in classification accuracy for real EEG signals. Two classification approaches described in section 3.2 were used. For the case of MLP, the feature extraction stage is made using the time and frequency descriptors expanded in section 3.1. On the other hand, when CNN approach is used as classifier, the input channels are currently presented as time-frequency maps or time-scale representations. The convolutional layers of the CNN generate an abstract representation of the input images making a dimensional reduction. The experiments were conducted on an Intel Core(TM) i7-7700HQ 2.80GHz with 16GB of RAM. Matlab 2017 was used to compute fastICA [77] and CWT maps, while python Tensorflow and Keras libraries were used to compute the CNN architecture.

Dataset The dataset was provided by Intelligent Data Analysis Group is the IVa from the BCI competition III was used in current study. This consists of five healthy subjects sat in a comfortable chair with arms resting. The subjects are labeled as *aa*, *al*, *av*, *aw*, and *ay* [87]. Visual cues indicated for 3.5 seconds which of the following two MI the subject should perform: right hand (class1) and foot (class2) MI. The presentation of target cues were intermitted by periods of random length, 1.75 to 2.25 seconds, in which the subject could relax. [88]. In Figure 4.1 is shown a diagram where it is illustrated the extraction of MI segments inside a trial, where the interest segment of MI occurs in time interval between 4.5 and 8 seconds. The recording was made using BrainAmp amplifiers and a 128 channel Ag/AgCl electrode cap from ECI. 118 EEG channels were measured at positions of the extended international 10–20-system. Signals were digitized at 1000 Hz with 16 bit (0.1 μV) accuracy.

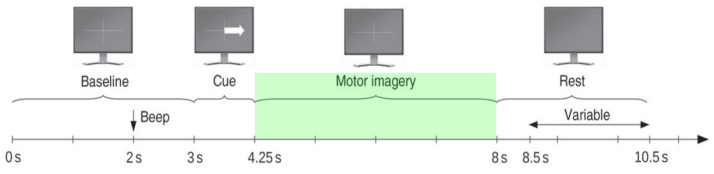


Figure 4.1: Extraction of Motor Imagery for each trial

According to the 10-20 sysyem, the electrodes associated with the sensorimotor regions in the motor

cortex are C_3 for the left hemisphere and C_4 for the right hemisphere. In this work, 18 channels were located in both hemispheres were selected. FC_5, FC_3, FC_1, C_5, C_3, C_1, CP_5, CP_3, CP_1 in left side, and FC_2, FC_4, FC_6, C_2, C_4, C_6, CP_2, CP_4,CP_6 in right side. The selected electrodes coincide with the regions reported in [73]. The two clusters *left* and *right* are filtered using a second-order IIR Butterworth, set as band-pass filter. To preserve the relevant information for MI, the selected cut-off frequencies were adjusted between 0.5 and 90 Hz.

4.1 Proposed method to sort the estimated sources

Due to its unsupervised and statistical nature, BSS does not require a priori information about MI classes, nor specific frequency bands, which is an advantage over other spatial filter approaches. Nonetheless, an inherent disadvantage of BSS algorithms is that for each processed trial, the order is not preserved, which limits its direct application in further classifier stages used in BCI, where the order of the input vectors must be conserved to avoid loss of the adjustment parameters for each new data entry.

One of the main contributions of this work is the criterion of sorting the estimated sources. A typical spectral profile of Movement Related Independent Components (MRIC) with significant components in the μ and β frequencies is used for sorting in each processed trial, thus ensuring that the sources estimated to be the most active in MI frequencies remain at the beginning of the array, while the artifacts are placed in the final positions.

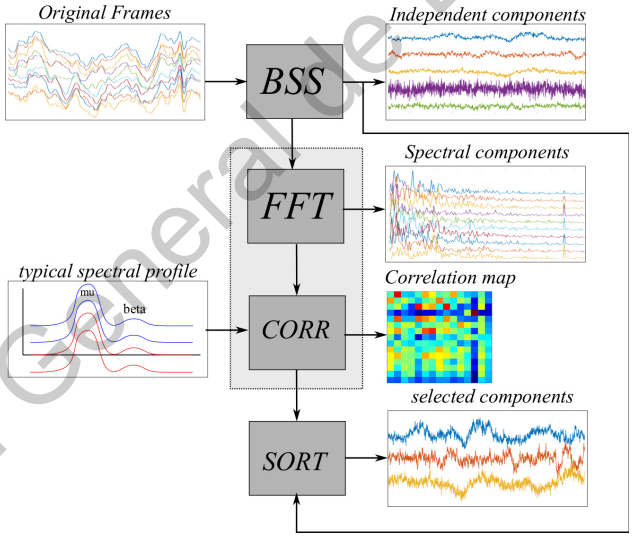


Figure 4.2: Proposed method. Original channels are passed through a BSS algorithm generating independent sources. Typical spectral profile are correlated with the spectral components of each trial to sort and select interest components.

In Figure 4.3 are shown the spectral components of estimated sources before and after apply the sort criterion in one trial. The components with more information in μ and β will generally be placed in the top positions, while the components least associated with these frequencies will be at the bottom as for example in Figure 4.3 (a), the first spectral component has more energy in frequencies over 20 Hz but without α and β contribution, is placed in the bottom in Figure 4.3 (b).

In contrast, the component located in the seventh position in Figure 4.3 (a), which has the most energy in the region μ, is located in the first position in Figure 4.3 (b).

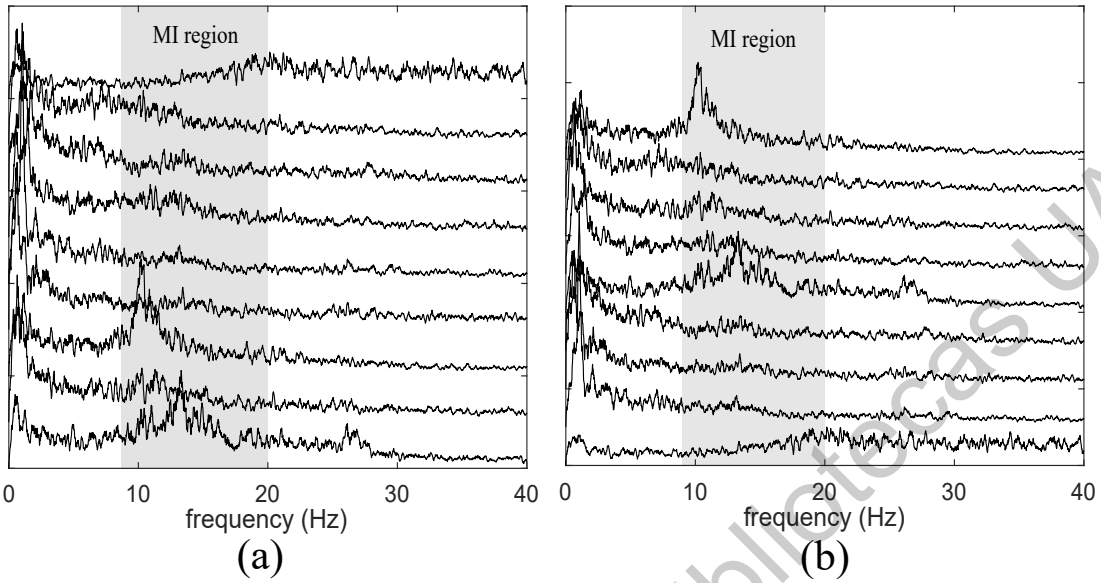


Figure 4.3: Analysis spectral components. (a) before sort; (b) after sort. The components with more MRIC frequencies are placed at the top after sorting.

4.2 Classification using descriptors and MLP

To compare the influence of BSS in the classification accuracy, three cases are studied, (a) NO BSS, (b) SOBIRO, and (c) fastICA. In feature extraction stage, the estimated sources generated in previous stage go through a series of descriptors in search of discriminant characteristics that allow to separate the classes. For this purpose, time-domain and frequency domain descriptors were used. To study the behaviour of descriptors according their nature, these have been separated into three main groups:

- Group 1 (time-domain descriptors): zero-crossing, energy, and entropy.
- Group 2 (frequency-domain descriptors): spectral centroid, spectral spread, spectral entyopy, spectral flux, and spectral flux.
- Group 3 (MFCC descriptors): 13 MFCC filters.

