UNIVERSITÀ POLITECNICA
DELLE MARCHE

FACULTY OF ENGINEERING

MASTER OF SCIENCE IN BIOMEDICAL ENGINEERING

Motor/Imaginary movement classification based on Fractal Dimension FD and event-related desynchronization ERD

Supervisor:

Prof.Camillo Porcaro

Candidate:

Zahra Mansour

February 2021

Thanks and gratitude...

All thanks and appreciation to ***professor Camillo Porcaro*** *and* ***Professor Laura Burattini*** *and all professors in Università Politecnica Delle Marche for their great support and precious effort.*

To the lady who taught me how to continue no matter how hard the life is, the one who sacrificed with her life to give me the happiness, nothing even death is able to separate her heart from mine, to the one who I miss the most... .

My mom

To my hero, my backbone, and my soulmate, the most patents and kindest person on the earth.

My Dad

To my oldest friend, the one who hold my hands from my early childhood and still.

My brother Abed

To my family in Sweden who support me like no one did, there is no words to explain how thankful I am. I could not do it without your kind souls.

Dr. Nahla's family

To my guardian angels, who comfort my soul with their prayers.

My uncles and my aunts

To my sisters, my friends , the stars in my dark nights.

Weddad, Nour, Faten, Dunia, Nisreen, Aya, and Ruba

To my friends in Italy who made this journey unforgettable and filled it with precious memories, you were more than family to me.

Elisa, Jessica , Elisea and Chiara

To my best friend, my super hero, my lord...

Jafar

To my little cutie pies, I hope that you will read this one day and understand how much I love you.

Adam, Khaled and Youssef

وعلى النبي وال البيت والشهداء مولاي صل وسلم دائما أبدا ....

CONTENTS

List of Figures

List of tables

Summary

Until recently, the dream of being able to control one's environment through thoughts had been in the realm of science fiction. However, the advance of technology has brought a new reality: Today, humans can use the electrical signals from brain activity to interact with, influence, or change their environments. The emerging field of brain-computer interface (BCI) technology may allow individuals unable to speak and/or use their limbs to once again communicate or operate assistive devices for walking and manipulating objects.

In this study, a PhysioNet data set have been used to create a BCI system, by investigating 1527 EEG signals during real and imagination movement tasks in time domain using Fractal Dimension (FD), and in frequency domain using Event related desynchronisation (ERD).

Different classifiers such as SVM, KNN and TREE have been applied with various combination of EEG channels. Moreover, the problem of selecting channels group have been discussed by proposing a new technique to select the optimal channels combination for each task and each classifier.

It has been found that the optimal channels selection technique is significantly able to improve the performance of each classifier. With accuracy equal to 98% for classifying between real and imagination of hand movement (se= 97%, sp=99%, AUC=0.99) obtained by using ERD with SVM. While the accuracy of classifying between the imagination of hand and feet movements equal to 91% (se=87% , sp=86% , AUC=0.93).

This study proposed different models of BCI system in order to select the best one with the highest accuracy to help removing the human communication boundaries and to improve the quality of the life for people with disabilities.

Chapter I

Origin of brain EEG signal

1. Chapter 1: Origin of EEG signal

1.1 overview of central nervous system

The nervous system is a sophisticated network of specialized cells which coordinates body activities according to the state of internal and external dynamic conditions. The nervous system is comprised of two main parts: The central nervous system (CNS) and the peripherical nervous system(PNS) [1].

CNS consists of spinal cord lying within the bony vertebral column and the brain enclosed within the skull. PNS is made up of the nerves that serve arms, legs, trunk skeletal muscles and internal organs as it shown in (Figure 1).

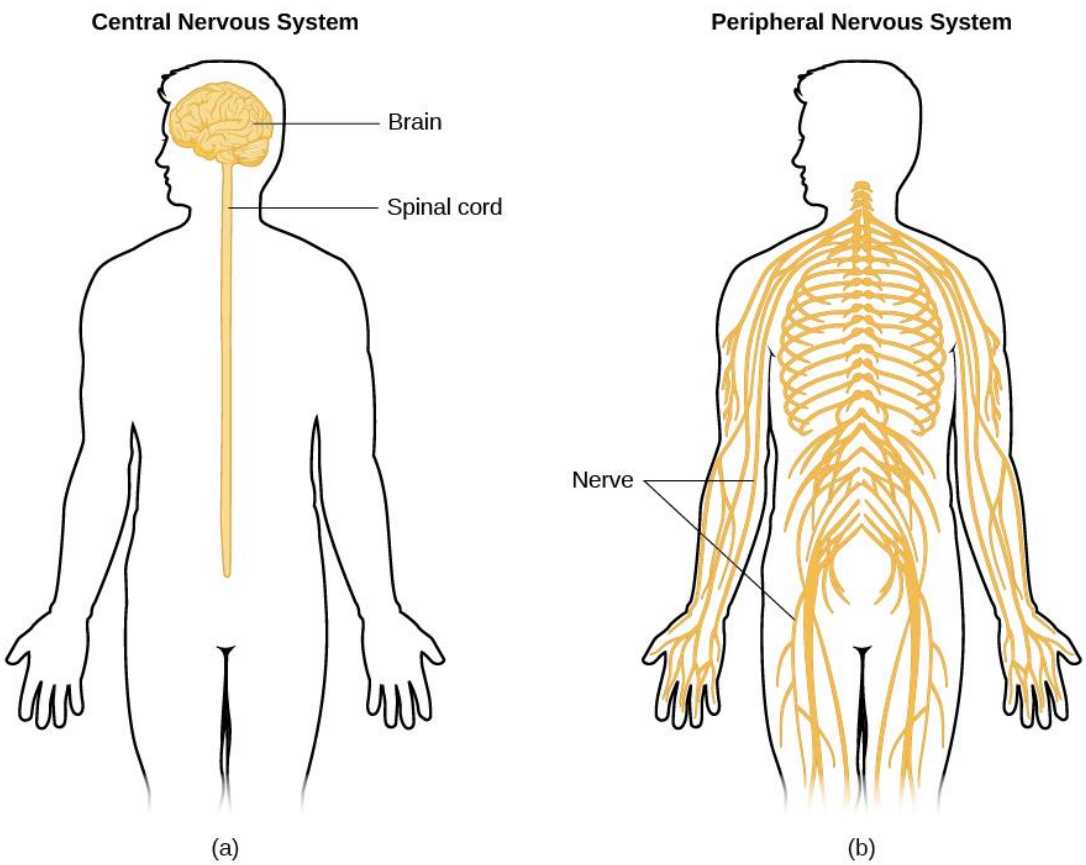

Figure 1 the different parts of the nervous system, (a) the central nervous system CNS which consists of the brain and the spinal cord, and (b) is the peripheral nervous system PNS which consists of the various nerves.

The smallest specialized intellectual unit of the nervous system is called the neuron. Each neuron is designed to transmit information to other neurons or other cells in the nervous system. The brain, as the largest and most complex part of CNS, consists of One hundred billion,

numbers of neurons. The brain owes its sophisticated functionalities, such as cognition, emotion, sensory and motor function, to the structural and functional proprieties of neurons. Although those phenomena arise in part from single neuron proprieties, mostly they are the cumulative outcome of networks in which vast number of neurons constitute local and global circuitries via electro-chemical interconnections, called synapses (the average number of synapses in the human brain is 10^{14}.

Since there are multiple subtypes of neurons and synapses, neural circuitries show much more functional and structural diversity than those of single neurons.

From anatomical and functional point of view, brain consists of three main parts: brainstem, cerebellum and cerebrum (Figure 2). The brainstem is actually a continuity of the spinal cord and it constitutes a connection among cerebral cortex, spinal cord and cerebellum. Main component of the brainstem is the thalamus which relays and integrates signals coming from sensory system to respective cortical areas. The cerebellum mainly serves as a coordinator of the somatic muscle system to provide balance and harmonious muscle activity. The cerebrum is the largest part of the brain and is accounted to be the part that actually distinguishes human from other high vertebrates [2].

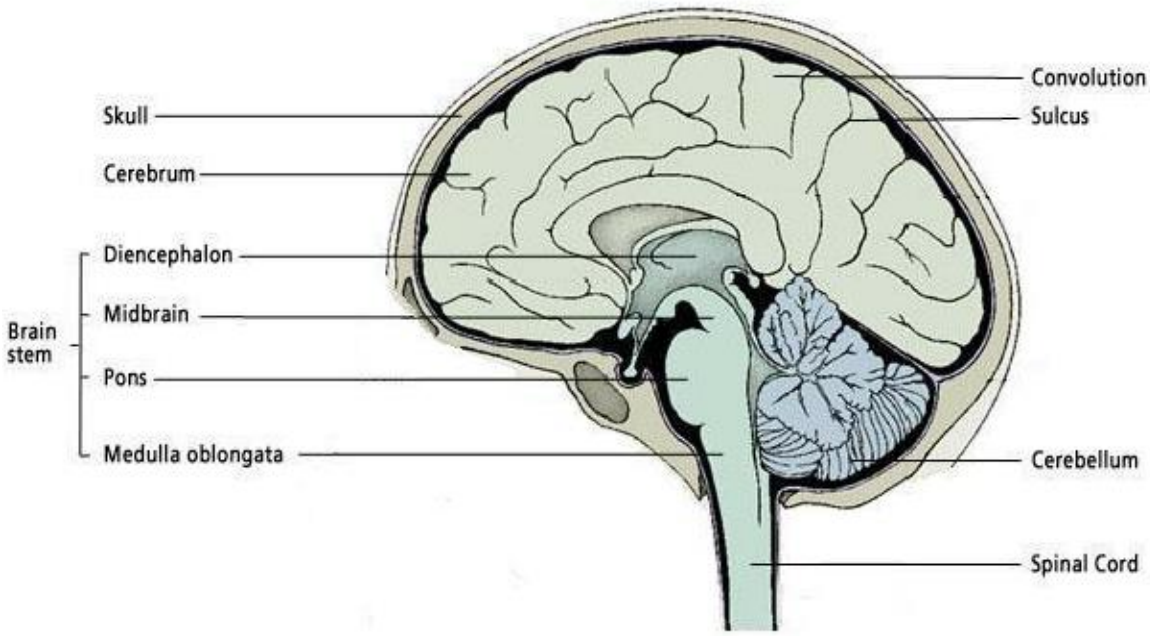

Figure 2 the main parts of the brain, Cerebrum, Cerebellum and Brain stem.

The outer cerebrum, or the cerebral cortex, is known as the heavily convulsed outer surface of the brain. These convolutions are made up of gyri (gyrus) or ridges and sulci (sulcus) or valley which provide convenient boundaries between the four part (frontal, parietal, temporal and occipital lobes) of the cerebral cortex. The frontal lobe is responsible for executive functions, thinking and planning of voluntary motor executions; and the parietal lobe is where visual and somatosensory information is integrated. At the side of the head is the occipital lobe which is involved in visual processing and perception. More detailed distinction of cortical areas is available as Brodmann's areas where the cerebral cortex is mapped out on the basis of functional and histological patterns [3] (Figure 3).

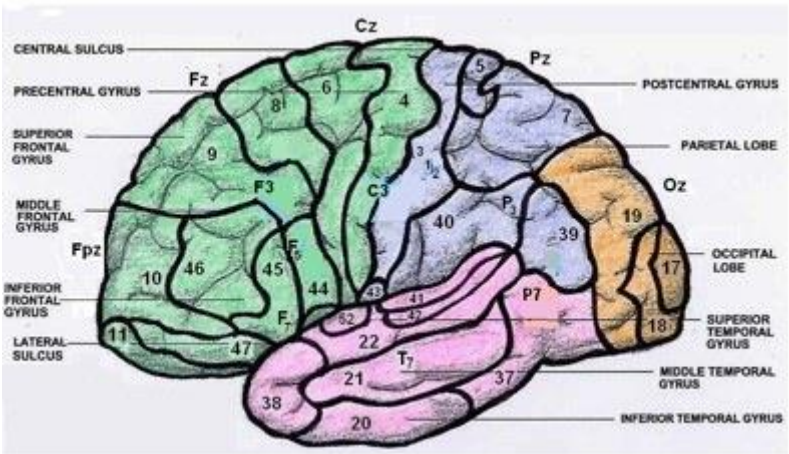

Figure 3 Brain lobes and Brodmann's area

1.2 Electroencephalography (EEG)

The first electrical neural activities were registered using simple galvanometers. In order to magnify very fine variations of the pointer a mirror was used to reflect the light projected to the galvanometer on the wall.

The d'Arsonval galvanometer later featured a mirror mounted on a movable coil and the light focused on the mirror was reflected when a current passed the coil. The capillary electrometer was introduced by Lippmann and Marey. The string galvanometer, as a very sensitive and more accurate measuring instrument, was introduced by Einthoven in 1903. This became a standard instrument for a few decades and enabled photographic recording.

It was 1920s that it was first systematically analyzed by the German psychiatrist Hans Berger who coined the term of Electroencephalogram (EEG).[4]

Basically, EEG denotes the spontaneous fluctuations of cortical potentials that are recorded over the scalp. Recording techniques of the electrical activity of the brain is labelled according to the position of the electrode used.

Unlike EEG , when the electrode are placed directly over the surface of the cerebral cortex, the recording is called electrocorticogram (ECoG). For all of these methods, the recorded potentials are the superposition of the volume conductor fields generated by a several neural circuitries. The major group of cortical neurons that take part in the electro-genesis of the scalp potentials are the *pyramidal cells* which are oriented vertically in the cortex and whose cell bodies are in triangular shape, with the apex directed to the cortical surface [5] (figure 4).

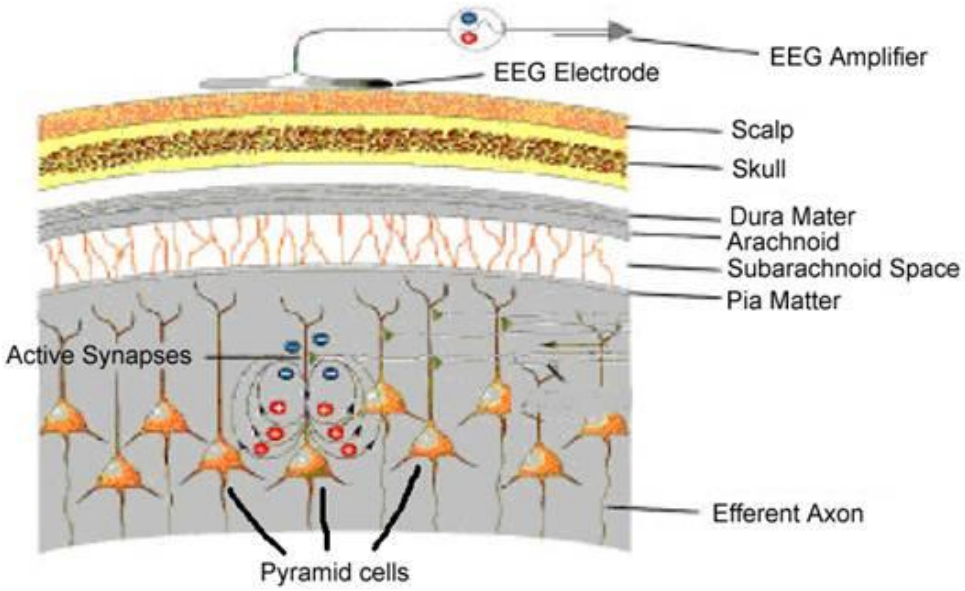

Figure 4 recording the post synaptic potential of the Pyramid cells using EEG electrode

When depolarizing synaptic input reaches to the dendrites of the pyramidal cells, sub-threshold current flows through the dendrites and the cell body, then it returns to the synaptic site via extracellular bathing medium. As it is indicated in (Figure 4), the extracellular medium about the soma becomes a source(+) and the dendritic site behaves like a sink(-) constituting a current dipole. Depending on the number and sign of the synaptic inputs, and also on the position and orientation of the pyramidal cells, those dipoles show a very dynamic behaviour. 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 [6].

The human head consists of different layers including the scalp, skull, brain , and many other thin layers in between. The skull attenuates the signals approximately one hundred times more than the soft tissue. On the other hand, most of the noise is generated either within the brain (internal noise) or over the scalp (system noise or external noise). Therefore, only large populations of active neurons can generate enough potential to be recordable using the scalp electrodes. These signals are later amplified greatly for display purposes.

Acquiring signals and images from the human body has become vital for early diagnosis of a variety of diseases. Such data can be in the form of electrobiological signals such as an electrocardiogram (ECG) from the heart, electromyogram (EMG) from muscles, electroencephalogram (EEG) from the brain, magnetoencephalogram (MEG) from the brain, electrogastrogram (EGG) from the stomach, and electrooculogram (or EOG) from eye nerves.

Measurements can also have the form of one type of ultrasound or radiography such as sonograph (or ultrasound image), computerized tomography (CT), magnetic resonance imaging (MRI) or functional MRI (fMRI), positron emission tomography (PET), and single photon emission tomography (SPET). Functional and physiological changes within the brain may be registered by either EEG, MEG, or fMRI.[7]

More recent EEG systems consist of a number of delicate electrodes, a set of differential amplifiers (one for each channel) followed by filters, and needle (pen)-type registers. The multichannel EEGs could be plotted on a paper or paper with a grid. Soon after this system came to the market, researchers started looking for a computerized system, which could digitize and store the signals. Therefore, to analyse EEG signals it was soon understood that the signals must be in digital form. This required sampling, quantization, and encoding of the signals. As the number of electrodes grows the data volume, in terms of the number of bits, increases. The

computerized systems allow variable settings, stimulations, and sampling frequency, and some are equipped with simple or advanced signal processing tools for processing the signals.

