Pontifícia Universidade Católica
do Rio de Janeiro

Alimed Celecia Ramos

Multiple Classifier System for Motor Imagery Task Classification

Dissertação de Mestrado

Dissertation presented to the Programa de Pós-graduação em Engenharia Elétrica of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Engenharia Elétrica

Advisor: Prof. Marley Vellasco

Rio de Janeiro
May 2017

Pontifícia Universidade Católica
do Rio de Janeiro

Alimed Celecia Ramos

Multiple Classifier System for Motor Imagery Task Classification

Dissertation presented to the Programa de Pós-graduação em Engenharia Elétrica of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Engenharia Elétrica. Approved by the undersigned Examination Committee

Profa. Marley Maria Bernardes Rebuzzi Vellasco
Orientadora
Departamento de Engenharia Elétrica – PUC-Rio

Profa. Karla Tereza Figueiredo Leite
UERJ

Prof. Antonio Mauricio Ferreira Leite Miranda de Sá
UFRJ

Prof. Jose Manoel de Seixas
UFRJ

Prof. Márcio da Silveira Carvalho
Vice Dean of Graduate Studies
Centro Técnico Científico – PUC-Rio

Rio de Janeiro, May 10th, 2017

Alimed Celecia Ramos

Graduated in Telecommunications and Electronics Engineering at ISPJAE (Instituto Superior Politécnico José Antonio Echeverría) in 2013. As researcher and developer in the Cuban Neuroscience Center (CNEURO) worked in the electronic design and development of a visual stimulator for visual nerve electrophysiological tests. Being part of the Laboratory of Applied Intelligence and Robotics participated in the Latin American Conference on Computational Intelligence 2016 (LA-CCI) presenting partial results of his current project, which is about Machine Learning algorithms applied to EEG processing and Brain Computer Interfaces (BCI). His main research areas are Machine Learning, Biological Signal Processing and Neurotechnology.

Bibliographic data

Celecia Ramos, Alimed

Multiple classifier system for motor imagery task classification / Alimed Celecia Ramos ; advisor: Marley Vellasco. – 2017.

177 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2017.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Sistema de múltiplos classificadores. 3. Interface cérebro computador. 4. Imaginação motora. 5. EEG. 6. Processamento de sinais. I. Vellasco, Marley. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

To my family

Acknowledgments

To my Tutor Prof. Marley Vellasco, who offered me knowledge, assistance, guidance and critical advice during all this research. I cannot describe how much I learned from her in this time as her student.

To my family, for all the support offered to me, that helped me to get so far away.

To my girlfriend, who was a constant support during my living in a foreign country, which she taught me to love.

To Prof Nival Nunes de Almeida, for his support and teachings, which showed me that teaching is also a vocation.

To my friends of the Cuban community in Rio, for all the good moments.

To the professors of the PUC Rio University that I have the pleasure to learn from them and assisted me when I needed, as Prof. Karla Figueiredo, Marco A. Meggiolaro, Marco Aurelio Cavalcanti Pacheco, Eduardo Costa, and many others.

To PUC Rio and CAPES program, for offering me the opportunity to grow as student and professional

Abstract

Celecia, Alimed; Vellasco, Marley Maria Bernardes Rebuzzi (Advisor). **Multiple Classifier System for Motor Imagery Task Classification.** Rio de Janeiro, 2017. 177p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica de Rio de Janeiro.