For classification of MI, a MLP with seven neurons in the hidden layer was used. To train the network, gradient descent with momentum and adaptive learning rate was employed. Inputs were divided in 60 % for training, 20 % for validation, and 20 % for test once the net has been trained.

Analysis without BSS preprocessing The case (a) NO BSS is used as a point of reference for comparing the influence of BSS separation. In this case, the three proposed groups of descriptors are evaluated separately in the MLP. In table 4.1 are the classification accuracy for each subject of

the dataset. The time it takes the entire system to process each trial is shown in the tables, once the network has been trained.

Table 4.1: accuracy without BSS dataset IVa

subject	Group 1	Group 2	Group 3	All	time for each trial (s)
IVa-aa	52.5	50.7	40.6	53.0	0.1190
IVa-al	50.6	66.7	38.7	47.6	0.1200
IVa-av	62.2	21.7	**91.5**	72.5	0.1222
IVa-aw	50.5	45.4	35.0	44.2	0.1194
IVa-ay	52.2	26.3	61.3	23.3	0.1219
mean	53.6	42.1	53.4	48.1	0.1204
std	3.9	15.3	19.3	14.4	0.0015

The mean of classification accuracies showed in table 4.1 are below 50 %, where only for one subject (aa) was achieved a value of 91.5 %. This result reveals the need for additional preprocessing before feature extraction and classification, or else, the choice of other kind of feature extraction descriptors. Whereas the focus of this paper is on the preprocessing stage, BSS algorithms are implemented, keeping the same descriptors.

Analysis using SOBIRO When SOBIRO is applied and features are extracted from all estimated sources, the classification percentages are improved in comparison with the case where BSS is not applied. However, these percentages are still below 80 %. In table 4.2 are showed the classification accuracy using SOBIRO in the preprocessing stage.

Table 4.2: classification accuracy with SOBIRO in BSS stage dataset IVa

subject	Group 1	Group 2	Group 3	All	time for each trial (s)
IVa-aa	67.0	82.3	65.9	78.4	0.2079
IVa-al	89.6	78.5	61.5	77.3	0.1607
IVa-av	55.9	66.7	70.1	85.9	0.1621
IVa-aw	72.1	83.2	75.3	81.3	0.1614
IVa-ay	56.4	80.7	72.5	78.5	0.1600
average	68.2	78.3	69.0	78.5	0.1704
std	13.8	6.7	5.4	6.0	0.0209

Analysis using fastICA For this case the classification accuracy in the output of the MLP is considerably improved, achieving a classification value of 90.8% when all descriptors are used. and even though the processing time of this HOS algorithm is about 8.5 times greater than the SOS SOBIRO algorithm, the window size for each trial is about three seconds, which is why fastICA can be implemented in real time applications. In table 4.3 are showed the classification accuracy using fastICA in the preprocessing stage.

Table 4.3: classification accuracy with fastICA in BSS stage dataset IVa

subject	Group 1	Group 2	Group 3	All	time for each trial(s)
IVa-aa	78.5	90.3	76.0	91.1	1.4750
IVa-al	78.8	88.7	96.5	90.7	1.4464
IVa-av	78.1	87.6	92.2	91.0	1.4536
IVa-aw	88.3	89.2	92.8	92.3	1.4464
IVa-ay	89.4	90.1	90.4	89.0	1.4500
average	82.6	89.2	89.6	**90.8**	1.4543
std	5.7	1.0	7.9	1.2	0.0120

4.3 Classification using CNN

A recent approach that has given excellent results, mainly in computer vision is deep learning [89]. However, deep learning techniques have not been widely used for EEG-BCI applications, due to factors such as as noise, the correlation between channels, and the high dimensional EEG data [90]. Some works where deep learning has been used for MI classification have been proposed [91, 92, 93, 94, 95, 96, 97, 90] where CNN are the main architecture used for these cases. However, for MI-BCI based paradigm the datasets are small due to the fatigue where the participants are exposed in each session. Therefore, it has been difficult to use deep learning for this purpose [95]. The natural and most common representation of time-series signals when CNN is used, is a time-frequency image for each of them. In this case, the Continuous Wavelet Transform (CWT) is used. The CWT generates 2D maps from 1D time-series containing information of time and scale, with a logarithmic relationship with the frequential components.

Wavelet Transform EEG registers are initially time series, from which it is possible to obtain information related to the temporal evolution of some characteristics. However, in this space, it is not possible to know frequential information. On the other hand, the Fourier transformation makes it possible to identify information in the frequency domain, but the temporal evolution of the frequency components is unknown. The CWT generates 2D maps from 1D time series containing information of time and scale, with a logarithmic relationship with the frequential components. Unlike the Short Time Fourier Transform (STFT), CWT performs a multiresolution analysis [98]. CWT is described by

$$W_x(a,\tau) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-\tau}{a}\right) dt, \tag{4.1}$$

where a is the scale factor, ψ is the mother wavelet, and τ is the shift time of the mother wavelet on the $x(t)$ signal.

In Figures 4.4 and 4.5 are depicted some CWT maps of representative sources with μ and β frequencies, obtained with SOBIRO and fastICA respectively.

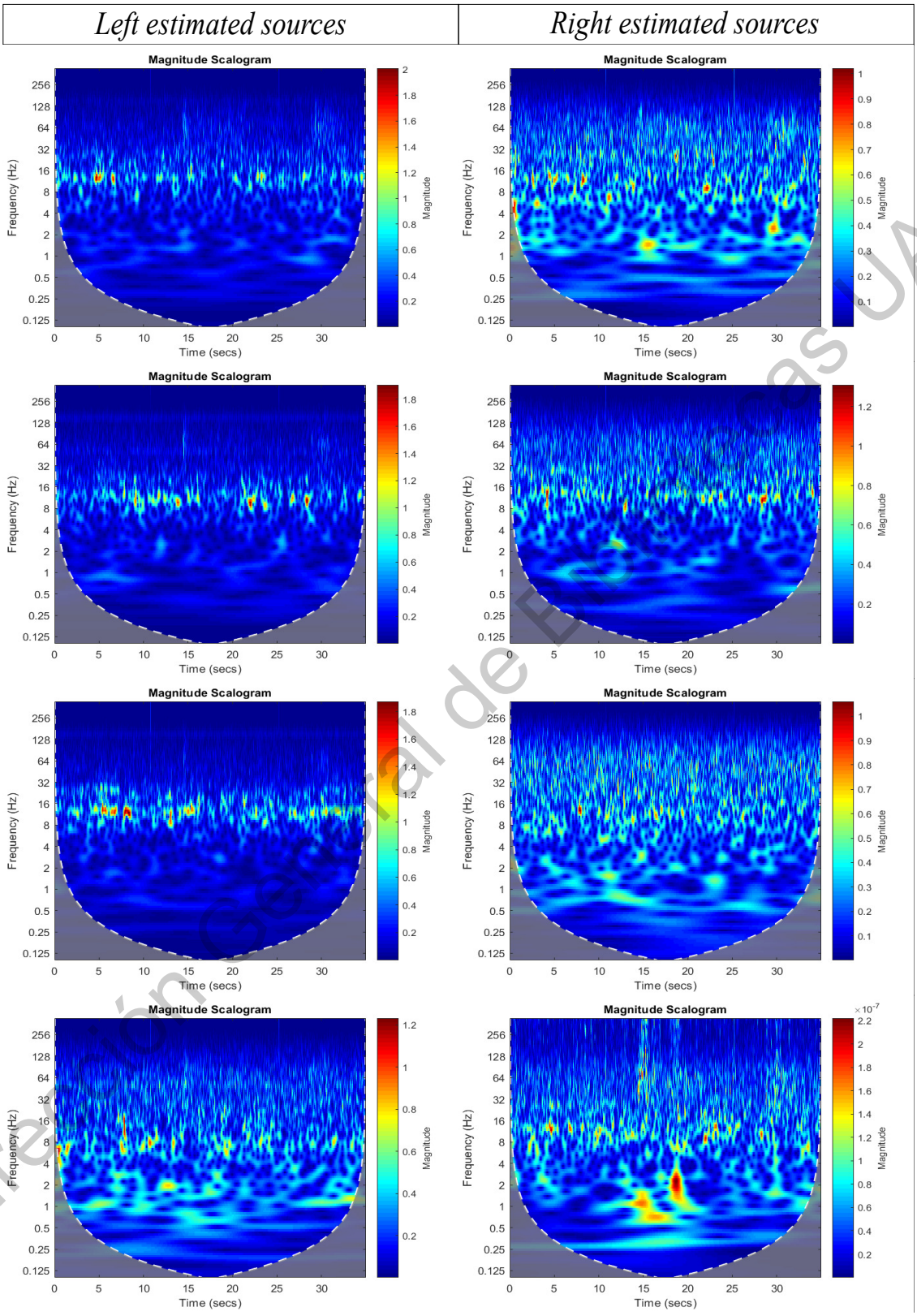


Figure 4.4: CWT of selected MRIC from SOBIRO algorithm.