The conversion from analogue to digital EEG is performed by means of multichannel analogue-to-digital converters (ADCs). Fortunately, the effective bandwidth for EEG signals is limited to approximately 100 Hz. For many applications this bandwidth may be considered to be even half of this value. Therefore, a minimum frequency of 200 samples/s (to satisfy the Nyquist criterion) is often enough for sampling the EEG signals. In some applications where a higher resolution is required for representation of brain activities in the frequency domain, sampling frequencies of up to 2000 sample/s may be used. [8]

In order to maintain the diagnostic information the quantization of EEG signals is normally very fine. Representation of each signal sample with up to 16 bits is very popular for the EEG recording systems. This makes the necessary memory volume for archiving the signals massive, especially for sleep EEG and epileptic seizure monitoring records. However, in general, the memory size for archiving the radiological images is often much larger than that used for archiving the EEG signals. [9]

The EEG recording electrodes and their proper function are crucial for acquiring high-quality data. Different types of electrodes are often used in the EEG recording systems, such as:

• disposable (gel-less, and pre-gelled types);

• reusable disc electrodes (gold, silver, stainless steel, or tin);

• headbands and electrode caps;

• saline-based electrodes;

• needle electrodes.

For multichannel recordings with a large number of electrodes, electrode caps are often used. Commonly used scalp electrodes consist of Ag–AgCl disks, less than 3 mm in diameter, with long flexible leads that can be plugged into an amplifier. Needle electrodes are those that have to be implanted under the skull with minimal invasive operations. High impedance between the cortex and the electrodes as well as the electrodes with high impedances can lead to distortion, which can even mask the actual EEG signals. Commercial EEG recording systems are often equipped with impedance monitors. To enable a satisfactory recording the electrode

impedances should read less than 5 kΩ and be balanced to within 1 kΩ of each other. For more accurate measurement the impedances are checked after each trial. [10]

Due to the layered and spiral structure of the brain, however, distribution of the potentials over the scalp (or cortex) is not uniform. This may affect some of the results of source localization using the EEG signals.

The International Federation of Societies for Electroencephalography and Clinical Neurophysiology has recommended the conventional electrode setting (also called 10–20 system) for 21 electrodes (excluding the earlobe electrodes).[11]

The 10–20 system avoids both eyeball placement and considers some constant distances by using specific anatomic landmarks from which the measurement would be made and then uses 10 or 20 % of that specified distance as the electrode interval. The odd electrodes are on the left and the even ones on the right.[12]

In many applications such as brain–computer interfacing (BCI) and study of mental activity, often a small number of electrodes around the movement-related regions are selected and used from the 10–20 setting system.

Chapter II

Brain Computer Interface BCI

2. Chapter 2: Brain Computer Interface (BCI)

2.1 Historical Review

Communication, or social interaction, is one of the key principles of human civilization. This quality enables one to share emotions, expectations, and creative thoughts amongst human beings. In the event that this communication is established through speech, gesture, or writing, human communication becomes easier and devoid of constraints. Nonetheless, people who are suffering from locked-in syndrome do not have the aforementioned options for interaction. Patients with locked-in syndrome could not interact or express themselves, although they are well-cognizant of things around them. Amyotrophic lateral sclerosis (ALS), cerebral palsy, brain stem stroke, multiple sclerosis, cerebral palsy, and spinal cord injury are the main causes of locked-in syndrome. It is almost impossible for a person who is affected by the locked-in syndrome to communicate with other persons, and hence, Brain–Computer Interface (BCI) is a promising means to furnish them with basic communication abilities. [13]

BCI is s a direct communication pathway between an enhanced or wired brain and an external device. BCIs are often directed at researching, mapping, assisting, augmenting, or repairing human cognitive or sensory-motor functions

Fundamentally, the human brain and devices are interfaced through the concept of BCI in which the users will have to generate a variety of brain waves that will be recognized and converted into commands to the devices [14]. In its earlier days, researchers intended to use this technology to develop assistive devices for medical purposes only. Nonetheless, the employment of this technology has expanded, and it has found its way into non-medical applications. It is discernible that over the last 15 years, a considerable number of original articles as well as reviews have been published on BCI.

Research on BCIs began in the 1970s at the University of California, Los Angeles (UCLA) under a grant from the National Science Foundation.
The year 1998 marked a significant development in the field of brain mapping when researcher Philip Kennedy implanted the first brain computer interface object into a human being. However, the BCI object was of limited function. The only benefit from this development was the use of a wireless di-electrode.
John Donoghue and his team of Brown University researchers formed a public traded company, Cybernetics, in 2001. The goal was to commercially design a brain computer interface, the so-

called Brain Gate. The company has come up with NeuroPort ™- its first commercial product. Columbia University Medical Centre researchers have successfully monitored and recorded electrical activity in the brain with improved precision. According to researchers, NeuroPort™ Neural Monitoring System enabled them to identify micro-seizure activity prior to epileptic seizures among patients. [15]

June 2004 marked a significant development in the field when Matthew Nagle became the first human to be implanted with a BCI, Cyberkinetics's BrainGate™.

In December 2004, Jonathan Wolpaw and researchers at New York State Department of Health's Wadsworth Centre came up with a research report that demonstrated the ability to control a computer using a BCI. In the study, patients were asked to wear a cap that contained electrodes to capture EEG signals from the motor cortex – part of the cerebrum governing movement.

Recently, studies in human-computer interaction through the application of machine learning with statistical temporal features extracted from the frontal lobe, EEG brainwave data has shown high levels of success in classifying mental states (Relaxed, Neutral, Concentrating), mental emotional states (Negative, Neutral, Positive) and thalamocortical dysrhythmia. [16]

2.2 Main components of BCI system

The working of the BCI system requires three modules that are signal acquisition module, signal processing, and application module. This section describes the working of each module. (Figure 5) shows the components of BCI and their interactions.

- ***Signal acquisition module***

The Signal Acquisition Module is liable for recording the electrophysiological signals that provide input to the BCI. These signals are recorded from the scalp or from the surface of the brain or neuronal activity. BCI might use either invasive methods or non-invasive methods for signal acquisition. Invasive methods are electrocardiograms (ECoG) and single-neuron recordings and have better signal quality as compared to non-invasive methods. Non-invasive methods are Electroencephalogram (EEG), Magnetoencephalogram (MEG), Positron Emission Tomography (PET), Functional Magnetic Resonance Imaging (fMRI) and Near-Infrared Spectroscopy (NIRs) [17].

The acquired signals are amplified to enhance the strength and are digitized before they are used by any of signal processing module

- ***Pre-processing***

The task of pre-processing is to prepare the recorded signals for processing by enhancing the signal –to- noise ratio (SNR). The part of EEG signal that comes from muscular activity of head, and eye movement generate electrical activity that is unrelated to the brain. Such part of signal is considered as artifact and should not be processed in order to preserve and exhibit the relevant information; therefore pre-processing is done to remove artifacts in EEG signals. In BCI research, the proper pre-processing of EEG signal is important in order to obtain high classification accuracy.

- ***Feature extraction***

After pre-processing the signal is fed into one or more type of feature extraction algorithms. This component extracts features in the time domain and frequency domain that encode messages or commands [18].

Wide varieties of feature extraction methods are used in BCI system.

some of these methods include amplitude measures, band power, Hjorth parameters, autoregressive models, and wavelets and spatial filters.

- ***Classification***

The task of the classification component is to translate the features provided by the feature extractor to a category of brain patterns; that is the independent variable is converted into the dependent variable. The classification algorithms may use linear methods like Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) or non-linear methods such as neural networks.

- ***Application module***

For most current BCIs, the output device is a computer screen and the output is the selection of targets, letters, or icons presented on it .

Some BCIs provide an output, such as cursor movement toward the item prior to its selection. The output generated by the output device is the feedback provided to the user to notify the user about the recognized brain activity pattern. This pattern is then used to sustain and enhance the accuracy and speed of communication. [19]

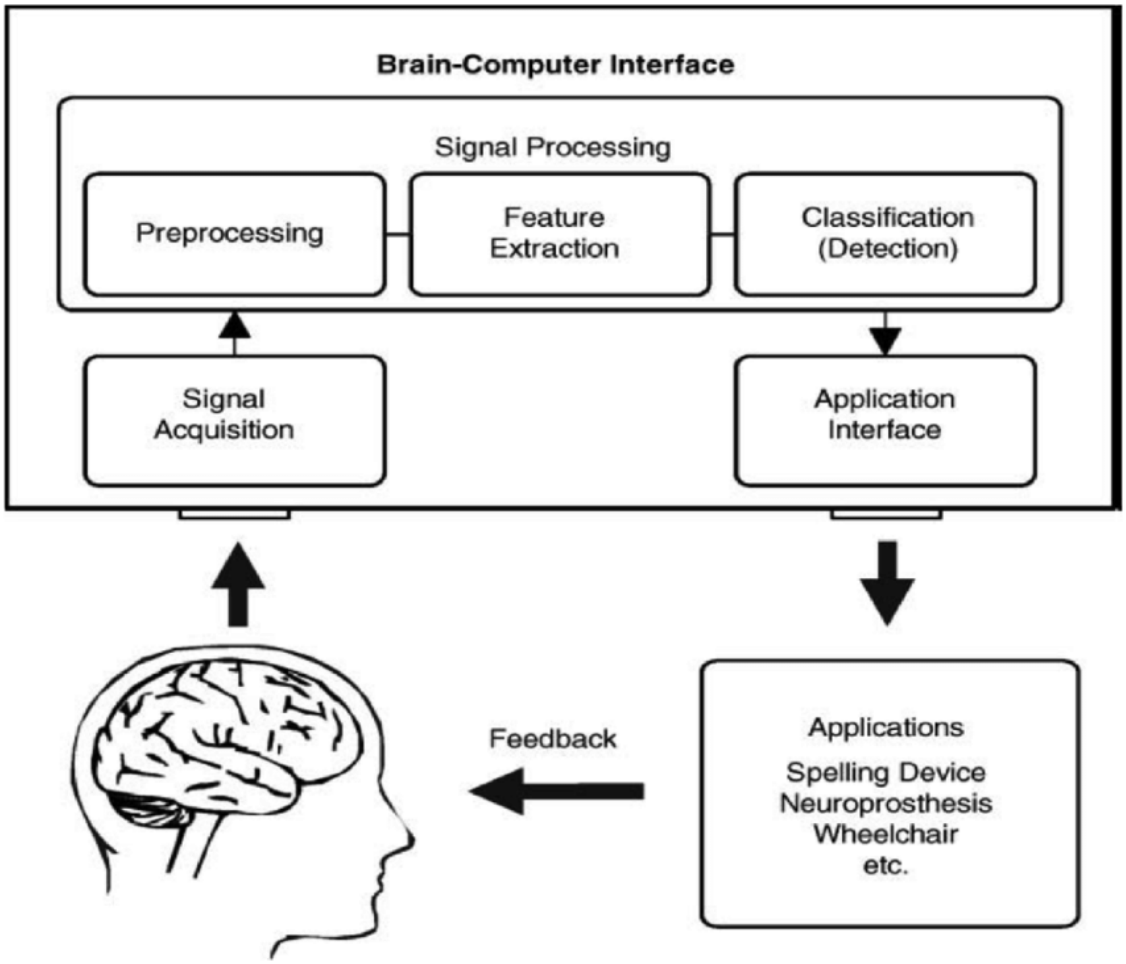

Figure 5 Main components of any BCI system.

2.3 the type of BCI systems

Generally, BCI frameworks may be separated into a number of classes. Figure 6 illustrates the three categorization schemes, namely by dependability, recording technique, and method of operation. Regarding dependability, BCI can be classed as either dependent or independent BCI. Dependent BCIs require some form of motor control by the user or healthy subjects, for instance, gaze control. MI-based BCIs are an ideal example of dependent BCI systems and have been extensively utilized. Conversely, independent BCIs do not require any form of motor control by the user; this type of BCI is ideal for stroke patients or severely impaired patients. an SSVEP-based independent BCI system was proposed to identify two different targets, and it was demonstrated to be successful [20].

With regard to recording method, BCI can be categorized into invasive and non-invasive. Microelectrode arrays are often required to be implanted inside the skull for invasive BCIs. Two common invasive modalities that have been reported in BCI research are intracortical recording and electrocorticography (ECoG). Conversely, if the brain signals are acquired by means of sensors placed on the scalp, it is known as non-invasive BCI. Amongst the non-invasive modalities often utilized are EEG, MEG, PET, fMRI, and fNIRS. In BCIs, EEG is the most widely employed non-invasive modality, where a variety of control signals, including SCP, SSVEP, MI, and P300, can be evoked.

Finally, BCI can have either a synchronous or asynchronous mode of operation. The interaction between the user and the system may be either time-dependent or time-independent. In the event that the interaction is carried out within a certain period of time upon a cue imposed by the system, then the system is known as synchronous BCI. In contrast, in asynchronous BCI, the subject can generate a mental task at any period of time to interact with the application. In comparison with asynchronous BCI, synchronous BCIs are not user-friendly, but designing such a system is much easier than for asynchronous BCI.[21]

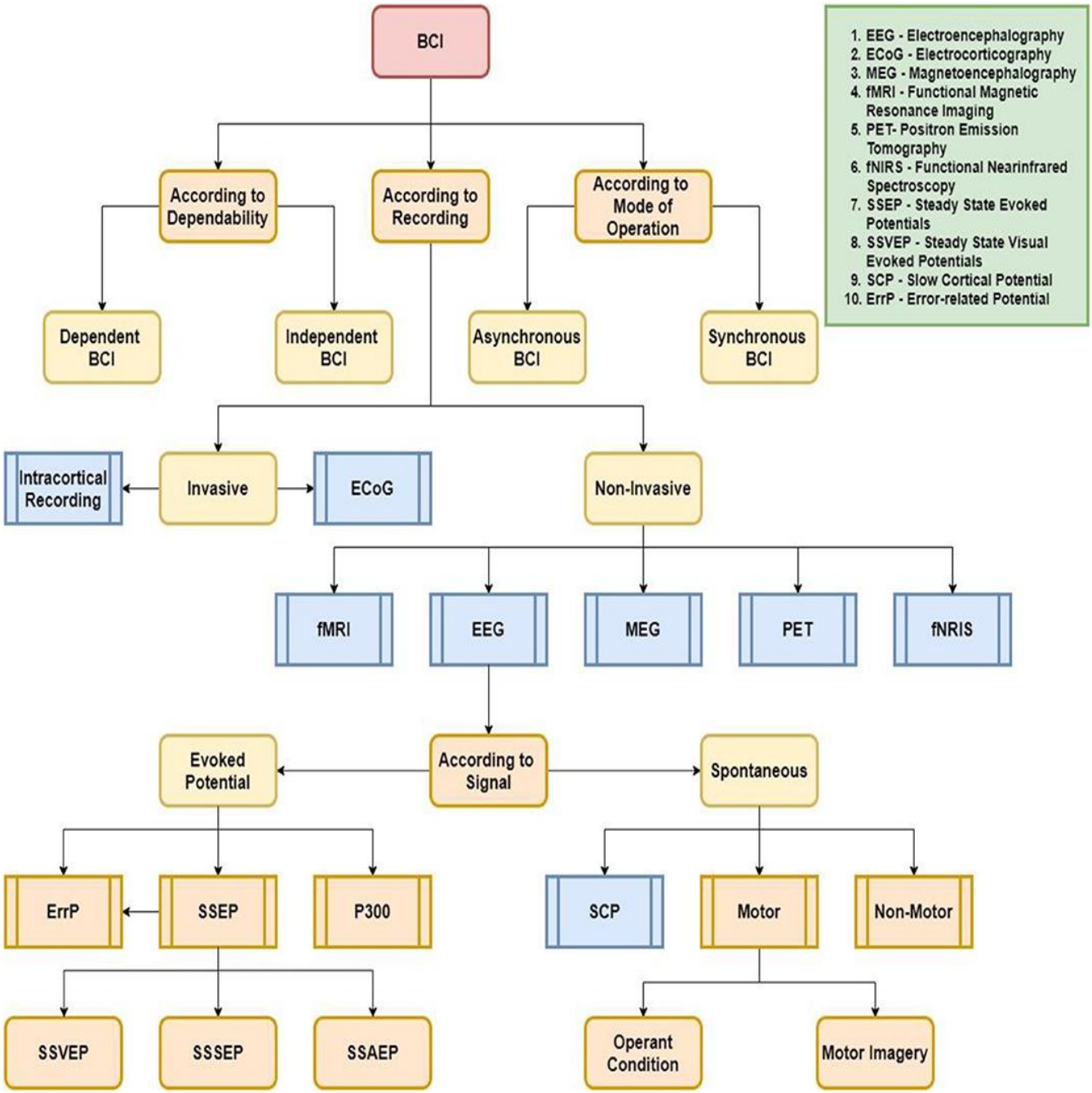

Figure 6 Classification of BCI systems in terms of dependability, recording method, and mode of operation.

Chapter III

Classification

3. Chapter 3: classification

3.1 General introduction about classification

classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known.

An algorithm that implements classification, especially in a concrete implementation, is known as a classifier.

Classification has many applications such as computer vision, medical imagining, speech recognition and biological classification. [22]

3.2 Support Vector Machine SVM classifier

A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems.

with associated learning algorithms that analyse data for classification and regression analysis. Developed at AT&T Bell Laboratories by Vapnik with colleagues (Boser et al., 1992, Guyon et al., 1993, Vapnik et al., 1997).

SVMs are one of the most robust prediction methods, being based on statistical learning frameworks or VC theory proposed by Vapnik and Chervonenkis (1974) and Vapnik (1982, 1995). [23]

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

SVM constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin, the lower the generalization error of the classifier.[24]

Whereas the original problem may be stated in a finite-dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products of

pairs of input data vectors may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function selected to suit the problem.[5] The hyperplanes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant, where such a set of vectors is an orthogonal (and thus minimal) set of vectors that defines a hyperplane.