Brain Computer Interfaces (BCIs) are artificial systems that allow the interaction between a person and their environment using the translated brain electrical signals to control any external device. An EEG neurorehabilitation system can combine portability and affordability with good temporal resolution and no health risks to the user. This system can stimulate the brain plasticity, provided that the system offers reliability on the recognition of the motor imagery (MI) tasks performed by the user. Therefore, the aim of this work is the design of a machine learning system that, based on the EEG signal from only C3 and C4 electrodes, can classify MI tasks with high accuracy, robustness to trial and inter-subject signal variations, and reasonable processing time. The proposed machine learning system has four main stages: preprocessing, feature extraction, feature selection, and classification. The preprocessing and feature extraction are implemented by the extraction of statistical, power and phase features of the frequency sub-bands obtained by the Wavelet Packet Decomposition. The feature selection process is effectuated by a Genetic Algorithm and the classifier model is constituted by a Multiple Classifier System composed by different classifiers and combined by a Multilayer Perceptron Neural Network as meta-classifier. The system is tested on six subjects from datasets offered by the BCIs Competitions and compared with benchmark works founded in the literature, outperforming the other methods. In addition, a real BCI system for neurorehabilitation is designed and tested, producing good results as well.

Keywords

Multiple Classifier System; Brain Computer Interface; Motor Imagery; EEG; Signal Processing; Fusion Techniques; MLP; Genetic Algorithms.

Resumo

Celecia, Alimed; Vellasco, Marley Maria Bernardes Rebuzzi. **Sistema de Múltiplos Classificadores para Classificação de Tarefas de Imaginação Motora.** Rio de Janeiro, 2017. 177p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica de Rio de Janeiro.

Interfaces Cérebro Computador (BCIs) são sistemas artificiais que permitem a interação entre a pessoa e seu ambiente empregando a tradução de sinais elétricos cerebrais como controle para qualquer dispositivo externo. Um Sistema de neuroreabilitação baseado em EEG pode combinar portabilidade e baixo custo com boa resolução temporal e nenhum risco para a vida do usuário. Este sistema pode estimular a plasticidade cerebral, desde que ofereça confiabilidade no reconhecimento das tarefas de imaginação motora realizadas pelo usuário. Portanto, o objetivo deste trabalho é o projeto de um sistema de aprendizado de máquinas que, baseado no sinal de EEG de somente dois eletrodos, C3 e C4, consiga classificar tarefas de imaginação motora com alta acurácia, robustez às variações do sinal entre experimentos e entre sujeitos, e tempo de processamento razoável. O sistema de aprendizado de máquina proposto é composto de quatro etapas principais: pré-processamento, extração de atributos, seleção de atributos, e classificação. O pré-processamento e extração de atributos são implementados mediante a extração de atributos estatísticos, de potência e de fase das sub-bandas de frequência obtidas utilizando a *Wavelet Packet Decomposition*. Já a seleção de atributos é efetuada por um Algoritmo Genético e o modelo de classificação é constituído por um Sistema de Múltiplos Classificadores, composto por diferentes classificadores, e combinados por uma rede neural Multi-Layer Perceptron. O sistema foi testado em seis sujeitos de bases de dados obtidas das Competições de BCIs e comparados com trabalhos benchmark da literatura, superando os resultados dos outros métodos. Adicionalmente, um sistema real de BCI para neurorehabilitação foi projetado, desenvolvido e testado, produzindo também bons resultados.

Palavras-chave

Sistema de Múltiplos Classificadores; Interface Cérebro Computador; Imaginação Motora; EEG; Processamento de Sinais; Técnicas de Fusão; MLP; Algoritmos Genéticos