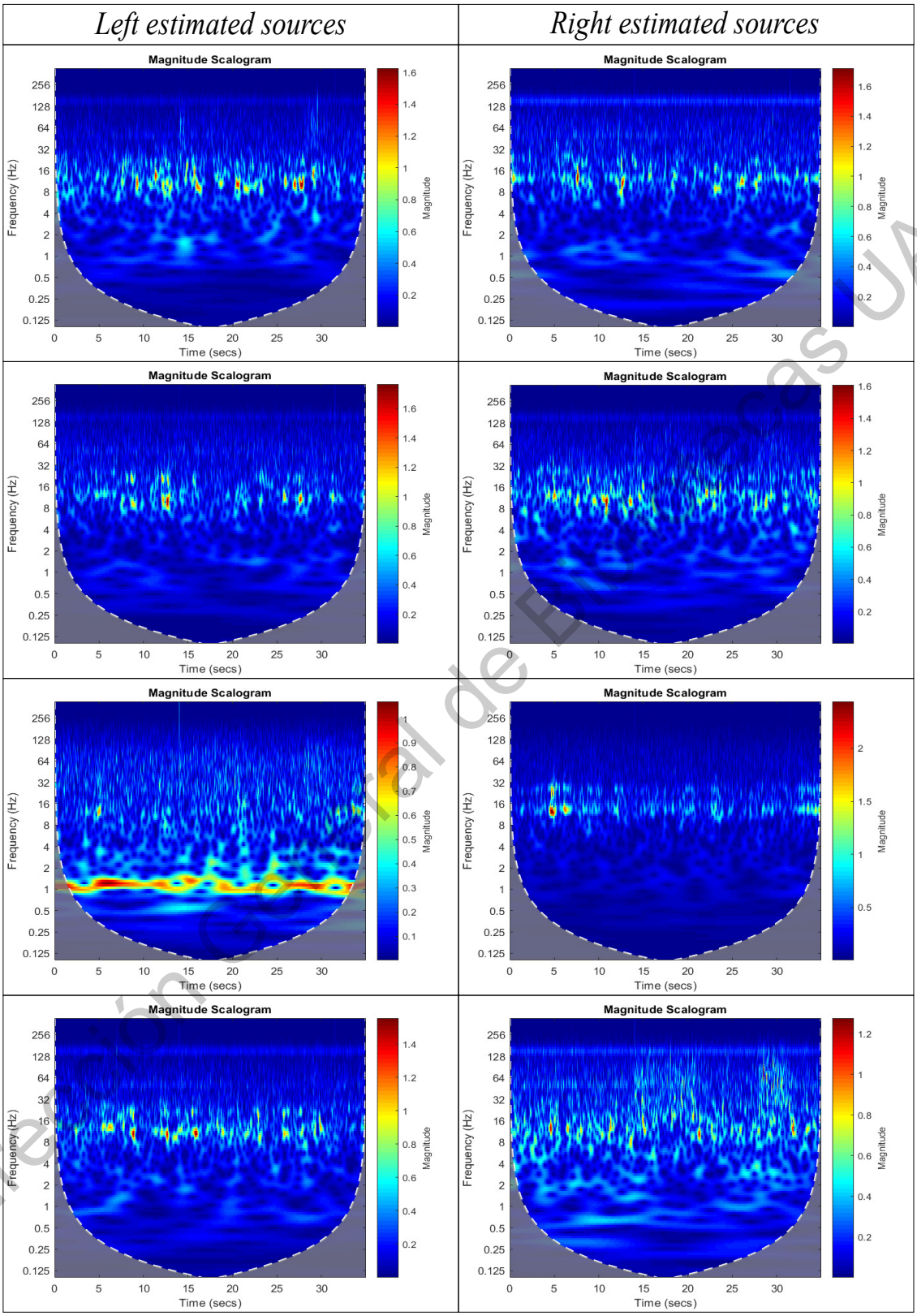


Figure 4.5: CWT of selected MRIC from fastICA algorithm.

Description of proposed approach is depicted in Figure 4.6. Clusters $left$ and $right$ are separated in $trials$, from $trial_1$ to $trial_N$. Each trial is preprocessed using a BSS algorithm, which generates equal number of estimated sources $\mathbf{s}'(t)$ from the input channels $\mathbf{x}(t)$. These sources were sorted using as criterion the correlation between their spectral components and the MRIC. This procedure helps to separate the sources and the unwanted artifacts that have low correlations with MRIC. Sorted trials of $\mathbf{s}'(t)$ are passed through a CWT block. The CWT is obtained using generalized Morse wavelets. Analytic wavelets are complex-valued wavelets whose Fourier transforms are supported only on the positive real axis. They are useful for analyzing modulated signals, which are signals with time-varying amplitude and frequency [99]. The window size of each CWT is 1.0 seconds, each window is computed using steps of 0.25 seconds. CNN architecture has as input the CWT figures and finally a fully connected Multilayer Perceptron (MLP) separates into two classes, *right hand MI* and *right foot MI*.

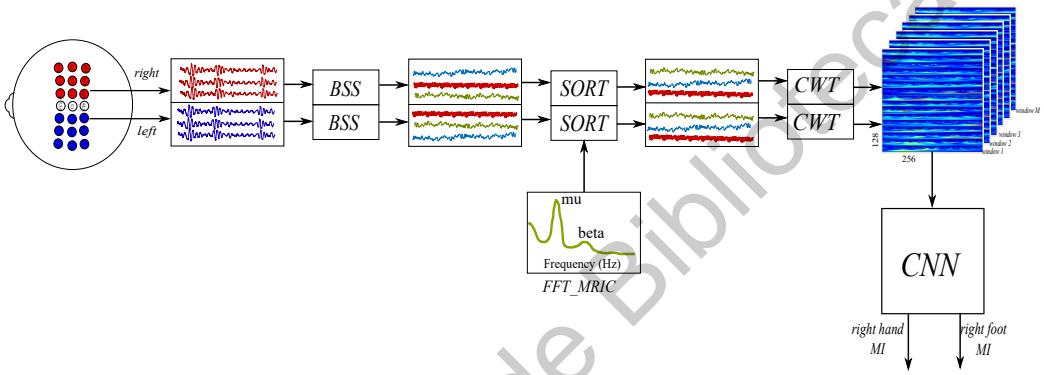


Figure 4.6: Proposed methodology. The left and right channels are prepossessed using a BSS algorithm, the MRIC sorts the estimated sources, in the CWT stage the images for each time window are obtained, finally the CNN separates the classes.

In Figure 4.7 a) is shown the CWT of a single estimated source $\mathbf{s}'_1$, in b) is shown an example of input containing all estimated sources stacked along y axis. Each figure is re-scaled to a size of 128×256 .

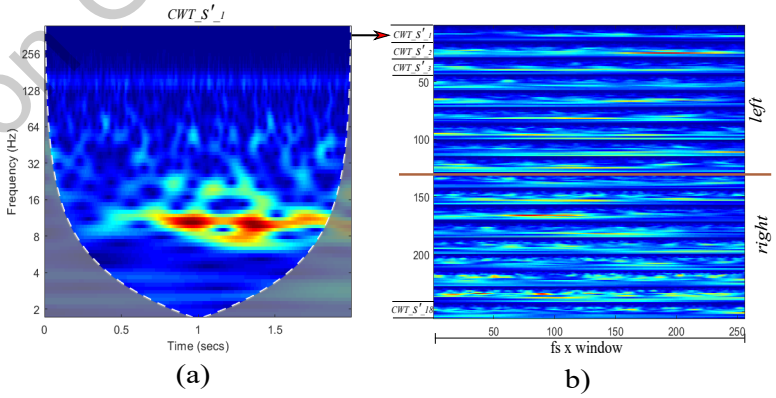


Figure 4.7: CWT maps for (a) one estimated source, (b) CWT stacked maps for left and right estimated sources.