SVMs can be used to solve various real-world problems such as Classification of images, Hand-written characters, biological and other sciences and text and hypertext categorization.

In addition to linear SVM, Non-linear SVM can be obtained by every dot product is replaced by a nonlinear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space high-dimensional; although the classifier is a hyperplane in the transformed feature space, it may be nonlinear in the original input space. [25]

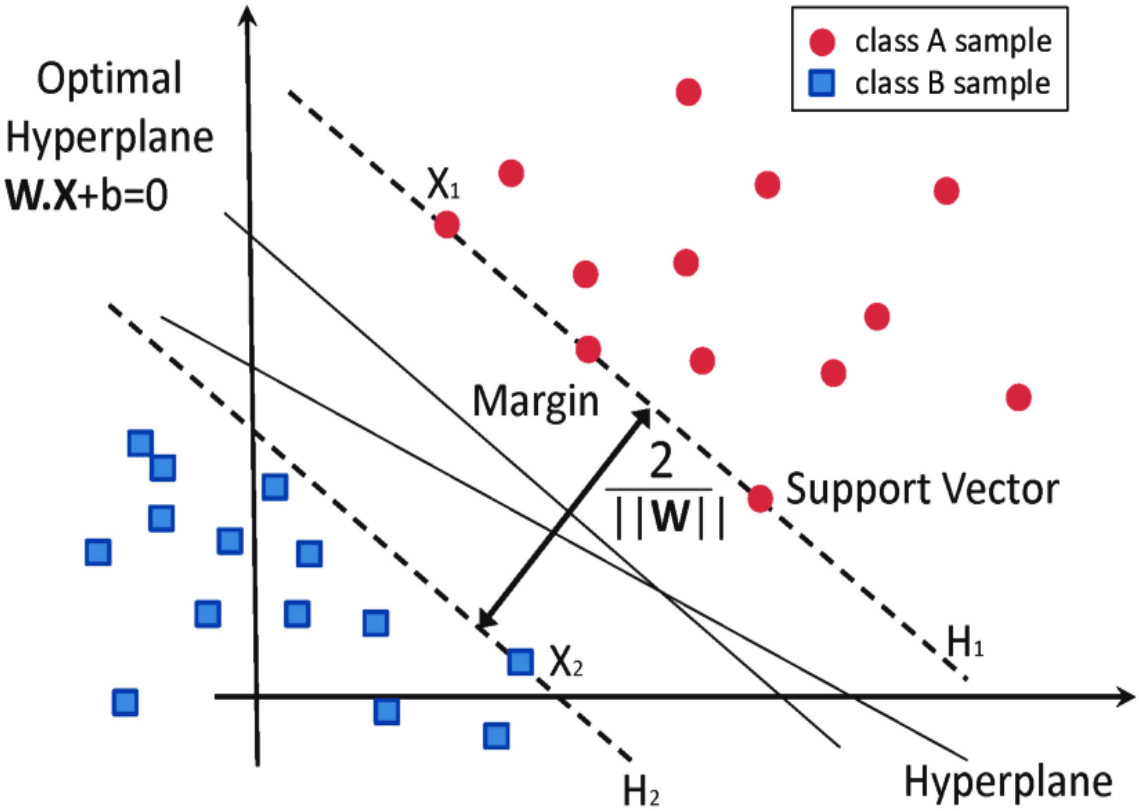

Figure 7 the theory behind SVM

3.3 K-Nearest Neighbors (KNN) classifier

k-nearest Neighbors algorithm (KNN) is a non-parametric machine learning method first developed by Evelyn Fix and Joseph Hodges in 1951, and later expanded by Thomas Cover. It is used for classification and regression. In both cases, the input consists of the k closest training examples in feature space. The output depends on whether KNN is used for classification or regression. [26]

KNN is a type of instance-based learning, where the function is only approximated locally and all computation is deferred until function evaluation. Since this algorithm relies on distance for classification, if the features represent different physical units or come in vastly different scales then normalizing the training data can improve its accuracy dramatically.

KNN is a special case of a variable-bandwidth, kernel density "balloon" estimator with a uniform kernel.

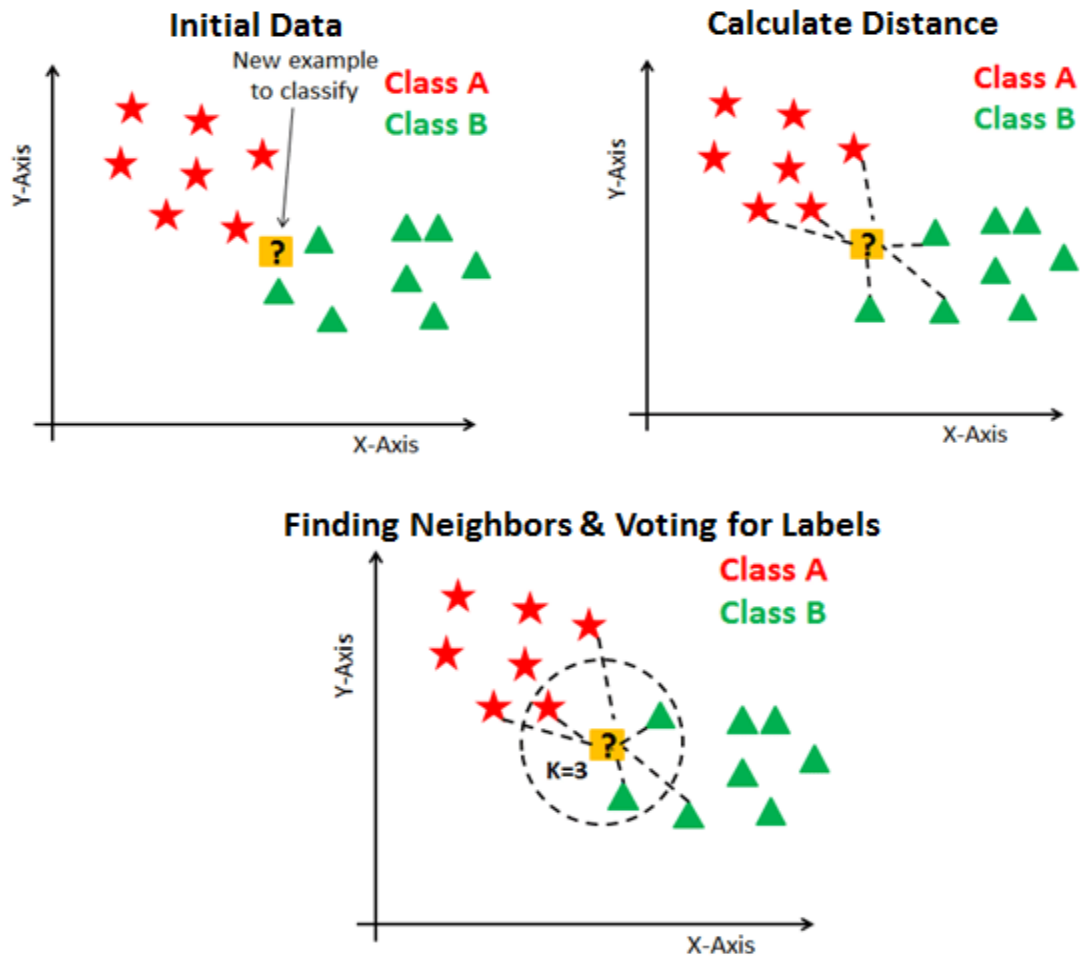

Figure 8 the theory behind KNN

The

test example to all stored examples, but it is computationally intensive for large training sets.

Using an approximate nearest neighbour search algorithm makes KNN computationally tractable even for large data sets. Many nearest neighbour search algorithms have been proposed over the years; these generally seek to reduce the number of distance evaluations actually performed.

KNN has some strong consistency results. As the amount of data approaches infinity, the two-class KNN algorithm is guaranteed to yield an error rate no worse than twice the Bayes error rate (the minimum achievable error rate given the distribution of the data). Various improvements to the KNN speed are possible by using proximity graphs. [27]

3.4 TREE classifier

Decision tree learning is one of the predictive modelling approaches used in statistics, data mining and machine learning. It uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves).

Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.

Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees. Decision trees are among the most popular machine learning algorithms given their intelligibility and simplicity.

Decision tree have many advantages such as it is requires little data preparation, it is Performs well with large datasets and Mirrors human decision making more closely than other approaches. [28]

In the same time it is suffered from several limitation for example small change in the training data can result in a large change in the tree and consequently the final predictions.

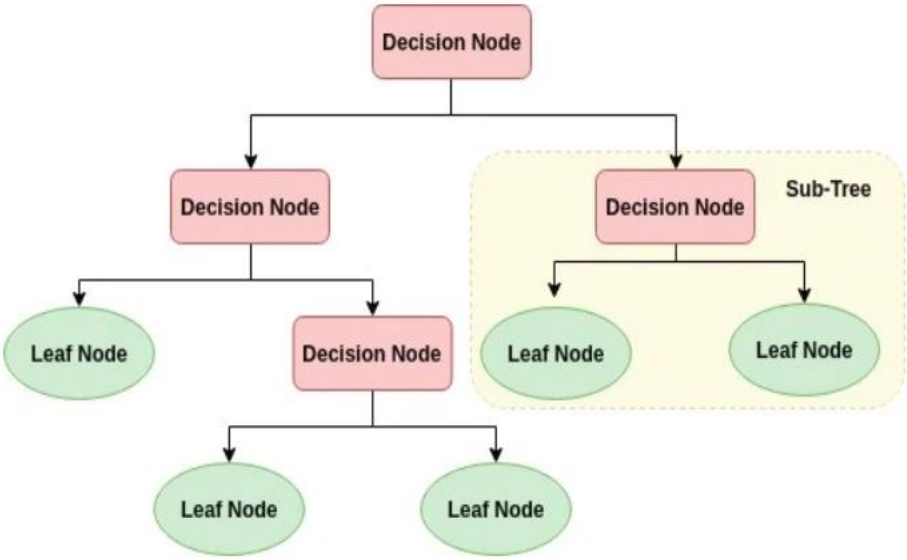

Figure 9 the theory behind the decision tree algorithm

3.5 Bayaze optimization

Bayesian optimization is typically used on problems of the form max x∈ f(x), where A is a set of points whose membership can easily be evaluated. Bayesian optimization is particularly advantageous for problems where f(x) is difficult to evaluate, is a black box with some unknown structure, relies upon less than 20 dimensions, and where derivatives are not evaluated.[29]

Since the objective function is unknown, the Bayesian strategy is to treat it as a random function and place a prior over it. The prior captures beliefs about the behaviour of the function. After gathering the function evaluations, which are treated as data, the prior is updated to form the posterior distribution over the objective function. The posterior distribution, in turn, is used to construct an acquisition function (often also referred to as infill sampling criteria) that determines the next query point.

Standard Bayesian optimization relies upon each x∈ A being easy to evaluate, and problems that deviate from this assumption are known as exotic Bayesian optimization problems. Optimization problems can become exotic if it is known that there is noise, the evaluations are being done in parallel, the quality of evaluations relies upon a trade-off between difficulty and accuracy, the presence of random environmental conditions, or if the evaluation involves derivatives.

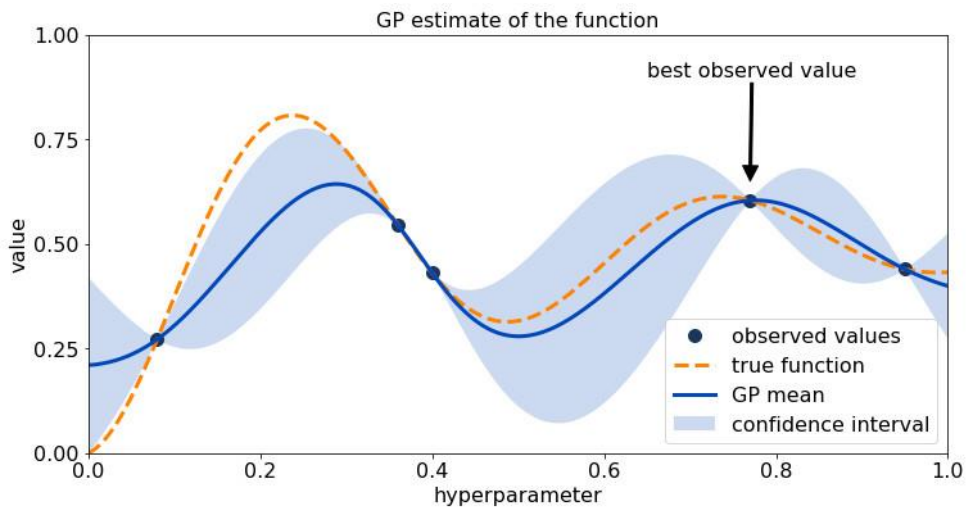

Figure 10 example related to optimise the parameters of the classification using Bayaze optimization

Chapter IV

real and imagination movement classification: an EEG study

4. Chapter 4: real and imagination movement classification: an EEG study

4.1 Introduction

The brain–computer interface (BCI) provides an information exchange and control channel independent of peripheral nerves and muscles for the brain and the outside world. It can directly read the physiological electrical signals in the human brain, analyse their meaning, and convert them into control signals to control external devices. Currently, neuroimaging techniques used in BCI systems commonly include functional magnetic resonance imaging (fMRI), electrocorticography (EcoG), magnetoencephalography (MEG) and electroencephalogram (EEG). EEG has become more and more popular because of its advantages such as low cost, easy portability and high temporal resolution [18].

Any EEG based BCI system include four main steps, starting from acquiring the EEG signal and amplifying and cleaning the data. Followed by the main core of any BCI system represented by data analysis which consist from feature extraction and classification procedure. Finally the output of the classification should be used based on the BCI application [30].

In any BCI system, cleaning the data from the different noises represents a critical stage, because any BCI system is highly effected by the data set quality. So that an automatic ICA procedure have been used to clean the data from the various artifacts. This method is introduced by [31] and widely applied in EEG signal processing because it is showed high ability to identify and classify artefactual non-cerebral activities.

The essence stage of any BCI system is represented by transforming the cleaned EEG system to matrix of features. the feature extraction process involves frequency filtering, windowing in which short segments are selected, feature extractor and the feature selection which outputs the selected features that are being fed into the classifier [32]. In BCI, frequency band power features and time domain features represent EEG signals. Band power features represent the power of EEG signals for a given frequency band averaged over a time window and time domain features are the combination of EEG signals from all channels. Motor Imagery BCI extensively uses band power features. Based on the literature found it has been noticed that most of the used or referenced techniques for feature extraction in motor imagery brain computer interfaces are Short Term Fourier Transform (STFT) [33] , Auto Regressive Model (AR) [34] , Wavelet Transform (WT) [35], and Common Spatial Pattern (CSP) [36].

The reason behind the suitability of the Common Spatial Pattern (CSP) features is that scholars have found that motor execution (ME) and motor imagery (MI) can change the neuronal activity in the primary sensorimotor areas . When humans execute or imagine the movement of unilateral limb, the power of mu and beta rhythms will decrease or increase in the sensorimotor area of the contralateral hemisphere and the ipsilateral hemisphere, respectively. The former case is called event-related desynchronization (ERD), and the latter event-related synchronization (ERS) [32].

The ERD/ERS patterns can be utilized as important features in the discrimination between right hand and left hand movement, and hand and foot movement. In the frequency bands varying between 9 and 14 Hz and between 18 and 26 Hz of EEG signals, the ERD/ERS patterns can provide best discrimination between left and right hand movement imagination, and the accuracy of online classification is more than 80% [32].

To discriminate between hand and foot movement, the online classification accuracy is even as high as 93%. Linear Discriminant Analysis (LDA) was used to classify ERS/ERD patterns associated with MI. Pfurtscheller et al. [37] used brain oscillations (ERS) to control an electrical driven hand orthosis (open or close) for restoring the hand grasp function. The subjects imagined left versus right hand movement, left and right hand versus no specific imagination, and both feet versus right hand, and achieved an average classification accuracy of approximately 65%, 75% and 95%, respectively.

In this study ERD in alpha (8-13) Hz, beta (14-25) Hz and (alpha + beta) range (8-25)Hz have been used to classify between the different tasks.

Another feature have been chosen to be implemented in this study is Fractal dimension (FD) , FD of a signal is a measure of its complexity and self-similarity in the time domain. [38]

FD is a statistical measure indicating the complexity of an object or a quantity that is self-similar over some region of space or time interval. It has been successfully used in various domains to characterize such objects and quantities [39] but its usage in motor imagery based BCI has been more recent . There are several FD estimation methods, some of which are not applicable to all types of data exhibiting fractal properties. In order to achieve a higher classification accuracy and speed, the FD estimation method that is most suitable to the data at hand should be chosen.

BCI operations have been said to depend mainly on effective interaction between two adaptive controllers the user who encodes his or her commands in the electrophysiological input provided to the BCI, and the computer which recognizes the command [40]. The development of translation algorithms solely relies on the classifiers like KNN, LDA, Neural Network and SVM. The performance of the classifier is highly effected by the EEG channels that are used in the BCI system.

In this study we aim to improve the quality of Motor Imagery system by creating a BCI system using features in time domain such FD and features in frequency domain such ERD with several channels combination and various classifier such as SVM, KNN, and TREE to classify between four real and imagery movement tasks. In order to compare between the different features and different channels combination and different classifier to evaluate their suitability to be used in a BCI system.

4.2 Materials and Method

4.2.1 Dataset

EEG Motor Movement/Imagery dataset called the PhysioNet was used [41]. This data set contains 1526 one and two seconds EEG signals acquired using the BCI2000 system from 109 volunteer subjects. The EEG signal were extracted from 64 channels complying with the international 10-20 system and sampled at 160 Hz. The placement of the electrode is shown in Figure 11.