Summary

List of Figures

List of Tables

List of Acronyms

ACO: Ant Colony Optimization

ANN: Artificial Neural Network

BCI: Brain Computer Interface

BMI: Brain Machine Interface

BKS: BehaECvior Knowledge Space

CFS: Correlation based Feature Selection

CFS-BE: Correlation based Feature Selection with Backward Elimination

CFS-FS: Correlation based Feature Selection with Forward Selection

CWT: Continuous Wavelet Transform

DWT: Discrete Wavelet Transform

ECoG: Electrocorticography

EEG: Electroencephalography

ELM: Extreme Learning Machines

EOG: Electrooculography

ERD: Event-related desynchronization

ERS: Event-related synchronization

FES: Functional Electrical Stimulation

fMRI: Functional Magnetic Resonance Imaging

GA: Genetic Algorithm

k-NN: k-Nearest Neighbor

KNN-cor: k-Nearest Neighbor with correlation

KNN-cos: k-Nearest Neighbor with cosine

KNN-e: k-Nearest Neighbor with Euclidean distance

KNN-m: k-Nearest Neighbor with Mahalanobis distance

LDA: Linear Discriminant Analysis

MEG: Magnetoencephalography

MI: Motor Imagery

MLP: Multi-Layer Perceptron

mRmR: Minimum Redundancy Maximum Relevance

NH: Near Hit

NM: Near Miss

PCA: Principal Component Analysis

PLV: Phase Locking Value

PNN: Probabilistic Neural Network

PSO: Particle Swarm Optimization

RBF: Radial Basis Function Networks

RMIPE: Robust Motor Intention Prediction Ensemble

SCP: Slow Cortical Potential

SNR: Signal to Noise Ratio

SSVEP: Steady State Visually Evoked Potential

SVM: Support Vector Machine

SVM-l: Support Vector Machine with linear kernel

SVM-q: Support Vector Machine with quadratic kernel

VEP: Visual Evoked Potential

WPD: Wavelet Packet Decomposition

WT: Wavelet Transform

1. Introduction

1.1 Motivation

Ancient Egyptians were the first documented people interested in the cerebral cortex and the brain injuries [1]. Since that time, the study of the brain, its functions and diseases has been one of the focuses of philosophers, doctors, and scientist communities. The difficulty in perfectly understanding the relationship between the brain areas and the human behavior and neurological disorders led to some initiatives, such as the BRAIN initiative from the USA government [2] and the Human Brain Project from the European Union [3]. Their main objective is to offer the necessary contribution to the scientific community for the fulfillment of the brain mapping task and the understanding of the whole human brain.

Those efforts were also extended toward the neurotechnology field and one of its most promising applications: Brain Computer Interfaces (or Brain Machine Interfaces, depending on the device that will be controlled). A Brain Computer/Machine Interface (BCI/BMI) is an artificial system that translates the brain electrical signals into control commands for any external device [4]. These systems have already been applied to a broad range of areas, such as robotics [5] [6], virtual reality environments [7], rehabilitation systems [8], assistive technologies [9], neuromarketing [10] and arts [11].

BCIs are classified into invasive and non-invasive [12], depending on how the brain electrical signal is acquired. The invasive methods employ electrical signals directly from ensembles of brain cells or multiple neurons, obtained by placing one or various grids of electrodes on the brain surface. The procedure entails a brain surgery, limiting its application to animals or patients during surgery procedures due to ethical principles. Some examples are electrocorticography (ECoGs) and single neuron recordings [13]. On the other hand, the non-invasive ones dispense a surgical approach, using signals obtained from electroencephalography (EEG), functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG) or optical imaging [14]. Except for EEG, all the other techniques are limited to specific facilities or present low

temporal resolution (limiting their use for BCI systems with lower communication speed).

The most used non-invasive signal acquisition method for BCI is the EEG. The main reasons are the relative affordability and portability of the EEG devices, the good temporal resolution that the signal provides, and the low health risks of its use. This technique employs electrodes situated over different areas of the scalp to measure the activity of the electric field of the human brain and the obtained signal is characterized by low amplitude (in the μV order). In BCI applications, EEG technique allows the identification of the changes occurred in the signal in response to a determined cognitive or imagined physical task. A recognized issue is the EEG susceptibility to noise and artifacts from different sources, their low spatial resolution, and the signal temporal variability [15] [16] [17].

The control of BCI systems is executed through tasks that, depending on their paradigm, lead to different patterns in the EEG signal. Paradigms usually for this tasks are: visual evoked potentials (VEPs), slow cortical potentials (SCPs), P300 evoked potentials and sensorimotor rhythms [18]. Those sensorimotor rhythms can be modulated by motor imagery tasks. Those tasks are based on the fact that the imagination and the realization of the movement of determined part of the body present a number of common features in the EEG [19]. Therefore, for BCI applications, the subject imagines the movement of predefined body parts (commonly hands, feets, and tongue) and the system translates this intent into commands to the external agent.