To evaluate the proposed methodology, SOBIRO, and fastICA were used again. In Figure 4.8 is showed the train and validation values for 20 epochs in cases (a) without BSS preprocessing, (b) with fastICA without the sort criterion, (c) with sorted SOBIRO, and (d) with sorted fastICA. The CNN architecture initially used was the same proposed in [1]. Nevertheless, the input data is organized in a different way, for which it was necessary to make some adjustments in the convolution stride and max-pooling size. The CNN architecture used in this work is showed in Table 4.4.

Table 4.4: CNN modified architecture inspired from [1]

Layer	Operation	Kernel	Stride	Output shape
1	Conv2D $\rightarrow$ 250	$(3,1)$	$(2,2)$	(63,256,250)
	Activation$\rightarrow$ $ReLU$			(63,256,250)
	Max-pooling$\rightarrow$ $(4,4)$			(15,64,250)
2	Conv2D $\rightarrow$ 150	$(1,2)$	$(1,1)$	(15,63,150)
	Activation$\rightarrow$ $ReLU$			(15,63,150)
	Max-pooling$\rightarrow$ $(3,3)$			(5,21,150)
3	Flatten			(15750)
	Dense$\rightarrow$ 2048			(2048)
	Activation$\rightarrow$ $ReLU$			(2048)
	Dropout$\rightarrow$ 0.4			(2048)
4	Dense$\rightarrow$ 2			(2)
	Activation$\rightarrow$ $softmax$			(2)

Analysis without BSS preprocessing in Figure 4.8 (a), the maximum accuracy value for train is near 0.8, while the validation accuracy is below 0.6. These results can be explained taking into account the reduced number of data for a deep learning approach, where large amounts of data are required to achieve an end-to-end system, where convolutional layers are able to find the determining patterns that allow for classifying movement intentions.

Analysis using SOBIRO In in Figure 4.8 (c), SOBIRO as BSS algorithm and sorted with the same explained criterion in Figure 4.6, the test accuracy achieved 0.73 values, improving the a) and b) responses. However, differences between train and test are considerably large which indicates overfitting.

Analysis using fastICA in Figure 4.8 (b), where fastICA is applied in each trial but without the sort step, the training accuracy reaches values close to 0.98, but the validation accuracy is below 0.60 and decrease for each epoch. This can be explained by the disorder of the estimated sources in each trial, where CNN learns training set, but fails to generalize in the test set. This result validates the hypothesis of the need to use a sort criterion for sources estimated through BSS. Finally, in Figure 4.8 (d), with the sorted HOS fastICA, both the training data and the validation data achieve values higher than 0.8, reducing the phenomenon of overfitting. This result validates several previous works where the superiority of the HOS algorithms over the SOS for BSS

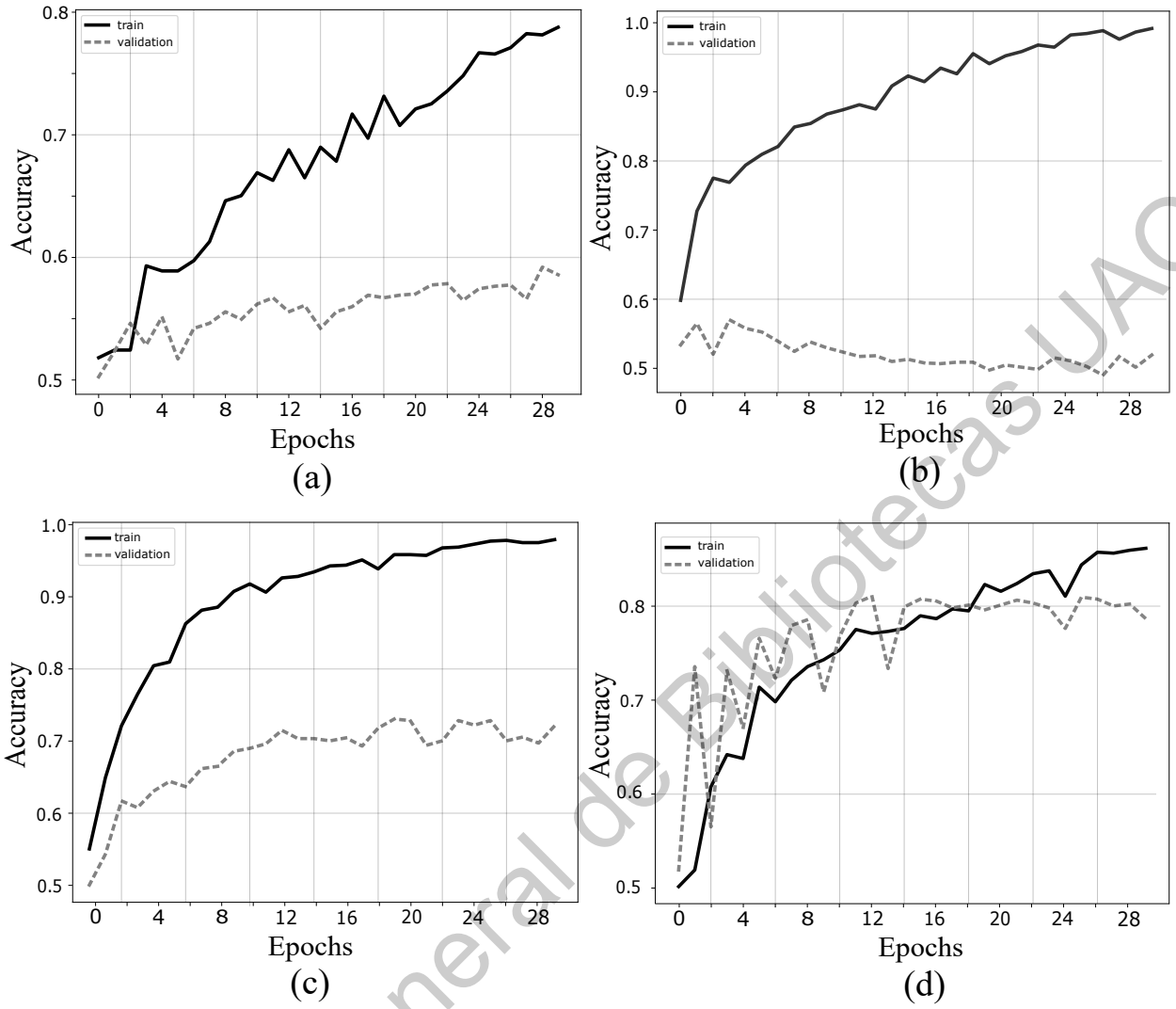


Figure 4.8: Train and test validation behaviour for subject *aa* in cases. (a) without BSS, (b) with no sorted fastICA, (c) with sorted SOBIRO, and (d) with sorted fastICA (30 epochs). with the sorted HOS fastICA, both the training data and the validation data achieve values higher than 0.8, reducing the phenomenon of overfitting in comparison to the other cases.

is reported. Results (c) and (d) are in accordance with numerous works reporting superiority of HOS-BSS algorithms over SOS-BSS algorithms for EEG preprocessing [8, 100, 101]. Sorted fastICA is then chosen as BSS algorithm before CWT generation and posterior CNN classification stages.

4.3.1 Comparison with other methods

In Table 4.5 are shown the validation accuracies for each k validation and each subject of dataset IVa of BCI competition III. The test set was divided into 10 equal parts for each cross validation. The maximum classification value was chosen in each case.

Table 4.5: 10-fold cross validation accuracy.