Each subject performed 14 experimental runs, which can be divided into two runs were recorded during rest condition with eyes close and eyes open and called baseline runs, and twelve runs were recorded during the performance of the following tasks:

- **Task 1** (Real Hand Movement **RHM**): executing the opening and closing of the left or right hand as a response to the target appearance on the left or the right of the screen respectively then the subject relaxes.

- **Task 2** (Imagination Hand Movement **IHM**): imagining the opening and closing of the left and right hand as a response to the target appearance on the left or the right of the screen respectively then the subject relaxes.

- **Task 3** (Real Fists Or Feet Movement **RFM**): executing the opening and closing of both fists or rising up and down both feet as a response to the target appearance on the top or bottom of the screen respectively then the subject relaxes..

- **Task 4**(Imagination Fists Or Feet Movement **IFM**): imagining executing the opening and closing of both fists or the rising up and down both feet as a response to the target appearance on the top or bottom of the screen respectively then the subject relaxes.

Each task was performed 3 times by each subject. So that each subject will perform 14 tasks (Eyes open, Eyes close, RHM, IHM, RFM, IFM, RHM, IHM, RFM, IFM, RHM, IHM, RFM, IFM).

Furthermore, the data of 100 subjects have been used in this study by excluding the data of subject number 38, 88, 89, 92, 93, 94, 100, 104, and 106 because it contain incorrectly annotated data.

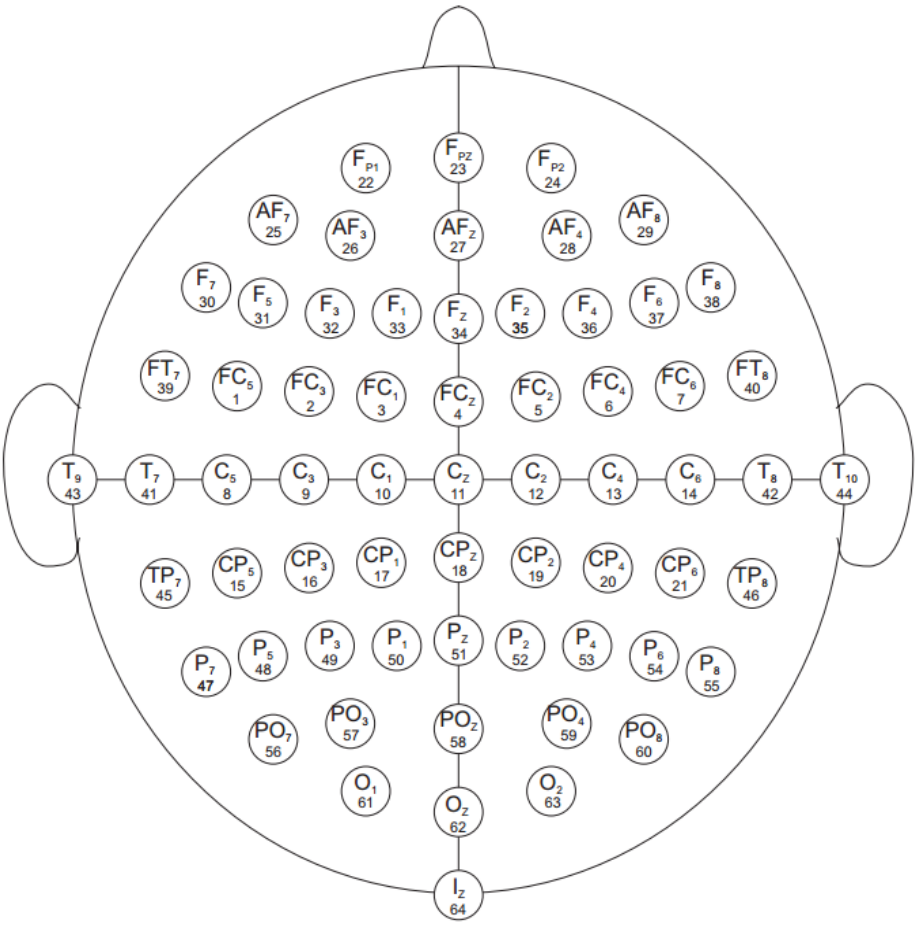

Figure 11 the location of the 64 EEG channels according to the 10-20 system

4.2.2 Methodology

To achieve the purpose of this study, three main steps should be performed, initially the data set should be cleaned from the internal and external noise and conducted in step called pre-processing, then the cleaned data should be used to extract features matrixes (ERD in alpha range, ERD in beta range, ERD in alpha + beta range and FD). Later on, the features matrixes will be the input of two types of the classification procedures based on the purpose of the study, classification between the different tasks, and classification between the different event in the same tasks, regardless the different classification purpose, each classification will be performed by using three different classifier SVM, KNN, and TREE. In addition to different channels compensation such as 64 channels, C3/C4, CP3/CP4, ROI include C3/C4, ROI include CP3/CP4 and optimization channels. Finally the accuracy will be used to compare between the variant models.

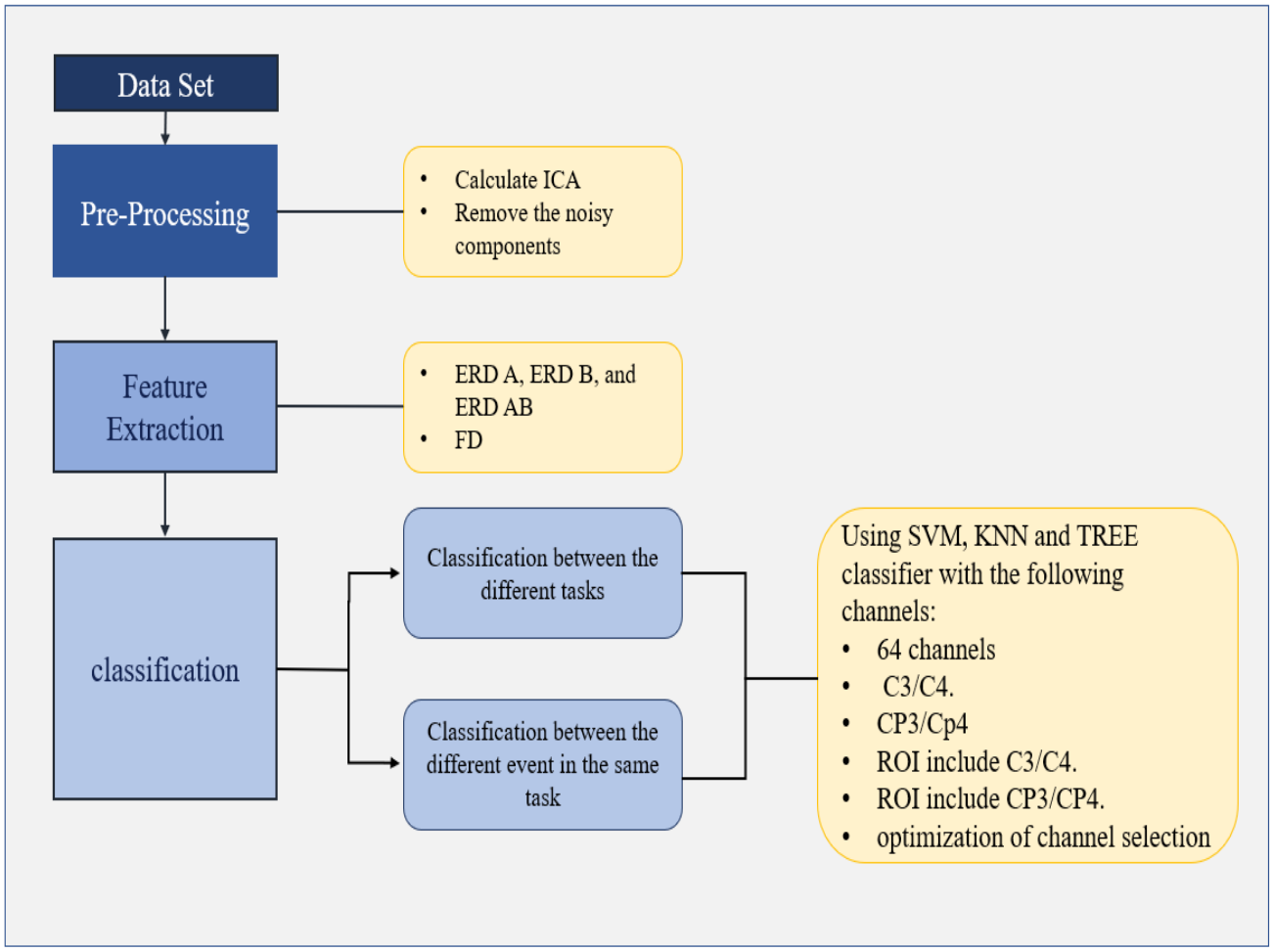

Figure 12 Schematic of the analysis pipeline

4.3 Pre-processing

Starting from the fact that each task was performed three time by each subject and include 30 event, the first step in the pre-processing procedure was to create a concatenation include the three trials of the same task by the same subjects with 90 events. To do that the *merge set function* in the EEGLAB v2019.1 has been used.

Moreover, to eliminate the environmental noise a band bass filer 1-70 Hz was applied. Followed by 60 Hz notch filter to eliminate the power line noise.

The data set is available in EDF+ formats, therefor connect the events with the corresponding data had to be done using *readedf function*.

Initially, the independent components of the data have been calculated by applying the independent components analysis ICA using Fastica function. Then SASICA toolbox have been used to select the noisy components and exclude them form the data and then the power spectrum density curve was plotted before and after the cleaning to check the cleaning quality and prevent the occurrence of overcleaning or undercleaning.

After the first cleaning, it have been noticed that the data set still include noise so that stronger cleaning have been done using Independent Component analysis (ICA), ICA like many other blind source separation (BSS) techniques, decomposes the EEG data into sources with independent time course on the basis of the statistical properties of the generated signal. we applied an automatic ICA procedure (an appropriately modified version of Barbati et al., 2004) to the Raw Data to identify and classify artefactual non-cerebral activities, i.e. eye movements, speech artifacts, environmental and channels noise.

This method is based on statistical and spectral characteristics of Independent Components(ICs) [31] . Briefly, it consists of three main steps:

1. Application of ICA for blind source separation (here we used fastICA .
2. Automatic detection of artefactual components, based on statistical characteristics (percentage of kurtosis-outlier segments to detect cardiac artifacts, global kurtosis coefficient to detect environmental noise and percentage of entropy outlier segments to detect ocular artifacts) and spectral IC characteristics (significant correlation between Power Spectral Density PSD EOG/EMG and ICs, $p<0.01$). More specifically, for the statistical indexes (kurtosis and entropy), these measure distributions were normalized with respect to all ICs (mean 0 and standard deviation 1). In this way, thresholds in terms of number of standard deviations from the mean were applied (standard

threshold set at ±1.64) and, if a certain percentage of segments (in our applications more than 33%) exceeded rejection thresholds, the corresponding IC was marked for rejection.

3. The final step was a control cycle on the 'discrepancy',i.e. on the difference between the original data and those reconstructed using only ICs retained after automatic artifact detection. The aim of the control cycle was to give a quick visual feedback on the quality of the automatic artifact identification. This step is based on visualizing PSD and ERP of the discrepancy. The feedback was positive when discrepancy contained only artifacts and noise. In case of negative feedback (i.e. presence of any brain activity in terms of brain rhythms or evoked activity in the discrepancy),the index thresholds used for artefactual IC detection was reduced and step (2) was repeated on these new thresholds.

In Figure 13 we can notice the effect of the cleaning on the quality of the data for subject 18.

The final step in the pre-processing procedure was to check the PSD of each signal manually to detect the noisy channels and replace them by the average of their adjacent channels.

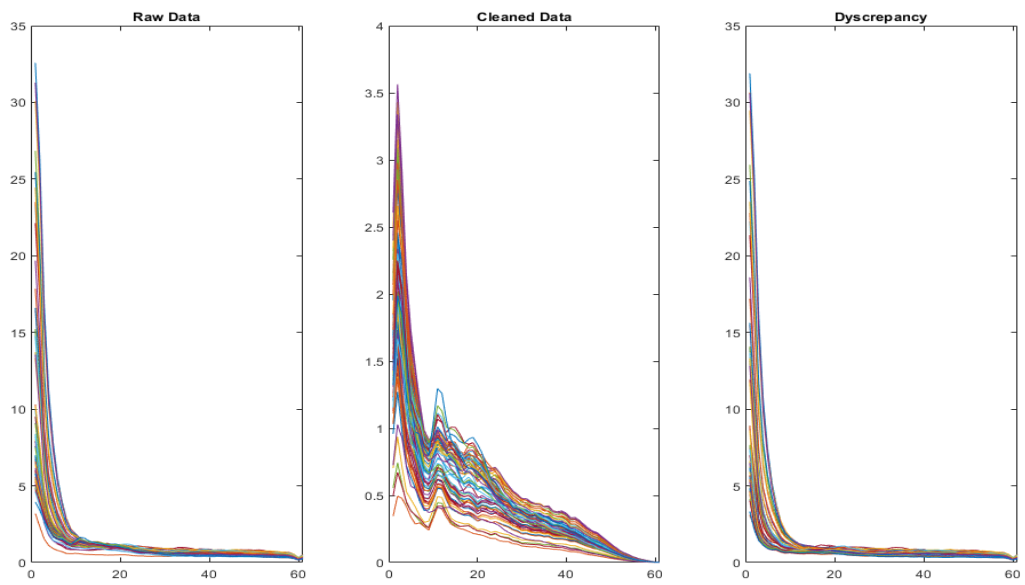

Figure 13 comparing the power spectrum density curve of the EEG signal of subject 18 before and after the cleaning and calculating the discrepancy to check the over cleaning or under cleaning.

4.4 Data analysis

4.4.1 Features extraction

4.4.1.1 Event related desynchronisation (ERD)

The EEG signal is composed of local and non-local rhythmic components. Local components can be considered as intrinsic activity of specific cortical areas. Non-local (i.e. global) EEG components are generated mainly in non-specific cortical areas and can be recorded over large parts of the scalp. Typical examples of local EEG rhythms are sensorimotor rhythms, including mu and central beta rhythms, and the occipital alpha rhythm. A characteristic of these local (i.e. intrinsic) rhythms is their close functional relationship to the state of the specific underlying networks. When such an area becomes activated, rhythmic activity in the alpha and lower beta band displays an amplitude attenuation or event-related desynchronisation (ERD).

The standard measure of ERD quantifies the induced change in signal band power as difference between a baseline prior to the event an post-event period. By convention an ERD corresponds to a negative value, i.e., a decrease in power, while event-related synchronization (ERS) refers to an increased signal power, changes of signal power are quantified only with respect to the deviation from a fixed, constant baseline level. [42]

The main advantage of the generalized ERD (gERD) measure is due to its ability to reliably study ERD responses even in the presence of dynamical cortical states, i.e., for the analysis of non-stationary dynamics. However, the generalized and the conventional measure of ERD yield identical results when analysing stationary dynamics and the superiority of gERD becomes evident only in case of non-stationary dynamics. One particular application field of gERD is the investigation of state conditional dependencies of ERD dynamics, i.e., investigation of explanatory factor of inter-trial variability[37].

For ERD computation, the EEG data of each channel were digitally filtered (finite impulse response filter) in the selected frequency bands, and the point-to-point inter-trial variance for the 8 s periods was computed as follows:

1. calculation of the mean across all filtered trials.
2. subtraction of the mean from each sample value and squaring of the differences.
3. calculation of the means squared differences across all trials.

The 1 s epoch before the presentation of the beep was defined as the reference interval and the ERD time course was computed as the percentage changes of the inter-trial variance (in time windows of -2 to 2 s) related to this reference interval.

This method of ERD computation was chosen to monitor non-phase-locked event-related EEG reactivity and to eliminate phase-locked EEG activities such as event-related potentials, due to the stimulation procedure.

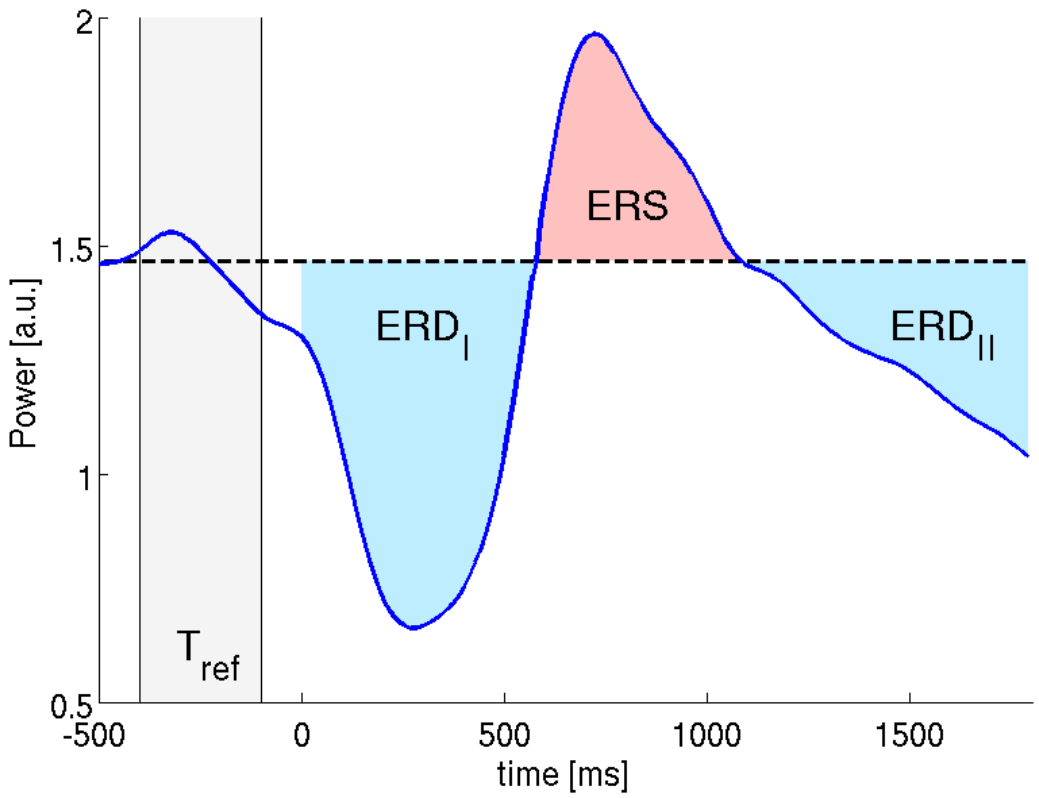

Figure 14 conventional ERD measures the relative deviation of the event related dynamics (solid) from a constant baseline level (dashed), the baseline is typically estimated as averaged power across a reference interval prior the event (Gray shaded region)

4.4.1.2 Fractal Dimension (FD)

FD is a highly sensitive measure for the detection of hidden information contained in physiological time series. in this study Higuchi's fractal dimension was calculated In practice, Higuchi's FD is a quantitative measure of signal dynamics. Linear methods commonly used for signal analysis, such as the well-known FFT and wavelet transformation (WT), are good choices if the analysed signals are stationary. However, neurophysiological processes are generally nonstationary and nonlinear by nature. Knowing that FD is an accurate numerical measure no matter what the properties (stationary, nonstationary, deterministic

or stochastic) of the analysed signal, it is reasonable to accept this advantage over widely used linear methods.