Subject	**Accuracy**											
	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	Average	Std
subject aa	94.79	100.00	96.87	89.58	100.00	100.00	100.00	98.95	98.95	98.95	97.81	3.34
subject al	91.66	94.79	94.79	98.95	85.41	87.50	97.91	100.00	98.95	94.79	94.47	4.96
subject av	95.83	97.91	100.00	98.95	97.91	88.54	68.75	100.00	100.00	100.00	94.78	9.79
subject aw	98.75	92.50	95.00	99.37	99.37	91.87	100.00	98.75	76.87	88.12	94.06	7.26
subject ay	85.41	97.91	85.41	79.16	96.87	95.83	95.83	100.00	96.87	88.54	92.18	6.98
Average											**94.66**	6.46

As is shown in Table 4.5, the average ranking percentage of 94.66%, with a standard deviation σ of 6.46. For the five subjects, the maximum k-fold accuracy average was 97.81% with σ of 3.34 in subject *aa*, and a minimum value of 92.18 with σ of 6.98 in subject *ay*. Table 4.6 shows some recent work that reports the same used dataset [54, 102, 68, 71].

Table 4.6: Comparison with other works using the IVa of BCI competition III dataset.

Author	**Method**	**Classifier**	**Accuracy** (%)	**Year**
Lu et al.	R-CSP with aggregation	R-CSP	83.90	2010
Siuly et al.	CT	LS-SVM	88.32	2011
Zhang et al.	Z-score	LDA	81.10	2013
Siuly et al.	OA	NB	96.36	2016
Kevric et al.	MSPCA, WPD, HOS	k-NN	92.80	2017
Taran et al.	TQWT	LS-SVM	96.89	2018
Proposed	sorted-fastICA-CWT	CNN	**94.21**	2019

4.3.2 Analysis of kernel size influence

The other distinctive part of proposed method is the adjustment of some relevant hyper-parameters. Taking as initial values those shown in Table 4.4, and changing the kernel size along height and width for each convolutional layer, the behaviour of validation accuracy for each case is analyzed. The kernel size is (y,x), where y-axis contain frequential and spatial information, while x-axis contain temporal information. In Figure 4.9 are depicted the validation accuracy for subject *aa*. The kernel sizes selected to the comparison in the first convolutional layer were $(i, 1)$, $(i, 3)$, $(i, 5)$, $(i, 7)$, with i taking odd values from 1 to 9.

According to the analysis of kernel size in the first convolutional layer, the size (7,5) (Figure 4.9 (c)) presented less overfitting and major classification accuracy. It is also noted that when the kernel has a size of in the size 3 for y-axis, while reaching a maximum value close to the best case, this kernel size generates a high overfitting after certain times. On the other hand, the the size 1 for y-axis generates the lowest maximum values in all combinations of x-axis.

These results are in accordance with other work where the size of the kernel is also studied, where they report that the vertical locations (frequency-space) is of great importance for the classification performance, while, in contrast, the horizontal locations (time) are not as significant [90].

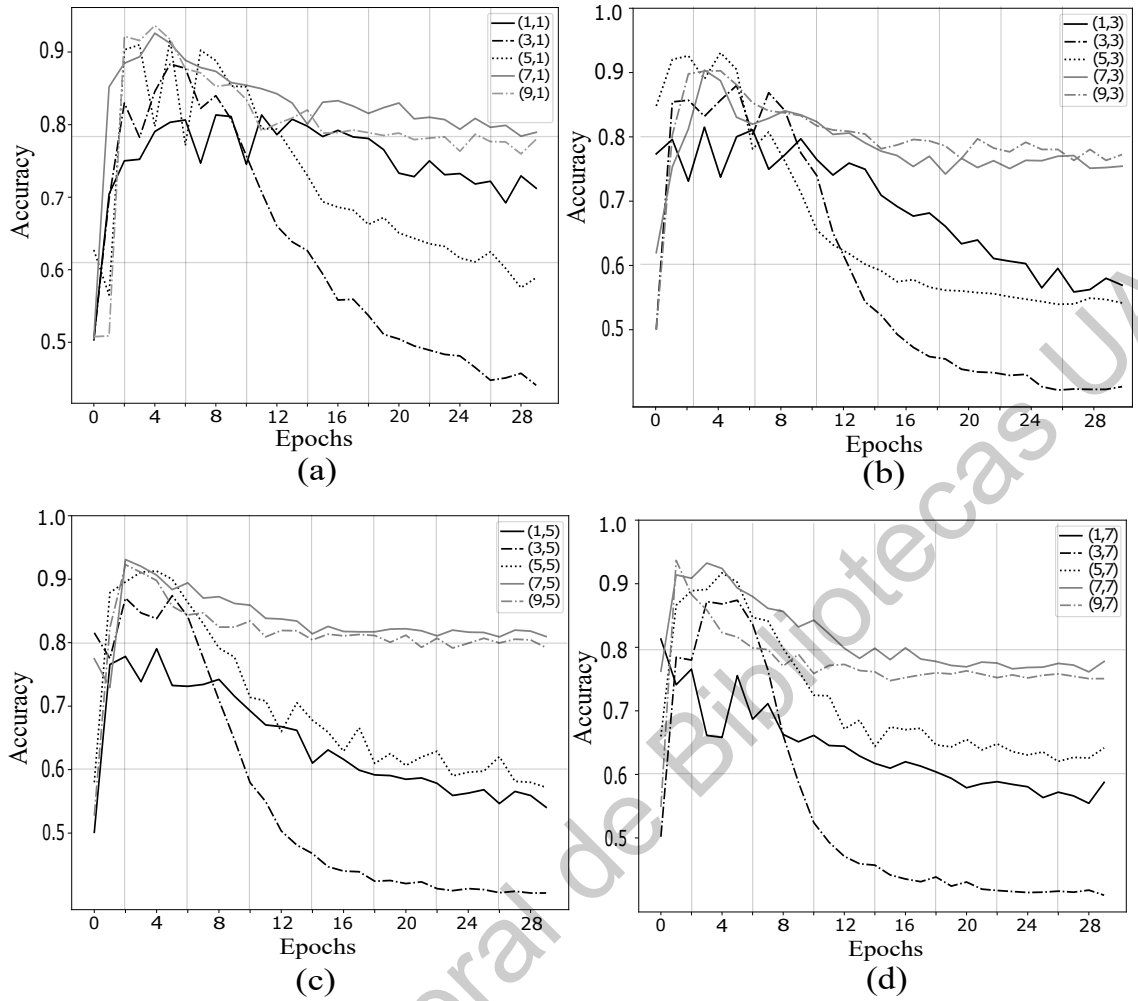


Figure 4.9: Analysis of kernel size in first convolutional layer. (a) (i,1), (b) (i,3), (c) (i,5), and (d) (i,7). The kernel size (7,5) in (c) presented less overfitting and major classification accuracy.

Once fixed the kernel size in first layer, a similar analysis was made for the kernel of the second convolutional layer. In Figure 4.10 are depicted the validation accuracy for subject *aa*. Taking into account that a (4,4) max-pooling has been previously applied, the kernel sizes selected were (j,1), (j,2), (j,3), and (j,4), with j taking values from 1 to 5.

At this case, the only y-axis where the CNN can achieve a stable accuracy throughout the epochs is 1 independently of x-axis size. In other cases, the validation accuracy decreases to 60% or even around 50% in case y-axis=5.

As is well known, currently deep learning approaches are state-of-the-art in many images processing and artificial vision. However, in contrast with two-dimensional static images, the EEG signals are dynamic time series, where the generalizable MI patterns in EEG signals are spatially distributed and mixed in the channels around the motor region. In addition to this, low signal-to-noise ratio could make learning features in an end-to-end fashion more difficult for EEG signals than for images [92]. On the other hand, deep learning approaches require a large amount of training data in order to obtain descriptors that allow discrimination between different classes. In particular case of

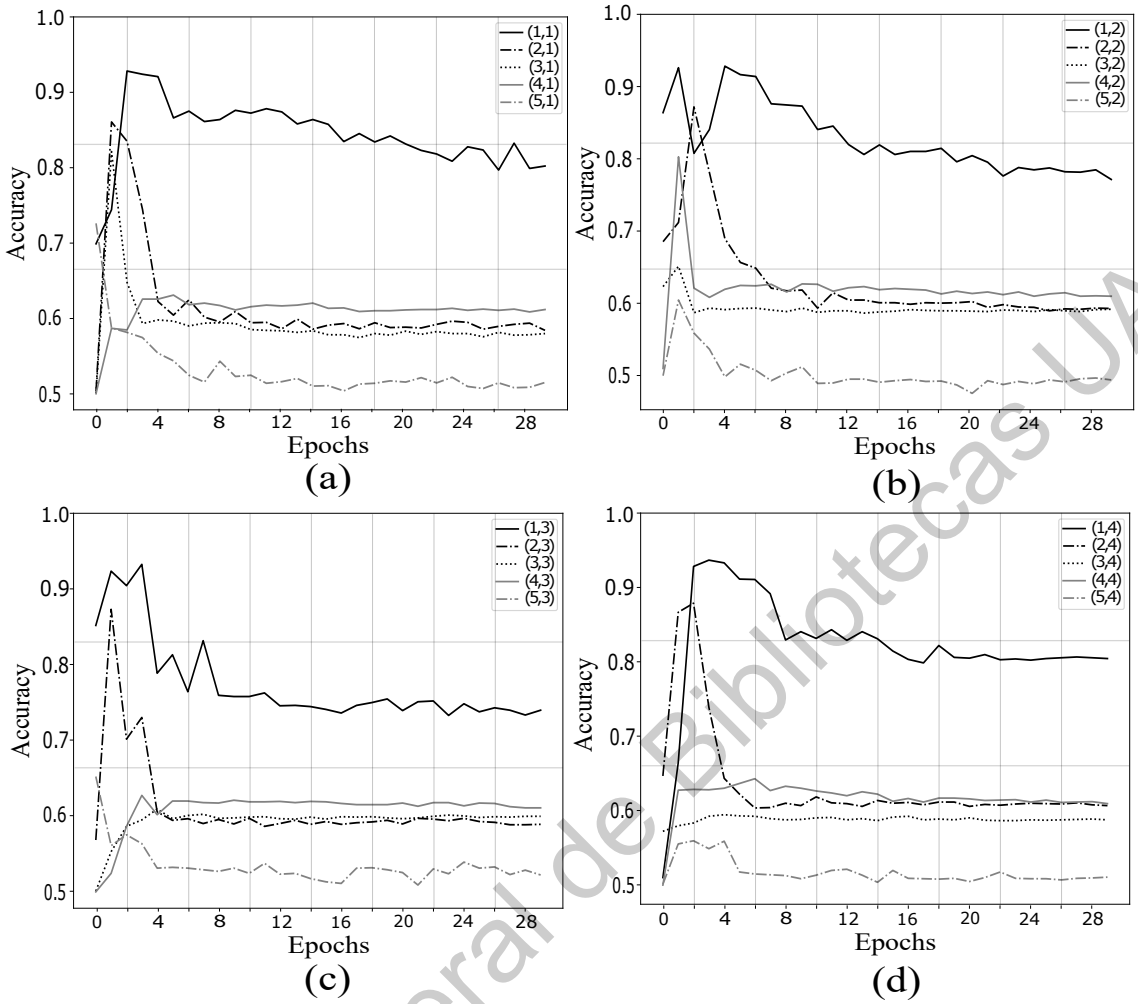


Figure 4.10: Analysis of kernel size in second convolutional layer. (a) (j,1), (b) (j,2), (c) (j,3), and (d) (j,4). The kernel size (1,1) in (a) presented less overfitting and major classification accuracy. The only y-axis where the CNN can achieve a stable accuracy throughout the epochs is 1.

EEG and MI, this is a limitation since the data must be processed independently for each subject, and due to fatigue, the MI databases are relatively small. To deal with this problem, some works proposed using deep learning have done data augmentation using some criteria to generate simulated data from the training set. This approach has yielded good results. However, the generation of artificial data can be risky without a rigorous methodology and thus generate false data that increases accuracy for a particular dataset.

CHAPTER 5

Conclusions and future works

In the first part of this dissertation, different SOS and HOS algorithms for BSS (SOBI, SOBIRO, infoMax, and fastICA) were evaluated to separate semi-synthetic signals. From these experiments, it can be concluded that the ability to separate the sources decreases significantly in both SOS and HOS approaches according to the amount of noise present in the mixture. Considering that EEG signals are a mixture with a meager signal-to-noise ratio, it is necessary to use machine learning approaches that have robust properties against noisy inputs.
The second part of the present dissertation was precisely focused on the classification stage, approaching this part from two perspectives, one based on extracting characteristics through descriptors in time and frequency domain and other based on convolutional layers. Two representative BSS algorithms were used in the preprocessing stage, SOS-SOBIRO, and HOS-fastICA. The obtained results showed better performance in classification accuracy for HOS-BSS algorithm for both classifiers MLP, and CNN.
One of the main contributions of this work is the criterion of sorting the estimated sources based on a spectral profile of Movement Related Independent Components with significant components in the μ and β frequencies. This approach helps to leave in the last positions the sources with a more substantial influence of artifacts and less μ and β components, reducing the complexity in the search for relevant patterns in the posterior extraction and classification stages. The use of CWT maps in the feature extraction stage allows obtaining a 2D representation of time series. In contrast with Short-Time Fourier Transform (STFT), CWT performs a multi-resolution analysis. According to the experimentation carried out, obtained results in this work are competitive with the state-of-the-art with a 94.21% in the k-fold cross-validation.

5.1 Future works

Regarding the architecture of CNN, it was found that the hyper-parameters related to the size of the kernel as well as the kernel stride in each convolutional layer have a significant influence on network performance, while the number of convolutions has less impact in final accuracy. Two future works derived: first, the development of a methodology that allows us to find the hyper-parameters close to the optimum and then, improve the current results. Second, to replace the BSS stage with some autoencoder architecture, for example, Variational Autoencoder (VAE), to obtain the estimated sources.

Bibliography

[1] Z. Zhang, F. Duan, J. Solé-Casals, J. Dinarès-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019.

[2] S. Sanei and J. A. Chambers, *EEG signal processing*. Wiley Online Library, 2007.

[3] V. Gandhi, *Brain-Computer Interfacing for Assistive Robotics*. Elsevier, 2015.

[4] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[5] C. J. James and C. W. Hesse, "Independent component analysis for biomedical signals," *Physiological measurement*, vol. 26, no. 1, p. R15, 2004.

[6] S. Hu, M. Stead, and G. A. Worrell, "Automatic identification and removal of scalp reference signal for intracranial eegs based on independent component analysis," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 9, pp. 1560–1572, 2007.

[7] R. R. Vázquez, H. Velez-Perez, R. Ranta, V. L. Dorr, D. Maquin, and L. Maillard, "Blind source separation, wavelet denoising and discriminant analysis for eeg artefacts and noise cancelling," *Biomedical Signal Processing and Control*, vol. 7, no. 4, pp. 389–400, 2012.

[8] N. Oosugi, K. Kitajo, N. Hasegawa, Y. Nagasaka, K. Okanoya, and N. Fujii, "A new method for quantifying the performance of eeg blind source separation algorithms by referencing a simultaneously recorded ecog signal," *Neural Networks*, vol. 93, pp. 1–6, 2017.

[9] M. Crespo-Garcia, M. Atienza, and J. L. Cantero, "Muscle artifact removal from human sleep eeg by using independent component analysis," *Annals of biomedical engineering*, vol. 36, no. 3, pp. 467–475, 2008.

[10] S. Romero, M. A. Mañanas, and M. J. Barbanoj, "A comparative study of automatic techniques for ocular artifact reduction in spontaneous eeg signals based on clinical target variables: a simulation case," *Computers in biology and medicine*, vol. 38, no. 3, pp. 348–360, 2008.

[11] "Ficha temática personas con discapacidad." `https:https://www.conapred.org.mx/userfiles/files`. Accessed: 2019-08-10.

[12] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 4, no. 2, p. R1, 2007.

[13] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for eeg-based communication," *Electroencephalography and clinical Neurophysiology*, vol. 103, no. 3, pp. 386–394, 1997.

[14] M. Essl and P. Rappelsberger, "Eeg cohererence and reference signals: experimental results and mathematical explanations," *Medical and Biological Engineering and Computing*, vol. 36, no. 4, pp. 399–406, 1998.

[15] T. Mima and M. Hallett, "Electroencephalographic analysis of cortico-muscular coherence: reference effect, volume conduction and generator mechanism," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1892–1899, 1999.