Higuchi's FD is a nonlinear measure of waveform complexity, here applied in the time domain. Discretised functions or signals can be analysed as segment of data X(1), X(2), . . . , X(N), where N is the total number of samples. From the starting time sequence, a new self-similar time series X_m^k can be calculated as

$$X_m^k: x(m), x(m+k), x(m+2k), \ldots., x(m + int\left(\frac{N-k}{k}\right)k) \qquad (1)$$

for m = 1, 2, . . . , k where m is the initial time; k is the time interval, k = 1, 2, . . . , k max; k max is a free parameter, and int(r) is the integer part of the number r.

The length, $L_m(K)$ of each curve X_m^k is calculated as

$$L_m(K) = \frac{1}{k}[\sum_{i=1,int(\frac{N-M}{k})} | X(m+ik) - X(m+(i-1)k | . \frac{N-1}{int(\frac{N-m}{k})}) \qquad (2)$$

where N is the length of the original time series X and (N − 1)/{int[(N − m)/k]k} is a normalization factor. $L_m(K)$ was averaged for all m forming the mean value of the curve length L(k) for each k = l, . . . , k max as

$$L(k) = \frac{\sum_{m=1}^{k} L_m(K)}{k} \qquad (3)$$

An array of mean values L(k) was obtained and the FD was estimated as

$$FD = ln\,(L(k)/ln(1/k)) \text{ for k = 1, 2, . . . , k max} \qquad (4)$$

the original curve or signal can be divided into smaller parts with or without overlap, called "windows". Then, the method for computing FD should be applied to each window where N should be seen as the length of the window. [43]

In that case, FD values are calculated for each window, with or without overlap. Individual FD values can be averaged across all windows for the entire curve (or data timeseries), and the mean FD value can be used as a measure of curve complexity.

4.4.2 classification

In order to find the classifier that give the best performance with the data set, three classifier were used, Support Vector Machine SVM [25], K-Nearest Neighbors KNN [26] and TREE [28] classifier.

The hyper parameters of each classifier have been optimised in a way that it minimize five-fold cross validation loss by using ***Bayesian optimization*** [29].

To cover the different purpose of the brain computer science, two different kinds of the classification have been done:

- ***Classification between the different tasks***

We used the data which has been acquired from RHM/ IHM and RFM/ IFM to classify between the EEG signal of the real and imagined movement of the same limp, while the data from RHM/ RFM to classify between the real movement of different limp, and the data from IHM/IFM to classify between two kinds of movement imagination of different limp.

Table 1 the goal of the classification for the classification between different tasks

The goal of the classification	The tasks companion that had been used
Classify between the real and imagination movement of the hand	RHM/ IHM
Classify between the real and imagination movement of the feet	RFM/ IFM
Classify between the real movement of the hand and the real movement of the feet	RHM/ RFM
Classify between the real and imagination movement of the feet	IHM/IFM

- ***Classification between the different events in the same task***

In this section, in RHM the classification distinguish between the real movement of the right and the left hand, in IHM the classification distinguish between the imagination of the movement of the right and the left hand, in RFM the classification distinguish between the real movement of the both hands and the real movements of the both feet, while in IFM the

classification distinguish between the imagination of the movement of the both hands and both feet.

Table 2 the goal of the classification for the classification between the different events in the same task.

The goal of the classification	The tasks companion that had been used
Classify between the real movement of the right and left hand	RHM (T1/T2)
Classify between the real and imagination movement of the right and left hand	IHM (T1/T2)
Classify between the real movement of both fest and both feet	RFM (T1/T2)
Classify between the real and imagination movement of both fest and both feet	IFM (T1/T2)

In general, 2 parameters can affect the performance of any classification procedure, the number of the features and the data set volume, in this paper those 2 parameters have been changed to get an optimal results.

4.4.2.1 channels selection

to detect the optimal groups of features which give the best performance of the model, various classification process have been performed with different group of channels, and it is include the using of the data of the 64 channels, using only the data of C3 and C4 channels, using only the data of CP3 and CP4 channels, using ROI include C3 and C4 (FC3, C5,C3, C1, CP3, C2 ,C4, C6,CP4) and using ROI include CP3 and CP4 (C3, CP5, CP3, CP1, P3, C13, CP2, CP4, CP6, P4) as it shown in figure 3 a, b, c and d respectively.

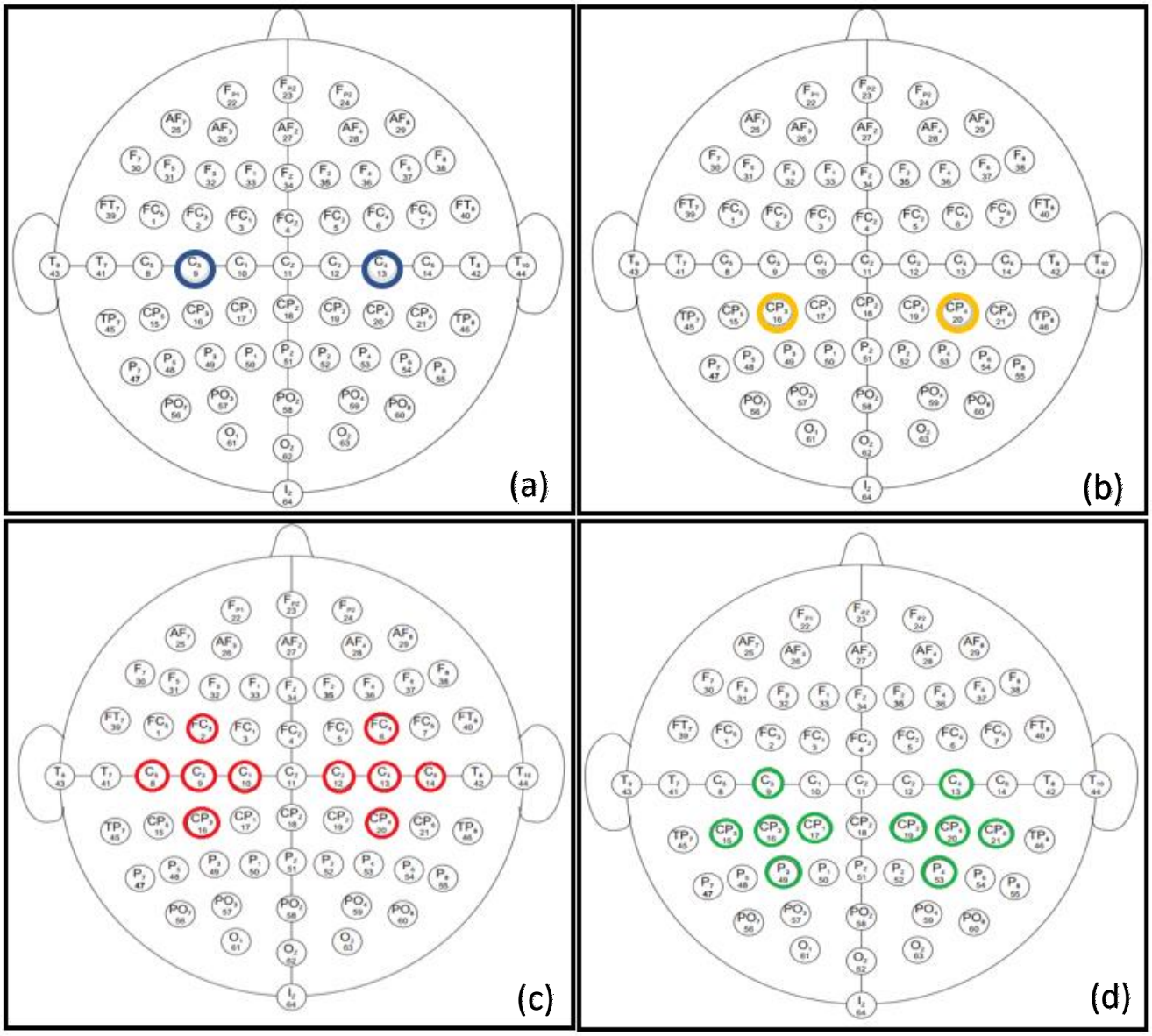

Figure 15 different combination of channels, (a) C3/C4, (b) CP3/CP4, (c) ROI include C3/C4 and (d) ROI include CP3/CP4

Moreover, a new feature optimization method have been proposed in this study. at variance to the previous channels collections which have been selected and fixed before the classification procedure, an optimization system have been created to select the optimal combination of channels which give the best model for each task, this optimization system work initially by compute the wight of each feature by using the relieff function and then give the descending order of the features based on their weight. Then 64 groups of features will be created, where the first group include only the first feature in the feature order vector, while the second group include the first and second features in the features order vector and so on until we reach the last group which include all the features. Later on a 64 classification procedure will be performed by using the 64 features group combination, and the features combination which give the highest accuracy will be selected as the optimal group of channels. The previous procedure is explained in Figure 16.

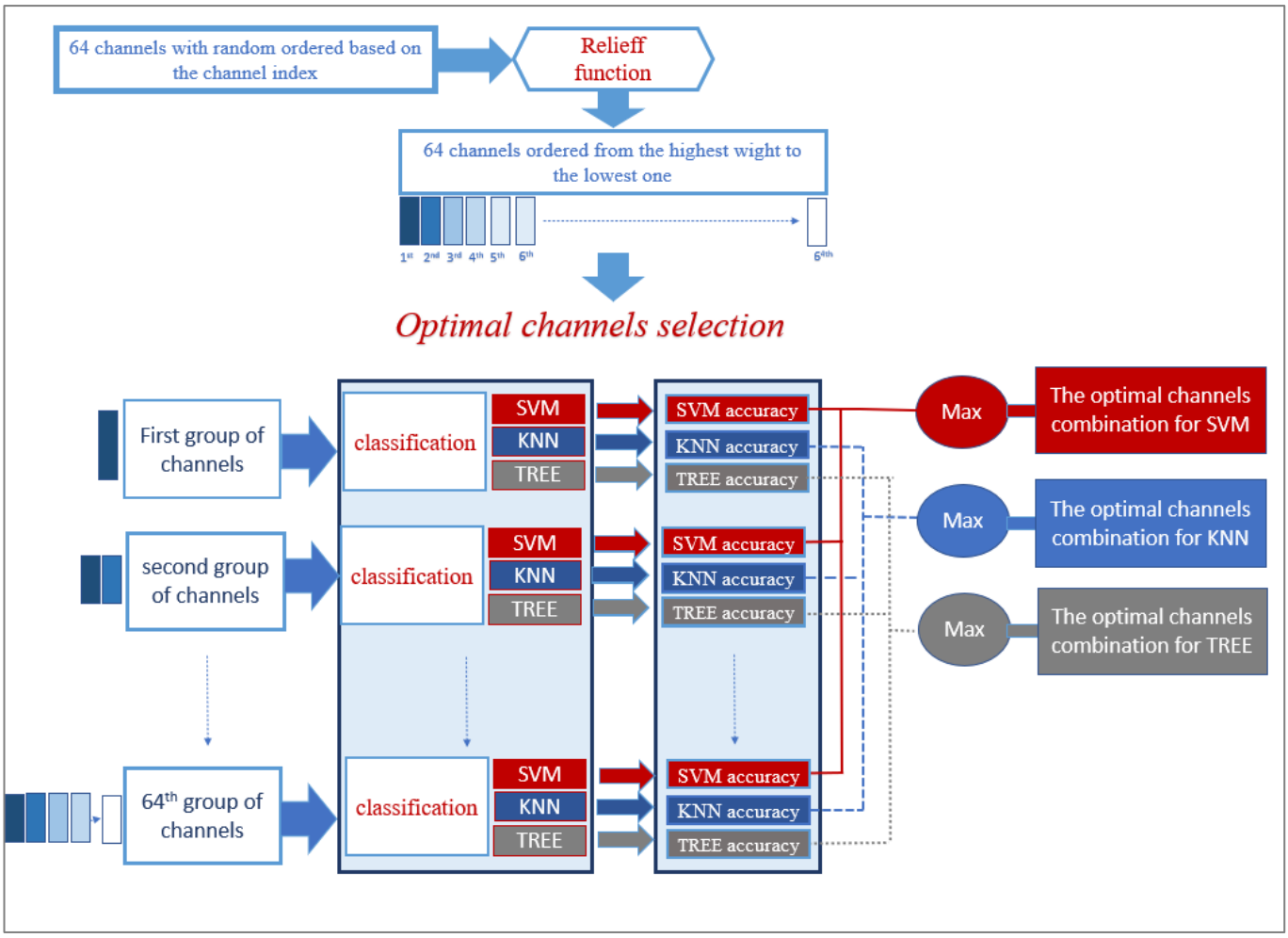

Figure 16 schematic explain the channel optimization process

4.4.2.2 Subject selection

The data set include 100 subjects and those subjects showed different response to the tasks. the ability to show clear EEG signal clearly vary a mong the subject group particularly during the imagination tasks, so on the average of the PSD in the range from 8 to 30 Hz which represents an independent variable has been selected to be the scale to evaluate each subject.

So that, for each task the PSD average for each subject have been calculated, then based on this value the subjects will be stored in descending order. Then the 100 subjects with the new order were divided into 5 groups each group contain 20 subjects. New classification model was created based on the best 20 subjects and compared to the model which have been created based on the 100 subjects.

It is been observed the different in PSD value among the subject during the different task, in addition to the observation that the PSD values for the different subject among IHM and IFM (the imagination tasks) are very close.

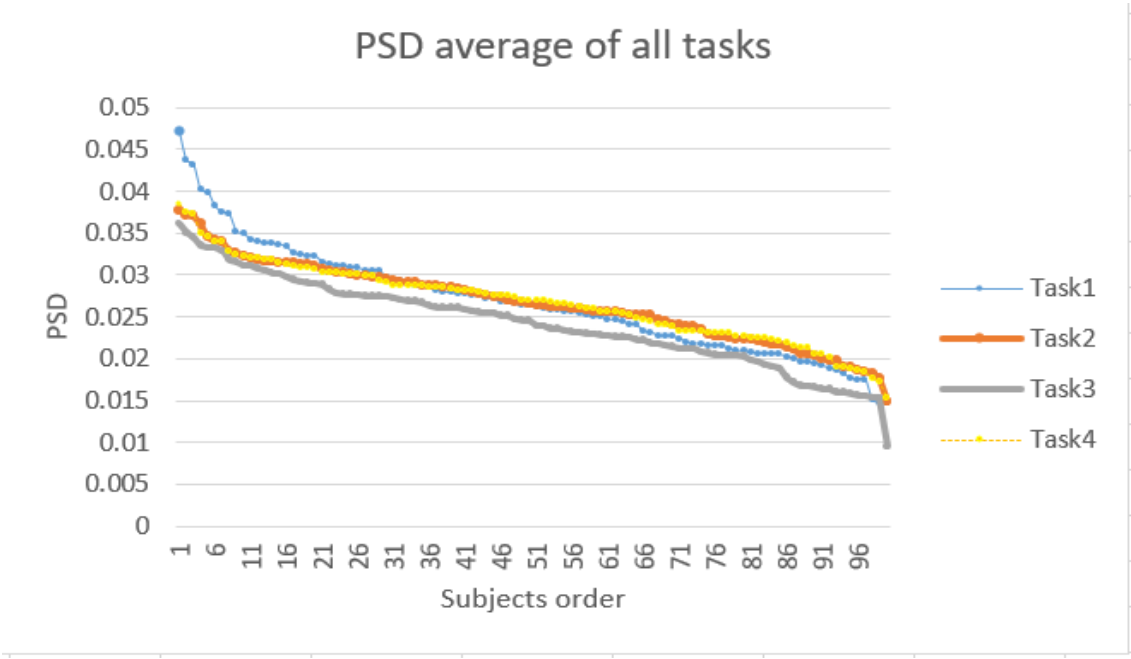

Figure 17 the 100subjects ordered from the highest PSD to the lowest one among the different tasks.

4.5 Results

4.5.1 the results of the classification between the different tasks

Initially, the classification between the tasks have been done by using the 64 channels and called 1st approach, the highest accuracy during the classification between RHM and IHM, RFM/ IFM, RHM/RFM and /IFM is equal to 97.3%, 92%, 95% and 89% respectively by using SVM and ERD AB.

While the second approach include the using of C3/C4 channels only, the highest accuracy during the classification between RHM and IHM, RFM/ IFM, RHM/RFM and IHM/IFM is equal to 84%, 84%, 80% and 56% respectively by using SVM and ERD AB.

The third approach include the using of CP3/CP4 channels only, the highest accuracy during the classification between RHM and IHM, RFM/ IFM, RHM/RFM and IHM/IFM is equal to 83%, 61%, 81% and 54% respectively by using SVM and ERD AB.

The fourth approach include the using of the ROI contains C3/C4, the highest accuracy during the classification between RHM and IHM, RFM/ IFM, IHM/IFM is equal to 92%, 82%, and 82% respectively by using SVM and ERD AB. And 96% to classify between RHM/RFM by using SVM and ERD B.

The fifth approach include the using of the ROI contains CP3/CP4 channels only, the highest accuracy during the classification between RHM and IHM, RFM/ IFM, and RHM/RFM and IHM/IFM is equal to 93.3%, 77%, 75% and 54% respectively by using SVM and ERD AB.

The fifth approach include the using of the ROI contains CP3/CP4 channels only, the highest accuracy during the classification between RHM and IHM, RFM/ IFM, and RHM/RFM and IHM/IFM is equal to 93.3%, 77%, 75% and 54% respectively by using SVM and ERD AB.

The sixth approach include the using of the optimization procedure, this method showed the highest accuracy among the different approach, , the highest accuracy during the classification between RHM and IHM, RFM/ IFM, and RHM/RFM and IHM/IFM is equal to 98.5%, 94%, 96% and 91% respectively by using SVM and ERD AB.

By comparing the different classifiers, SVM showed the highest accuracy during the different approaches, followed by KNN and TREE respectively.

By comparing the purpose of the classification procedure, distinguishing between RHM/IHM showed the highest accuracy among the different classifiers and different approaches. While distinguishing between IHM/IFM showed the lowest accuracy among the different classifiers and different approach.

Table 3 results of the classification between the different tasks (T1 is RHM, T2 is IHM, T3 is RFM and T4 is IFM)

	Features	SVM				KNN				TREE			
		T1/T2	T3/T4	T1/T3	T2/T4	T1/T2	T3/T4	T1/T3	T2/T4	T1/T2	T3/T4	T1/T3	T2/T4
64 channels	FD	88	83	89	84	84	75	89	75	74	73	73	71
	ERD A	85	84	87	83	72.5	70	83	67	66	62	66	60
	ERD B	95	91	94	82	86	73	***89***	67	81	71	79	60
	ERD AB	***97***	92	95	89	81	76	87	71	***90***	73	88	63
C3/C4	FD	70	73	87	74	69	73	89	75	71	70	74	71
	ERD A	50	59	62	56	51	57	61	57	52	55	60	54
	ERD B	72	61	65	58	72	63	63	60	70	62	58	56
	ERD AB	***84***	84	80	56	81	***82***	78	55	***84***	81	77	55
CP3/CP4	FD	71	82	70	**83**	72	74	68	**76**	71	75	65	**72**
	ERD A	56	58	62	57	54	56	60	55	52	58	58	57
	ERD B	70	65	68	54	70	67	65	55	68	65	65	55
	ERD AB	***83***	61	81	54	***82***	60	78	54	***80***	60	75	55
ROI include C3/C4	FD	86	80	92	84	78	81	83	**77**	72	70	72	**74**
	ERD A	78	83	87	79	81	79	88	68	61	61	67	60
	ERD B	92	86	***96***	75	78	77	82	68	74	69	78	58
	ERD AB	92	82	94	82	***89***	79	87	73	***86***	66	83	60
ROI include CP3/CP4	FD	84	76	86	**75**	74	71	83	**74**	72	70	70	**74**
	ERD A	69	75	79	67	63	66	71	66	56	56	62	56
	ERD B	86	79	91	65	75	66	87	65	74	68	73	60
	ERD AB	***93***	77	78	75	***90***	71	77	71	***82***	61	60	68
Optimized channels	FD	91	95	82	86	93	80	94	79	75	78	79	76
	ERD A	89	88	91	85	80	76	89	75	73	69	72	69
	ERD B	95	93	95	86	85	82	90	70	84	76	83	65
	ERD AB	***98***	94	96	91	88	85	88	***90***	***93***	87	91	68

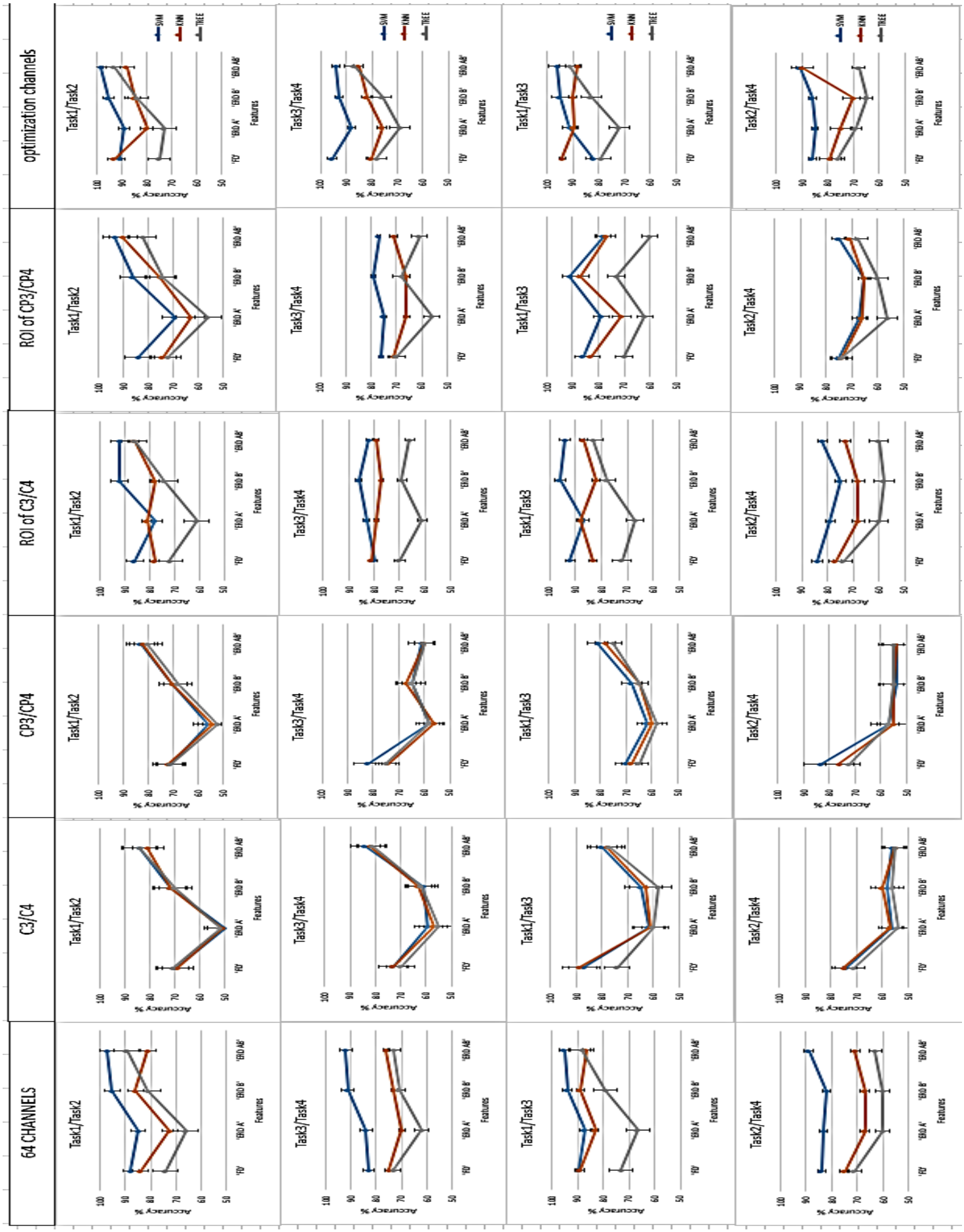

Figure 18 classification results between the different tasks using 6 different approaches for channels selection

- **Receiver Operating Characteristic (ROC) curve and Confusion Matrix**

By drawing the ROC curve for the best model to classify between RHM and IHM by using the optimal channels combination of the ERD AB and SVM it have been shown that the sensitivity = 97%, specificity=99% and Area under the curve (AUC)=0.99.

While the confusion matrix presents that true Positive (TP) = 194/200 , True Negative (TN) = 199/200, False Positive (FP)= 1/200, and False Negative (FN)= 6/200.

By drawing the ROC curve for the best model to classify between RFM and IFM by using the optimal channels combination of the ERD AB and SVM it have been shown that the sensitivity = 92%, specificity=90% and AUC =0.96.

While the confusion matrix presents that TP = 184/200 , TN= 181/200, FP= 19/200, and FN= 16/200.

By drawing the ROC curve for the best model to classify between RHM and RFM by using the optimal channels combination of the ERD AB and SVM it have been shown that the sensitivity = 91%, specificity=99% and AUC = 0.97.

While the confusion matrix presents that TP = 181/200 , TN= 197/200, FP= 3/200, FN= 19/200.

By drawing the ROC curve for the best model to classify between RFM and IFM by using the optimal channels combination of the ERD AB and SVM it have been shown that the sensitivity = 87%, specificity=86% and AUC = 0.93.

While the confusion matrix presents that TP = 174/200 , TN= 171/200, FP= 29/200, and FN= 26/200.

Table 4 sensitivity, specificity and AUC for classification between different tasks

	Sensitivity	**Specificity**	**AUC**
RHM/IHM	97%	99%	0.99
RFM/IFM	92%	90%	0.96
RHM/RFM	91%	99%	0.97
IHM/IFM	87%	86%	0.93

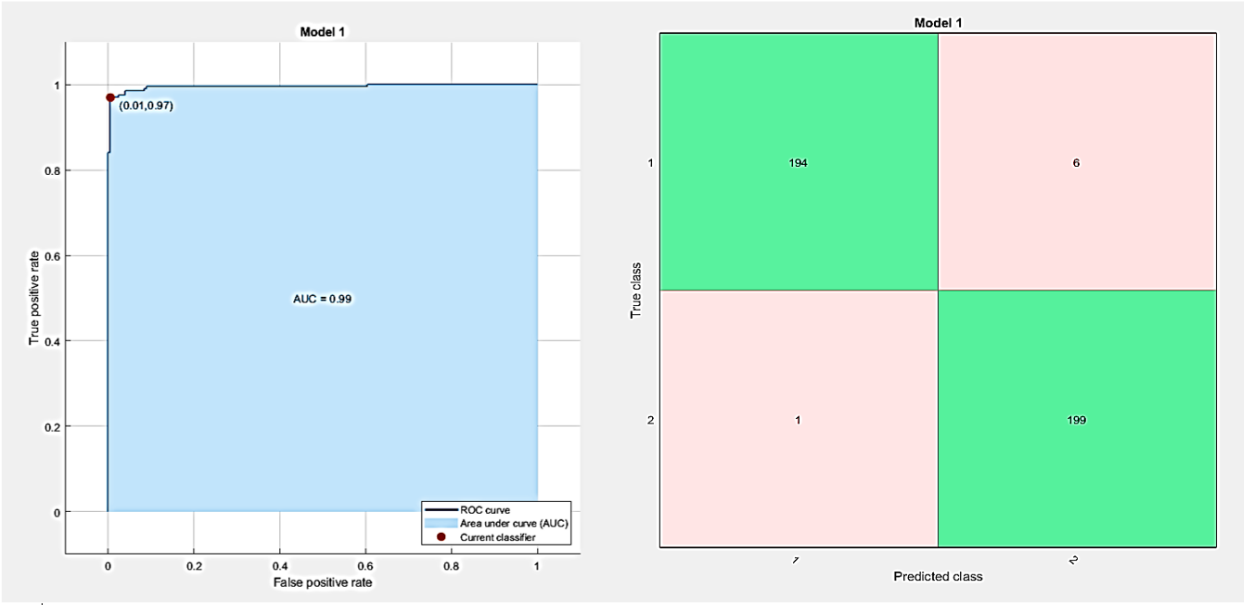

Figure 19 the ROC curve and the confusion matrix for the classification between RHM and IHM by using the optimal channels and ERD AB with SVM.

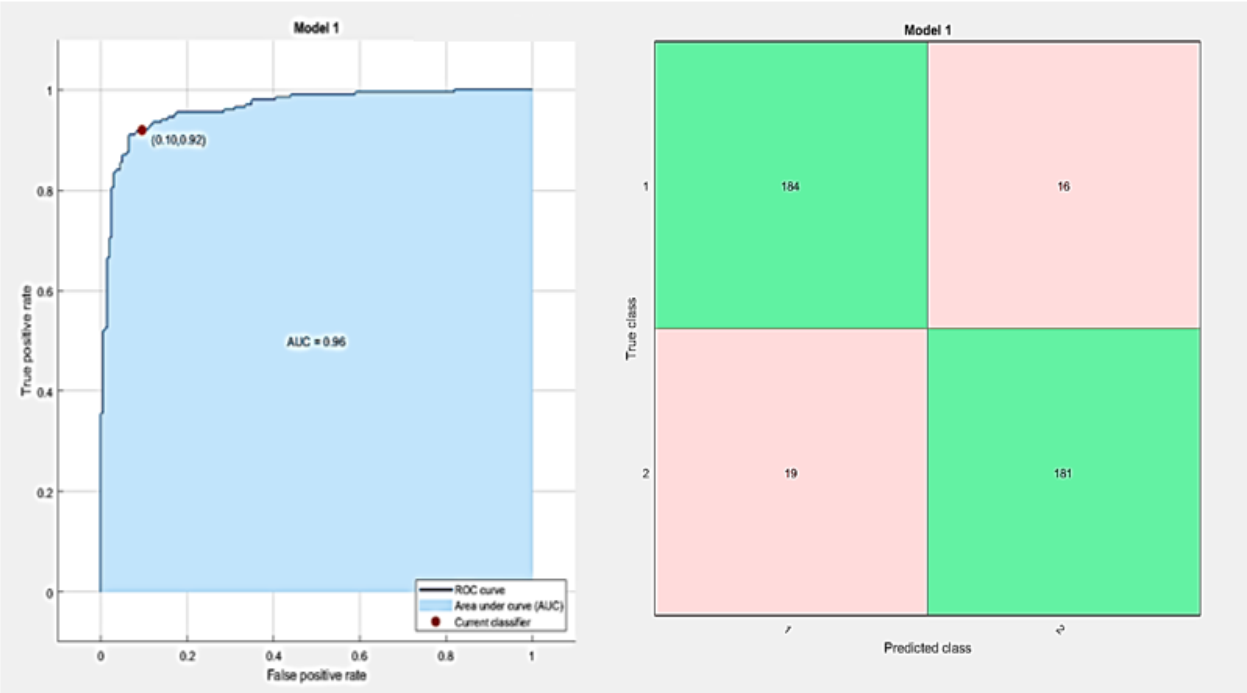

Figure 20 the ROC curve and the confusion matrix for the classification between RFM and IFM by using the optimal channels and ERD AB with SVM.

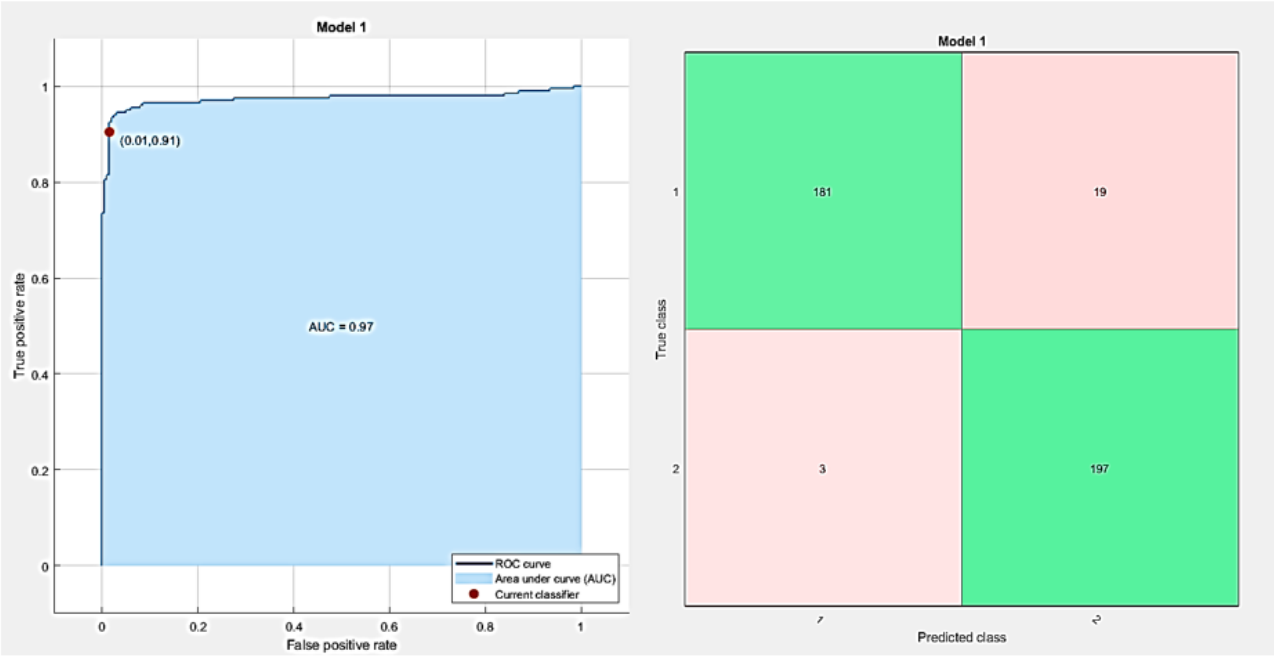

Figure 21 the ROC curve and the confusion matrix for the classification between RHM and RFM by using the optimal channels and ERD AB with SVM.