[16] K. A. Ludwig, R. M. Miriani, N. B. Langhals, M. D. Joseph, D. J. Anderson, and D. R. Kipke, "Using a common average reference to improve cortical neuron recordings from microelectrode arrays," *Journal of neurophysiology*, vol. 101, no. 3, pp. 1679–1689, 2009.

[17] U. R. Acharya, S. V. Sree, A. P. C. Alvin, and J. S. Suri, "Use of principal component analysis for automatic classification of epileptic eeg activities in wavelet framework," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9072–9078, 2012.

[18] R. Kottaimalai, M. P. Rajasekaran, V. Selvam, and B. Kannapiran, "Eeg signal classification using principal component analysis with neural network in brain computer interface applications," in *2013 IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICECCN)*, pp. 227–231, IEEE, 2013.

[19] K. Polat and S. Güneş, "Artificial immune recognition system with fuzzy resource allocation mechanism classifier, principal component analysis and fft method based new hybrid automated identification system for classification of eeg signals," *Expert Systems with Applications*, vol. 34, no. 3, pp. 2039–2048, 2008.

[20] T. D. Lagerlund, F. W. Sharbrough, and N. E. Busacker, "Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition," *Journal of clinical neurophysiology*, vol. 14, no. 1, pp. 73–82, 1997.

[21] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, vol. 2014, 2014.

[22] R. Vigário, J. Sarela, V. Jousmiki, M. Hamalainen, and E. Oja, "Independent component approach to the analysis of eeg and meg recordings," *IEEE transactions on biomedical engineering*, vol. 47, no. 5, pp. 589–593, 2000.

[23] R. N. Vigário, "Extraction of ocular artefacts from eeg using independent component analysis," *Electroencephalography and clinical neurophysiology*, vol. 103, no. 3, pp. 395–404, 1997.

[24] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis," *Neuroimage*, vol. 34, no. 4, pp. 1443–1449, 2007.

[25] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in neural information processing systems*, pp. 145–151, 1996.

[26] N. P. Castellanos and V. A. Makarov, "Recovering eeg brain signals: artifact suppression with wavelet enhanced independent component analysis," *Journal of neuroscience methods*, vol. 158, no. 2, pp. 300–312, 2006.

[27] J. Iriarte, E. Urrestarazu, M. Valencia, M. Alegre, A. Malanda, C. Viteri, and J. Artieda, "Independent component analysis as a tool to eliminate artifacts in eeg: a quantitative study," *Journal of clinical neurophysiology*, vol. 20, no. 4, pp. 249–257, 2003.

[28] M. P. Tarvainen, J. K. Hiltunen, P. O. Ranta-aho, and P. A. Karjalainen, "Estimation of nonstationary eeg with kalman smoother approach: an application to event-related synchronization (ers)," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 3, pp. 516–524, 2004.

[29] M. Rohál'ová, P. Sykacek, M. Koskaand, and G. Dorffner, "Detection of the eeg artifacts by the means of the (extended) kalman filter," *Meas. Sci. Rev*, vol. 1, no. 1, pp. 59–62, 2001.

[30] E. Veslin, M. Dutra, L. Bevilacqua, L. Raptopoulos, W. Andrade, and J. Soares, "Decoding elbow movement with kalman filter using non-invasive eeg," in *2019 IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI)*, pp. 1–6, IEEE, 2019.

[31] B. H. Jansen, J. R. Bourne, and J. W. Ward, "Autoregressive estimation of short segment spectra for computerized eeg analysis," *IEEE Transactions on Biomedical Engineering*, no. 9, pp. 630–638, 1981.

[32] A. Gupta, V. Bhateja, A. Mishra, and A. Mishra, "Autoregressive modeling-based feature extraction of eeg/eog signals," in *Information and Communication Technology for Intelligent Systems*, pp. 731–739, Springer, 2019.

[33] A. Malafeev, X. Omlin, A. Wierzbicka, A. Wichniak, W. Jernajczyk, R. Riener, and P. Achermann, "Automatic artefact detection in single-channel sleep eeg recordings," *Journal of sleep research*, vol. 28, no. 2, p. e12679, 2019.

[34] B. Hjorth, "Eeg analysis based on time domain properties," *Electroencephalography and clinical neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.

[35] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, "Time domain parameters as a feature for eeg-based brain–computer interfaces," *Neural Networks*, vol. 22, no. 9, pp. 1313–1319, 2009.

[36] V. Srinivasan, C. Eswaran, Sriraam, and N, "Artificial neural network based epileptic detection using time-domain and frequency-domain features," *Journal of Medical Systems*, vol. 29, no. 6, pp. 647–660, 2005.

[37] J. F. D. Saa and M. S. Gutierrez, "Eeg signal classification using power spectral features and linear discriminant analysis: A brain computer interface application," in *Eighth Latin American and Caribbean Conference for Engineering and Technology*, pp. 1–7, LACCEI Arequipa, 2010.

[38] Q. Wei, M. Xiao, and Z. Lu, "A comparative study of canonical correlation analysis and power spectral density analysis for ssvep detection," in *2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics*, vol. 2, pp. 7–10, IEEE, 2011.

[39] C. Kim, J. Sun, D. Liu, Q. Wang, and S. Paek, "An effective feature extraction method by power spectral density of eeg signal for 2-class motor imagery-based bci," *Medical & biological engineering & computing*, vol. 56, no. 9, pp. 1645–1658, 2018.

[40] K. C. Veluvolu, Y. Wang, and S. S. Kavuri, "Adaptive estimation of eeg-rhythms for optimal band identification in bci," *Journal of Neuroscience Methods*, vol. 203, no. 1, pp. 163–172, 2012.

[41] N. P. Castellanos and V. A. Makarov, "Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis," *Journal of Neuroscience Methods*, vol. 158, no. 2, pp. 300–312, 2006.

[42] A. Klein, T. Sauer, A. Jedynak, and W. Skrandies, "Conventional and wavelet coherence applied to sensory-evoked electrical brain activity," *IEEE transactions on biomedical engineering*, vol. 53, no. 2, pp. 266–272, 2006.

[43] Q. Xu, H. Zhou, Y. Wang, and J. Huang, "Fuzzy support vector machine for classification of eeg signals using wavelet-based features," *Medical engineering & physics*, vol. 31, no. 7, pp. 858–865, 2009.

[44] S. Kumar, A. Sharma, and T. Tsunoda, "Subject-specific-frequency-band for motor imagery eeg signal recognition based on common spatial spectral pattern," in *Pacific Rim International Conference on Artificial Intelligence*, pp. 712–722, Springer, 2019.

[45] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcsp) in brain-computer interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2390–2397, IEEE, 2008.

[46] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms," *IEEE Transactions on biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2010.

[47] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 5392–5395, IEEE, 2006.

[48] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (sbcsp) for brain-computer interface," in *2007 3rd International IEEE/EMBS Conference on Neural Engineering*, pp. 204–207, IEEE, 2007.

[49] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1991–2000, 2008.

[50] S. Kumar, A. Sharma, and T. Tsunoda, "An improved discriminative filter bank selection approach for motor imagery eeg signal classification using mutual information," *BMC bioinformatics*, vol. 18, no. 16, p. 545, 2017.

[51] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier, and M. Pregenzer, "Current trends in graz brain-computer interface (bci) research," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 2, pp. 216–219, 2000.

[52] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Aggregation of sparse linear discriminant analyses for event-related potential classification in brain-computer interface," *International journal of neural systems*, vol. 24, no. 01, p. 1450003, 2014.

[53] W. Zhou, Y. Liu, Q. Yuan, and X. Li, "Epileptic seizure detection using lacunarity and bayesian linear discriminant analysis in intracranial eeg," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 12, pp. 3375–3381, 2013.

[54] S. Siuly and Y. Li, "Improving the separability of motor imagery eeg signals using a cross correlation-based least square support vector machine for brain–computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 4, pp. 526–538, 2012.

[55] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "Bci competition 2003-data set iib: support vector machines for the p300 speller paradigm," *IEEE Transactions on biomedical Engineering*, vol. 51, no. 6, pp. 1073–1076, 2004.

[56] T. N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Scholkopf, "Support vector channel selection in bci," *IEEE transactions on biomedical engineering*, vol. 51, no. 6, pp. 1003–1010, 2004.

[57] R. Singla and B. Haseena, "Bci based wheelchair control using steady state visual evoked potentials and support vector machines," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, no. 3, pp. 46–52, 2013.