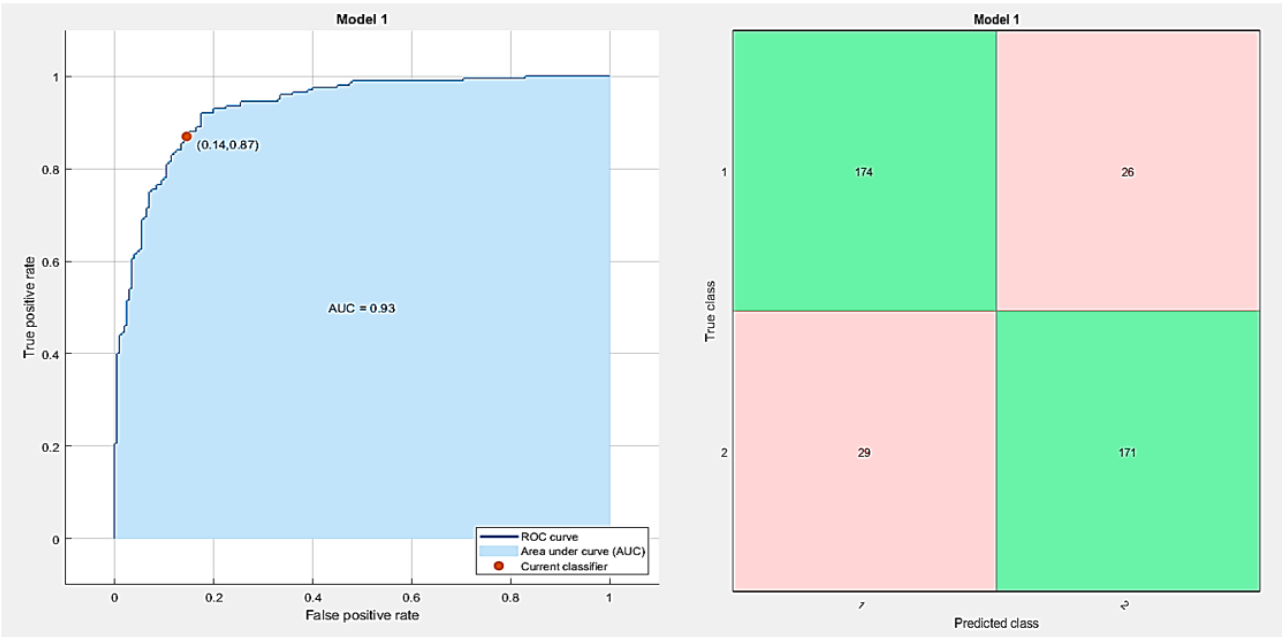

Figure 22 the ROC curve and the confusion matrix for the classification between IHM and IFM by using the optimal channels and ERD AB with SVM.

4.5.2 the results of the classification between the different events in the same task

By comparing the accuracy of classification between the tasks and the accuracy of classification between the different events of the same task, it can be clearly noticed that classification between different events of the same task is more difficult mission, Thus to improve the accuracy we create 2 types of models, one have been created based on the entire subjects group, and the other have been created based only on the data of the best 20 subjects.

- **Using the data of 100 subjects**

By comparing between the different features, the highest accuracy was reached by using FD and it is equal to 79%, 71% and 75% for the classification of the events on RHM, RFM and IFM respectively using KNN classifier. And accuracy equal to 74% % for the classification of the events on IHM using SVM classifier by using the optimal channels selection.

By comparing between the channels combination using ERD A as a features the highest accuracy is reached by using the optimal channels selection with values equal to 75%, 69%, 62% for the classification of the events on RHM, IHM and RFM respectively using KNN classifier, and 70% for classifying the event in IFM using SVM.

by comparing between the channels combination using ERD B as a features the highest accuracy is reached by using the optimal channels selection with values equal to 63%, 72%, 67% for the classification of the events on RHM, IHM and IFM respectively using KNN classifier, and 63% for classifying the event in RFM using SVM.

by comparing between the channels combination using ERD AB as a features the highest accuracy is reached by using the optimal channels selection with values equal to 58%, 71%, 62% for the classification of the events on RFM, and IFM respectively using KNN classifier, and 64% and 72% for classifying the events in IFM and IHM respectively using SVM.

Table 5 results of classification between different event in the same task for 100 subject model

		RHM			IHM			RFM			IFM		
		SVM	KNN	Tree	SVM	KNN	Tree	SVM	KNN	Tree	SVM	KNN	Tree
FD	64 channel	72	78	62	70.5	69	64	67	67	55	65	68	65
	Opt channels	*75*	**79**	72	***74***	72	68	67	***71***	65	71	***75***	67
	ROI C3/C4	72	65	60	63	67	60	63	65	55	66	65	65
	ROI CP3/CP4	62	65	59	60	62	57	60	60	53	66	67	66
ERD A	64 channel	56	58	54	55	62	60	49	51	52	63	62	61
	Opt channels	66	***75***	61	69	***69***	61	60	***62***	60	***70***	66	67
	ROI C3/C4	55	58	58	60	58	50	58	55	57	65	63	64
	ROI CP3/CP4	64	67	59	62	65	55	59	60	59	68	65	65
ERD B	64 channel	52	53	52	68	62	54	54	57	56	64	62	54
	Opt channels	62	***63***	59	69	***72***	62	60	62	***63***	66	***67***	63
	ROI C3/C4	54	53	53	60	65	58	53	59	58	63	62	58
	ROI CP3/CP4	53	57	59	63	70	54	55	60	62	65	65	60
ERD AB	64 channel	56	58	50	67	69	60	54	52	50	63	67	62
	opt channels	***64***	61	57	***72***	72	66	56	***58***	60	70	***71***	66
	ROI C3/C4	53	56	48	65	60	50	52	54	55	65	65	63
	ROI CP3/CP4	57	57	49	66	68	59	55	55	58	68	68	65

- **Using the best 20 subjects**

By comparing between the different features, the highest accuracy was reached by using FD and it is equal to 85%, 90%, 85% and 87% for the classification of the events on RHM, IHM, RFM and IFM respectively using KNN classifier and the optimal channels selection.

by comparing between the channels combination using ERD A as a features the highest accuracy is reached by using the optimal channels selection with values equal to 83%, and 80% for the classification of the events on RFM, IFM respectively using KNN classifier, and 80% and 80% for classifying the event in RHM and IHM respectively using SVM.

by comparing between the channels combination using ERD B as a features the highest accuracy is reached by using the optimal channels selection with values equal to 78%, 78%, 83% and 81% for the classification of the events on RHM, IHM , RFM and IFM respectively using KNN classifier .

by comparing between the channels combination using ERD AB as a features the highest accuracy is reached by using the optimal channels selection with values equal to 75%, 83%, 81% for the classification of the events on RHM, RFM ,and IFM respectively using KNN classifier, and 80% and 72% for classifying the event in IHM using SVM.

Table 6 results of classification between different event in the same task for 20 subject model

		RHM			IHM			RFM			IFM		
		SVM	**KNN**	**Tree**	**SVM**	**KNN**	**Tree**	**SVM**	**KNN**	**Tree**	**SVM**	**KNN**	**Tree**
FD	**64 channel**	67	55	57	67	65	52.5	72	62	55	70	60	60
	Opt channels	83	***85***	70	88	***90***	67	83	***85***	70	82	***87***	70
	ROI C3/C4	75	80	53	80	75	65	75	76	58	75	65	52
	ROI CP3/CP4	70	70	52	67	72	62	70	65	43	67	65	52
ERD A	**64 channel**	80	73	60	60	73	45	68	75	55	65	70	50
	Opt channels	***80***	78	75	***80***	78	60	78	***83***	65	79	***80***	73
	ROI C3/C4	70	63	65	73	60	50	67	72.5	45	60	67	48
	ROI CP3/CP4	73	63	70	75	70	55	67	75	60	77	75	55
ERD B	**64 channel**	68	75	63	70	68	60	73	70	55	65	78	62.5
	Opt channels	75	**78**	*75*	78	***78***	70	81	***83***	70	80	***81***	75
	ROI C3/C4	55	60	55	70	55	47	77	73	53	75	68	63
	ROI CP3/CP4	60	65	58	77	68	48	78	81	60	77	73	65
ERD AB	**64 channel**	65	67	55	75	70	57.5	65	75	58	67	72	65
	Opt channels	73	**75**	*73*	***80***	73	70	78	***83***	73	80	***81***	75
	ROI C3/C4	55	60	60	75	65	55	60	67	55	70	75	60
	ROI CP3/CP4	65	65	63	75	77	45	68	67	65	75	75	72

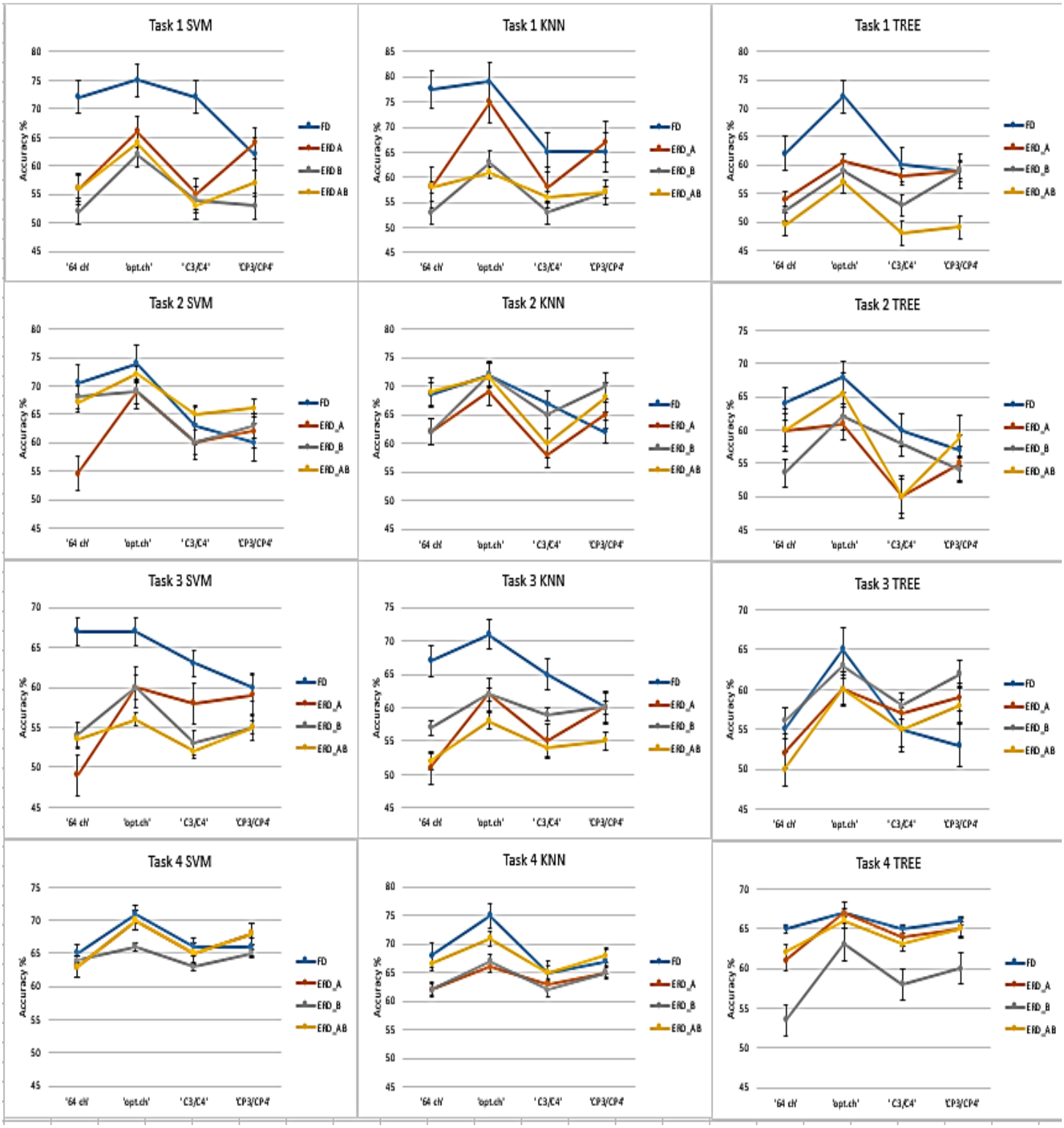

Figure 23 classification results to distinguish between the different event in the same task using 100 subjects and four approach for features selection

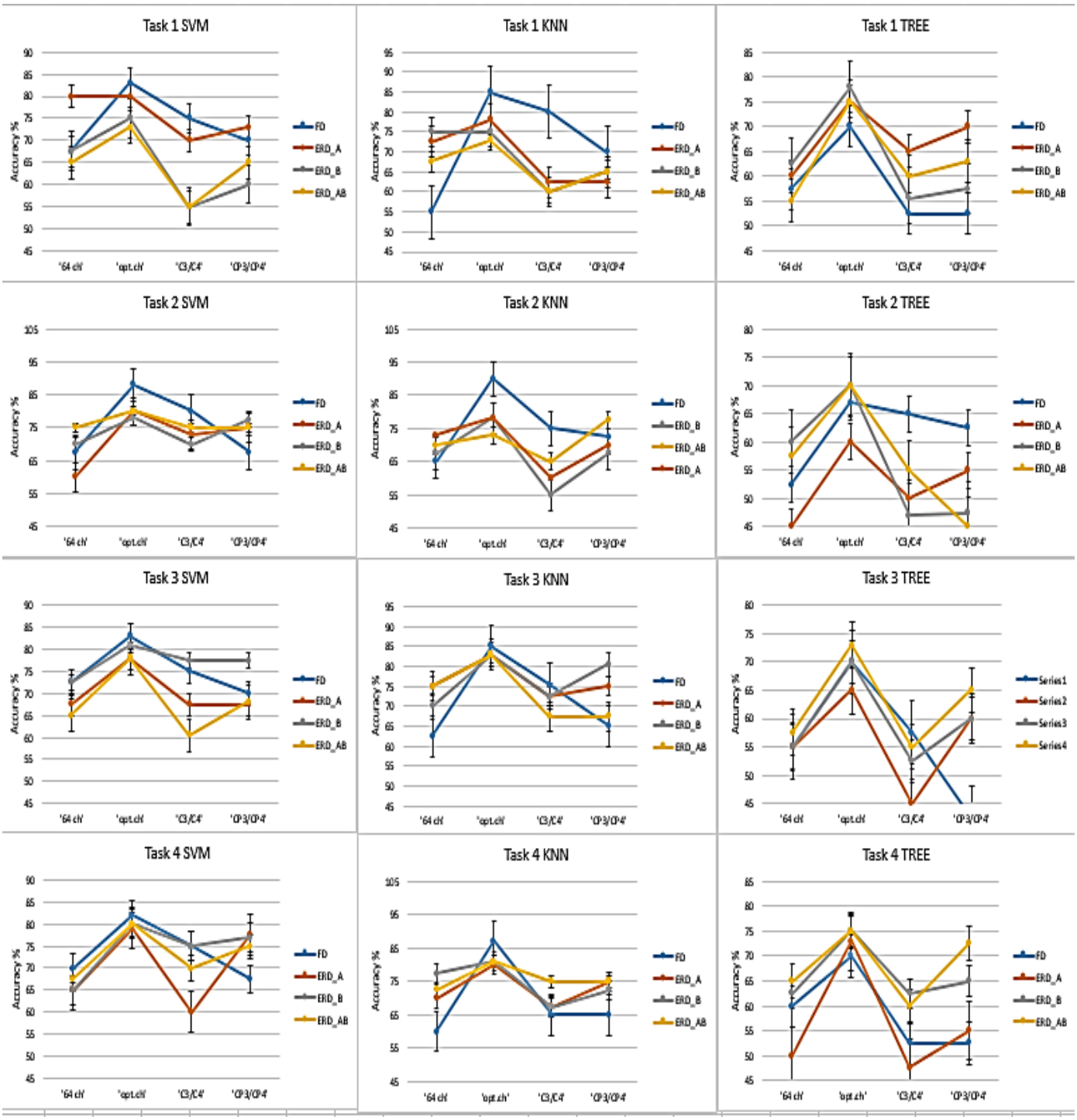

Figure 24 classification results to distinguish between the different event in the same task using the best 20 subjects and four approach for features selection represented in a way that compares the features effects

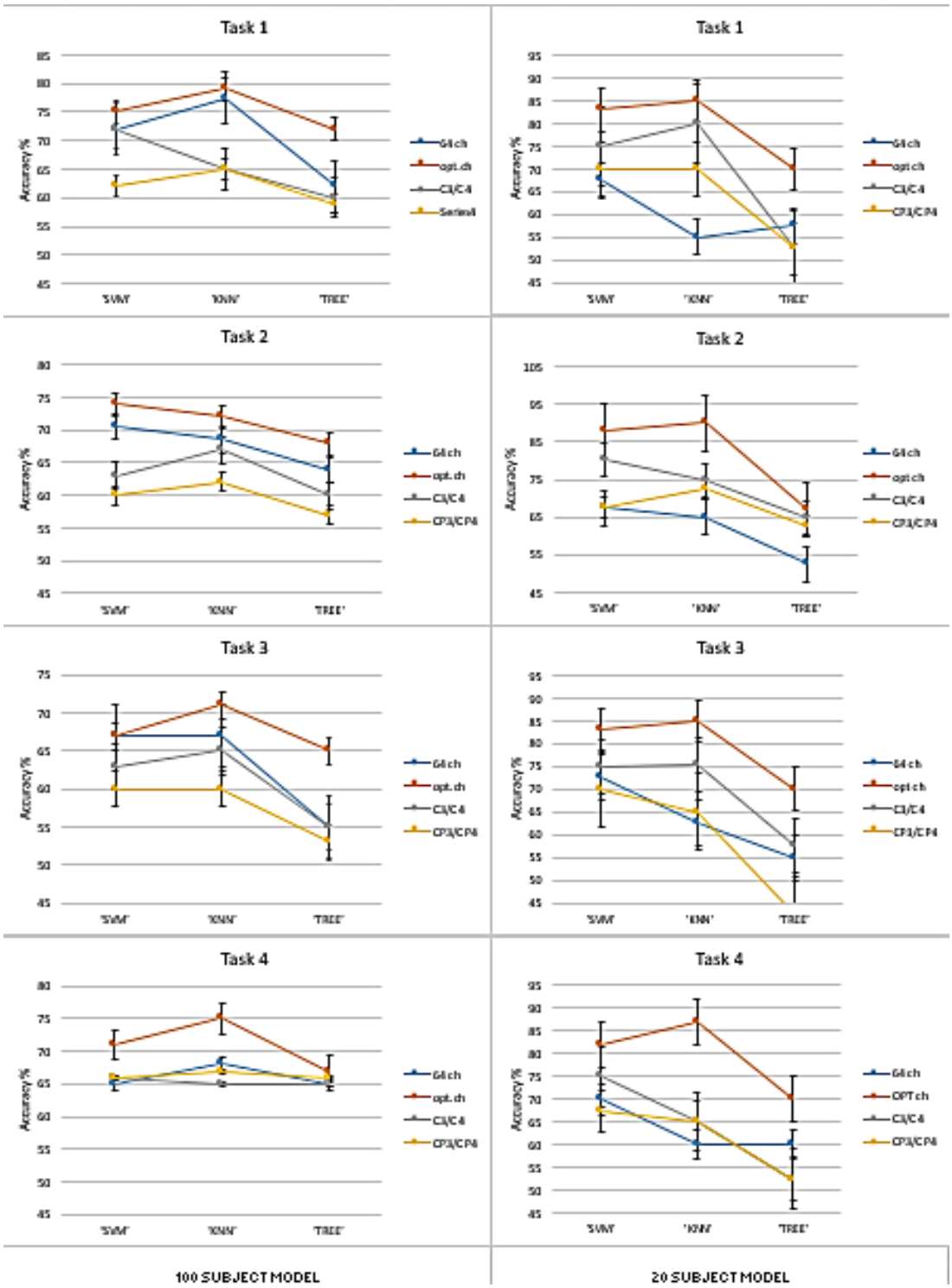

Figure 25 classification results to distinguish between the different represented in a way that compare the classifiers performance

4.6 Discussion and conclusion

This study aimed to improve the accuracy of motor imagery classification to help people with limited motor abilities interact with their environment using brain–computer communication. a PhysioNet data set have been used [41], the 100 subject data set include EEG signals acquired during the performance of four different tasks, RHM is a real movement of the right or left hand, IHM is the imagination of the movement of the right or left hand, RFM is the real movement of both feet or both hands and IFM is the imagination of the movement of both feet or both hands. The aim of this study was to compare between different set of features (ERD in alpha, Beta and Alpha + beta range and FD) and different combination of EEG channels (64 channels, C3/C4, CP3/CP4, ROI include C3/C4, ROI include CP3/CP4 and optimal channels combination) and different classifier (SVM, KNN, and TREE) in order to evaluate their performance in a BCI system.

For classifying between the different tasks. ERD in alpha plus Beta range give the highest accuracy be selecting the channels using the optimal channels combination and SVM as a classifier. (table 3 , Figure 18)

Among the different tasks, classifying between RHM and IHM give the highest accuracy 98%, while comparing between IHM and IFM give the lowest accuracy 91% by using the same features and the same manner to select the channels. This result confirm the fact that classifying between imagination task represent the most challenging mission. (table 3)

Moving to the level of classifying between the different events in the same tasks reflects the high influence of the subject performance of the task on the classification accuracy. In particular for the imagination Tasks (IHM and IFM) a noticeable number of subjects were not able to perform the imagination task in a good way which highly effect the classification procedure.

So that, creating a model by selecting 20 subjects with the best task performance increase the accuracy of the model. (table 5 and table 6). This observation was confirmed for the same data set in a previous study [44].

The highest accuracy for classifying between the different events in the same tasks was obtained by using FD and the optimal channels combination and KNN classifier. This method showed better performance for classifying the events in the imagination task (90% for IHM, 87% for IFM) than real movement tasks (85% for both RHM and RFM).

By comparing the accuracy of our model with other studies who used the same data set, we obtained higher accuracy than Root et al 2020 [45], where CNN have been used to classify between the right and the left hand imagination of the movement of IHM and they got accuracy equal to 83.8 % while in this study using FD and the optimal channels combinations with KNN classifier we obtained accuracy equal to 90% to classify the same task.

Several deep learning study used CNN to classify between events in IHM in addition to event in IFM, Lune et al 2020 [44], select only 10 subjects and got accuracy equal to 96%. Dose et al 2018 [46] got accuracy equal to 86%, and Karácsony et al 2019 [47] reached 85.9% while we classify using 100 subjects and reach accuracy equal to 90% using the optimal channels combination and ERD AB and KNN as a classifier.

Moreover, Alomari et al, 2013 [48] used MRCP and NN to classify between RHM and IHM only on the first 6 subjects and the obtained accuracy equal to 89% , while in this study by using the data of 100 subjects by using ERD AB as a feature and optimal channels combination and SVM classifier we obtained accuracy equal to 98% to classify between the same tasks.

Comparing to another study which used time domain feature, Sleight et al 2009 [49], used EEG averaged power and SVM to classify between RHM and IHM and they got accuracy equal to 69% while in this study by using ERD AB and optimal channels combination and SVM classifier we obtained accuracy equal to 98%.

From channels combination point of view, in this study we proposed a new mothed to obtain the optimal channels combination. This method able to obtain the highest accuracy comparing to fixed channels selection. However in some specific BCI cases, it is necessary to use to minimum number of channels. Therefore, in this study we proposed a various fixed channels selection with 2 channels and 10 channels able to give acceptable accuracy. So the number of channels can be compromised based on the application purpose.

From computation time point of view, our model represents huge improvements in testing time (5s per sample) comparing to other system which used conventional neural network (107 s per sample) [44]. Moreover even for training time, our model require around 60 min to calculate the optimal channels combination for each feature and each classifier and training the model while for systems with deep learning algorithm several days are required for training the model.

Moreover, in this study we confirmed the fact that the performance of the classifier is highly effected by the purpose of the study[50]. For classifying between the different tasks SVM showed the best performance comparing to the other classifiers (Figure 18, table 3). In contrast KNN give the highest accuracy for classifying between the different events in the same task (Figure 25, table 6).

Chapter V

Conclusion

5. Conclusion

In this study, motor imagery data set includes four real and imagination hand and feet movements tasks (RHM, IHM, RFM, IFM) was used to create a BCI system employing several techniques to investigate the EEG signal based on its attitude in different domains. FD and ERD were tested using SVM, KNN and TREE classifiers.

Furthermore, to examine the influence of the channels on the model accuracy, 6 groups of channels have been created, 64 channel C3 and C4 channels, CP3 and CP4 channels, using ROI include C3 and C4 and using ROI include CP3 and CP4, in addition to new optimization technique proposed for the first time to select the suitable group of channels for each task .

Two kinds of classification procedure have been done based on the purpose of the study, classification between the different tasks and classification between the different event in the same task.

For classification between different tasks, ERD in (8-25) Hz showed the highest accuracy using SVM and the optimal channels combination. To classify between RHM/IHM, RFM/IFM, RHM/RFM and IHM/IFM with accuracy equal to 98%, 94%, 96% and 91% respectively.

For classification between different events in the same task two steps have been done, creating the classification model using 100 subjects and create model using only 20 subjects by selecting the subjects with best test performance based on the value of their PSD.

In the 100 subjects model, The highest accuracy achieved by using FD and the optimal channels selection. With values equal to 79%, 74%, 71% and 75% to classify between the events in RHM, IHM, RFM and IFM respectively.

In the 20 subjects model, The highest accuracy achieved by using FD and the optimal channels selection. With values equal to 85%, 90%, 85% and 87% to classify between the events in RHM, IHM, RFM and IFM respectively.

This study proposed that selecting the channels with optimization procedure improve the performance of the classification process. Moreover, classifying on the events level is more influenced by the quality of the data and creating the model starting from data with good quality will improve the quality of the resultant model.

Chapter VI

References

6. References

1. Bougea A, Anagnostou E, Konstantinos G, George P, Triantafyllou N, Kararizou E. A Systematic Review of Peripheral and Central Nervous System Involvement of Rheumatoid Arthritis, Systemic Lupus Erythematosus, Primary Sjögren's Syndrome, and Associated Immunological Profiles. Int J Chronic Dis. 2015;2015:1–11.

2. Du W, Li S, Wang Z. Research on the human brain, the external brain and the public external brain. J Phys Conf Ser. 2019;1168.

3. Judaš M, Cepanec M, Sedmak G. Brodmann's map of the human cerebral cortex - Or Brodmann's maps? Transl Neurosci. 2012;3:67–74.

4. Kumar JS, Bhuvaneswari P. Analysis of electroencephalography (EEG) signals and its categorization - A study. Procedia Eng. 2012;38:2525–36.

5. Subha DP, Joseph PK, Acharya U R, Lim CM. EEG signal analysis: a survey. J Med Syst. 2010;34:195–212.

6. Ihc F, Deshmukh MD, Kamble S. International Journal of Advanced Research in Survey on Hierarchical Document Clustering Techniques. 2013;3:157–61.

7. Savelainen A. An introduction to EEG artifacts. Perception [Internet]. 2010;1–67. Available from: http://www.sciencedirect.com/science/article/B8KVH-4SW8G48-2/2/72b76e2ec9f9d2337906fa5bc0ec731c

8. Mehmood RM, Lee HJ. Exploration of Prominent Frequency Wave in EEG Signals from Brain Sensors Network. Int J Distrib Sens Networks. 2015;2015.

9. Ma P, Gao Q. EEG Signal and Feature Interaction Modeling-Based Eye Behavior Prediction Research. Comput Math Methods Med. 2020;2020.

10. Read GL, Innis IJ. Electroencephalography (Eeg) . Int Encycl Commun Res Methods. 2017;1–18.

11. Teplan M. Fundamentals of EEG measurement. Meas Sci Rev. 2002;2:1–11.

12. Lopez-Gordo MA, Sanchez Morillo D, Pelayo Valle F. Dry EEG electrodes. Sensors (Switzerland). 2014;14:12847–70.

13. Lance BJ, Kerick SE, Ries AJ, Oie KS, McDowell K. Brain-computer interface

technologies in the coming decades. Proc IEEE. 2012;100:1585–99.

14. Rashid M, Sulaiman N, P. P. Abdul Majeed A, Musa RM, Ahmad AF, Bari BS, et al. Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review. Front Neurorobot. 2020;14:1–35.

15. Aggarwal S, Chugh N. Signal processing techniques for motor imagery brain computer interface: A review. Array [Internet]. Elsevier Ltd; 2019;1–2:100003. Available from: https://doi.org/10.1016/j.array.2019.100003

16. Fabio B. Brain Computer Interfaces for communication and control. Front Neurosci. 1900;4:767–91.

17. Gerven M Van, Farquhar J, Schaefer R, Vlek R, Geuze J, Nijholt A, et al. The brain-computer interface cycle. J Neural Eng. 2009;6.

18. Hassanien AE, Azar AT. BCI: Current Trends and Applications [Internet]. 2015. Available from: http://link.springer.com/10.1007/978-3-319-10978-7

19. Nicolas-Alonso LF, Gomez-Gil J. Brain computer interfaces, a review. Sensors. 2012;12:1211–79.

20. Anupama HS, Cauvery NK, Lingaraju GM. Brain Computer Interface and Its Types - A Study. Int J Adv Eng Technol [Internet]. 2012;3:739–45. Available from: http://www.e-ijaet.org/media/0001/78I8-IJAET0805886-BRAIN-COMPUTER-INTERFACE.pdf

21. Forslund P. A Neural Network Based Brain – Computer Interface for Classification of Movement Related EEG. 2003;11.

22. Bilski A. A review of artificial intelligence algorithms in document classification. Int J Electron Telecommun. 2011;57:263–70.

23. Farhat NH. Photonit neural networks and learning mathines the role of electron-trapping materials. IEEE Expert Syst their Appl. 1992;7:63–72.

24. Awad M, Khanna R. Efficient learning machines: Theories, concepts, and applications for engineers and system designers. Effic Learn Mach Theor Concepts, Appl Eng Syst Des. 2015;1–248.

25. Evgeniou T, Pontil M. Support vector machines: Theory and applications. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2001;2049

LNAI:249–57.

26. Zhang Z. Introduction to machine learning: K-nearest neighbors. Ann Transl Med. 2016;4.

27. Cai Y, Ji D, Cai D. A KNN Research Paper Classification Method Based on Shared Nearest Neighbor. NTCIR-8 Work Meet. 2010;336–40.

28. Sharma H, Kumar S. A Survey on Decision Tree Algorithms of Classification in Data Mining. Int J Sci Res. 2016;5:2094–7.

29. Anwar H, Qamar U, Muzaffar Qureshi AW. Global optimization ensemble model for classification methods. Sci World J. 2014;2014.

30. Papanastasiou G, Drigas A, Skianis C, Lytras M. Brain computer interface based applications for training and rehabilitation of students with neurodevelopmental disorders. A literature review. Heliyon [Internet]. Elsevier Ltd; 2020;6:e04250. Available from: https://doi.org/10.1016/j.heliyon.2020.e04250

31. Honório BCZ, Sanchetta AC, Leite EP, Vidal AC. Independent component spectral analysis. Interpretation. 2014;2:SA21–9.

32. Tang Z, Sun S, Zhang S, Chen Y, Li C, Chen S. A brain-machine interface based on ERD/ERS for an upper-limb exoskeleton control. Sensors (Switzerland). 2016;16:1–14.

33. Coyle D, Prasad G, McGinnity TM. A time-frequency approach to feature extraction for a brain-computer interface with a comparative analysis of performance measures. EURASIP J Appl Signal Processing. 2005;2005:3141–51.

34. Huan NJ, Palaniappan R. Neural network classification of autoregressive features from electroencephalogram signals for brain-computer interface design. J Neural Eng. 2004;1:142–50.

35. Aydemir O, Kayikcioglu T. Wavelet transform based classification of invasive brain computer interface data. Radioengineering. 2011;20:31–8.

36. Lei X, Yang P, Xu P, Liu TJ, Yao DZ. Common Spatial Pattern Ensemble Classifier and Its Application in Brain-Computer Interface. J Electron Sci Technol China [Internet]. 2009;7:17–21. Available from: papers3://publication/uuid/313E8F8B-6AC9-4DAF-A03A-8B6B18966379

37. Pfurtscheller G, Neuper C, Flotzinger D, Pregenzer M. EEG-based discrimination between

imagination of right and left hand movement. Electroencephalogr Clin Neurophysiol. 1997;103:642–51.

38. Güçlü U, Güçlütürk Y, Loo CK. Evaluation of fractal dimension estimation methods for feature extraction in motor imagery based brain computer interface. Procedia Comput Sci. 2011;3:589–94.

39. Čukić M, Pokrajac D, Stokić M, Simić S, Radivojević V, Ljubisavljević M. EEG machine learning with Higuchi's fractal dimension and sample entropy as features for successful detection of depression. arXiv. 2018;1–34.

40. Osalusi B, Abraham A, Aborisade D. EEG Classification in Brain Computer Interface (BCI): A Pragmatic Appraisal. 2018;8:1–11.

41. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. Current Perspective. 2000;

42. Neuper C, Wörtz M, Pfurtscheller G. Chapter 14 ERD/ERS patterns reflecting sensorimotor activation and deactivation. Prog Brain Res. 2006;159:211–22.

43. Smits FM, Porcaro C, Cottone C, Cancelli A, Rossini PM, Tecchio F. Electroencephalographic fractal dimension in healthy ageing and Alzheimer's disease. PLoS One. 2016;11.

44. Lun X, Yu Z, Chen T, Wang F, Hou Y. A Simplified CNN Classification Method for MI-EEG via the Electrode Pairs Signals. Front Hum Neurosci. 2020;14:1–14.

45. Roots K, Muhammad Y, Muhammad N. Fusion convolutional neural network for cross-subject eeg motor imagery classification. Computers. 2020;9:1–9.

46. Dose H, Møller JS, Iversen HK, Puthusserypady S. An end-to-end deep learning approach to MI-EEG signal classification for BCIs. Expert Syst Appl [Internet]. Elsevier Ltd; 2018;114:532–42. Available from: https://doi.org/10.1016/j.eswa.2018.08.031

47. Karácsony T, Hansen JP, Iversen HK, Puthusserypady S. Brain computer interface for neuro-rehabilitation with deep learning classification and virtual reality feedback. ACM Int Conf Proceeding Ser. 2019;

48. H. M, Samaha A, AlKamha K. Automated Classification of L/R Hand Movement EEG Signals using Advanced Feature Extraction and Machine Learning. Int J Adv Comput Sci Appl.

2013;4:207–12.

49. Sleight J, Pillai P, Mohan S. Classification of executed and imagined motor movement EEG signals. 2009;1–10. Available from: http://www-personal.umich.edu/~shiwali/images/MohanPillaiSleight.pdf

50. Singaravelan S, Arun R, Arun Shunmugam D, Ruba Soundar K, Mayakrishnan R, Murugan D. Analysis of Classification Algorithms on Different Datasets. Rev Innov Compet. 2018;4:41–54.