[58] J. Liu, L. Zhang, C. Li, and Z. Xiao, "A semi-supervised support vector machine classification method based on parameter optimization for a motor imagery based bci system," in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 457–462, IEEE, 2019.

[59] M. A. Rahman, M. M. Haque, A. Anjum, M. N. Mollah, and M. Ahmad, "Classification of motor imagery events from prefrontal hemodynamics for bci application," in *Proceedings of International Joint Conference on Computational Intelligence*, pp. 11–23, Springer, 2020.

[60] M. A. M. Joadder, J. J. Myszewski, M. H. Rahman, and I. Wang, "A performance based feature selection technique for subject independent mi based bci," *Health information science and systems*, vol. 7, no. 1, p. 15, 2019.

[61] E. G. M. Kanaga, J. Sthuti, C. B. Sharon, and J. K. Reethu, "A comparitive analysis of classification algorithms on different stimuli for eeg based bci applications," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1424–1428, IEEE, 2019.

[62] X. Tang, T. Wang, Y. Du, and Y. Dai, "Motor imagery eeg recognition with knn-based smooth auto-encoder," *Artificial Intelligence in Medicine*, p. 101747, 2019.

[63] K. Prince, P. Grace, J. Premalatha, D. Marshiana, K. Raghavan, and S. Kumar, "Application of clustering techniques on statistical features of eeg signals for seizure detection.," *Indian Journal of Public Health Research & Development*, vol. 10, no. 7, 2019.

[64] S. Abdulla, M. Diykh, R. L. Laft, K. Saleh, and R. C. Deo, "Sleep eeg signal analysis based on correlation graph similarity coupled with an ensemble extreme machine learning algorithm," *Expert Systems With Applications*, vol. 138, p. 112790, 2019.

[65] Z. Y. Chin, K. K. Ang, C. Wang, C. Guan, and H. Zhang, "Multi-class filter bank common spatial pattern for four-class motor imagery bci," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 571–574, IEEE, 2009.

[66] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, and K. K. Ang, "A new discriminative common spatial pattern method for motor imagery brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 11, pp. 2730–2733, 2009.

[67] Y. Wang, Y.-T. Wang, and T.-P. Jung, "Translation of eeg spatial filters from resting to motor imagery using independent component analysis," *PloS one*, vol. 7, no. 5, p. e37665, 2012.

[68] J. Kevric and A. Subasi, "Comparison of signal decomposition methods in classification of eeg signals for motor-imagery bci system," *Biomedical Signal Processing and Control*, vol. 31, pp. 398–406, 2017.

[69] S. Taran, V. Bajaj, D. Sharma, S. Siuly, and A. Sengur, "Features based on analytic imf for classifying motor imagery eeg signals in bci applications," *Measurement*, vol. 116, pp. 68–76, 2018.

[70] M. R. Islam, T. Tanaka, and M. K. I. Molla, "Multiband tangent space mapping and feature selection for classification of eeg during motor imagery," *Journal of neural engineering*, vol. 15, no. 4, p. 046021, 2018.

[71] S. Taran and V. Bajaj, "Motor imagery tasks-based eeg signals classification using tunable-q wavelet transform," *Neural Computing and Applications*, pp. 1–8, 2018.

[72] J. Luo, Z. Feng, and N. Lu, "Spatio-temporal discrepancy feature for classification of motor imageries," *Biomedical Signal Processing and Control*, vol. 47, pp. 137–144, 2019.

[73] B. Zhou, X. Wu, J. Ruan, L. Zhao, and L. Zhang, "How many channels are suitable for independent component analysis in motor imagery brain-computer interface," *Biomedical Signal Processing and Control*, vol. 50, pp. 103–120, 2019.

[74] S. Guan, K. Zhao, and S. Yang, "Motor imagery eeg classification based on decision tree framework and riemannian geometry," *Computational intelligence and neuroscience*, vol. 2019, 2019.

[75] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "Eeg-based brain-computer interfaces using motor-imagery: Techniques and challenges," *Sensors*, vol. 19, no. 6, p. 1423, 2019.

[76] M. M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.

[77] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[78] M. Congedo, C. Gouy-Pailler, and C. Jutten, "On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics," *Clinical Neurophysiology*, vol. 119, no. 12, pp. 2677–2686, 2008.

[79] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," in *IEE Proceedings F-Radar and Signal Processing*, vol. 140, pp. 362–370, IET, 1993.

[80] O.-E. C. Javier, P. K. Daniela, B. G.-B. Giovanni, and R.-R. Juvenal, "Blind source separation problem algorithms for audio and biomedical signals," in *2018 XIV International Engineering Congress (CONIIN)*, pp. 1–7, IEEE, 2018.

[81] A. Belouchrani, K. Abed-Meraim, J. Cardoso, and E. Moulines, "Second-order blind separation of temporally correlated sources," in *Proc. Int. Conf. Digital Signal Processing*, pp. 346–351, Citeseer, 1993.

[82] J. Bigot, M. Longcamp, F. Dal Maso, and D. Amarantini, "A new statistical test based on the wavelet cross-spectrum to detect time–frequency dependence between non-stationary signals: Application to the analysis of cortico-muscular interactions," *NeuroImage*, vol. 55, no. 4, pp. 1504–1518, 2011.

[83] T. Giannakopoulos and A. Pikrakis, *Introduction to audio analysis: a MATLAB® approach*. Academic Press, 2014.

[84] S. Katagiri, *Handbook of neural networks for speech processing*, vol. 171. Artech House Boston, 2000.

[85] W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka, "Parallel distributed processing model with local space-invariant interconnections and its optical architecture," *Applied optics*, vol. 29, no. 32, pp. 4790–4797, 1990.

[86] "A comprehensive guide to convolutional neural networks — the eli5 way." `https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53`. Accessed: 2019-09-09.

[87] "Data set iva." http://www.bbci.de/competition/iii/desc_IVa.html. Accessed: 2019-08-10.

[88] B. Blankertz, K.-R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birbaumer, "The bci competition iii: Validating alternative approaches to actual bci problems," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 14, no. 2, pp. 153–159, 2006.

[89] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[90] M. Dai, D. Zheng, R. Na, S. Wang, and S. Zhang, "Eeg classification of motor imagery using a novel deep learning framework," *Sensors*, vol. 19, no. 3, p. 551, 2019.

[91] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted boltzmann machines," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 25, no. 6, pp. 566–576, 2016.

[92] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[93] Z. Tang, C. Li, and S. Sun, "Single-trial eeg classification of motor imagery using deep convolutional neural networks," *Optik-International Journal for Light and Electron Optics*, vol. 130, pp. 11–18, 2017.

[94] Z. Tayeb, J. Fedjaev, N. Ghaboosi, C. Richter, L. Everding, X. Qu, Y. Wu, G. Cheng, and J. Conradt, "Validating deep neural networks for online decoding of motor imagery movements from eeg signals," *Sensors*, vol. 19, no. 1, p. 210, 2019.

[95] X. Zhang, L. Yao, X. Wang, J. Monaghan, and D. Mcalpine, "A survey on deep learning based brain computer interface: Recent advances and new frontiers," *arXiv preprint arXiv:1905.04149*, 2019.

[96] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of eeg motor imagery signals," *Journal of neural engineering*, vol. 14, no. 1, p. 016003, 2016.

[97] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of neural engineering*, 2019.

[98] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.

[99] "Morse wavelets." https://la.mathworks.com/help/wavelet/ug/morse-wavelets.html. Accessed: 2019-08-10.

[100] M. Klemm, J. Haueisen, and G. Ivanova, "Independent component analysis: comparison of algorithms for the investigation of surface electrical brain activity," *Medical & biological engineering & computing*, vol. 47, no. 4, pp. 413–423, 2009.

[101] L. Albera, A. Kachenoura, P. Comon, A. Karfoul, F. Wendling, L. Senhadji, and I. Merlet, "Ica-based eeg denoising: a comparative analysis of fifteen methods," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 60, no. 3, pp. 407–418, 2012.

[102] H. Wang, Y. Zhang, *et al.*, "Detection of motor imagery eeg signals employing naïve bayes based learning process," *Measurement*, vol. 86, pp. 148–158, 2016.