Usually, due to the nature of the EEG and the motor imagery tasks, the datasets obtained are characterized by a high dimensionality, with a great number of features and a low number of instances [20]. This represents a problem for the majority of classification models used to recognize the desired action or command to be performed, which together with the great variability (within different trials of the same subject and inter-subjects) and the non-stationarity of the EEG, results in a difficult challenge for the classification of BCI systems [21] [22].

Neurorehabilitation is a new science field supported by the principles of brain plasticity and Hebbian learning [23]. Brain plasticity is the capacity of the brain and the nervous system to reorganize its structure, functions, and connections in response to determined training [24]. This training can be based on

the Hebbian learning principle, that asserts that two neurons or groups or neurons may reconnect if they are simultaneously activated, which can occur if they are connected to the same circuit of interconnected neurons [23]. Therefore, Hebbian learning defines a framework for recovery of disabled neural circuits and their function restoration.

In a BCI neurorehabilitation application, the subject activates the neural circuits correspondent to the imagined movement and, by the feedback received from the external device (typically exoskeleton or orthosis), the Hebbian learning and brain plasticity are stimulated. The application of these systems has presented promising results [24]. Consequently, a functional BCI system for neurorehabilitation can be of primary importance to assist the recovery of the more than 9 million strokes survivors yearly in the world [25].

In particular, hand impairment is a common after-stroke consequence [26]. The rehabilitation strategies for this kind of motor deficiency are performed through active, repetitive, and task-oriented hand movement [27]. In a neurorehabilitation system for hand grasping rehabilitation, repetitive exercises of hand grasping motor imagery are performed by the subject, activating a robotic device (orthosis or exoskeleton) that performs the desired task.

Motor Imagery (MI) tasks activate many brain areas, depending on the nature of the task [28], but fundamentally the areas are: the primary motor cortex, premotor and supplementary motor areas, the posterior parietal cortex, and the prefrontal areas. For the sensorimotor rhythms, the primary sensorimotor cortex is the principal source of EEG signals [29], using as main electrodes C3 and C4 of the 10-20 international system [30] [31] [32] with good classification results. One of the benefits of this reduced number of electrodes is that it can cut down the price of the acquisition device.

The nature of the rehabilitation process and its influence on the patient indicates several conditions for the machine learning system to comply. Firstly, the system must offer high classification accuracy in order to not discourage the patient during the rehabilitation exercises caused by model induced misclassifications in the task. Another important characteristic is the robustness of the model to artifacts, time and subject variations of the EEG signals. Finally, the training and tuning time of the model should be as short as possible, since it has to be tuned separately for each patient due to the characteristics of the signals.

However, despite all the efforts in the search for a model that fulfills the previous conditions, it is still impossible to conclude that such a model is superior for this signal. Moreover, usually the conclusions of the effectuated comparisons are subject to the specific features and processing algorithms.

The combination of classifiers is an approach not commonly employed for this kind of biological signal processing, despite the possibility of representing a model with improved robustness and accuracy [33]. The few works found in the literature that apply this model on motor imagery [32] [34] [35] [36] generally does not exploit the possibility of combining classifiers of different learning principles, which can boost the diversity of the model. The one that explores this approach [34] limit its analysis to only one subject, leaving aside the analysis of the response of the model to other subjects and the effect of different acquisition conditions as frequency variation and artifact presence. In addition, the fusion algorithms employed are of low complexity, which limits the response of the model in accuracy and robustness. Therefore, a multiple classifier system that offers robustness to artifacts, time and subject variations, and reasonable computing time, can be the first step towards a functional low-cost neurorehabilitation device based on a BCI system with MI classification.

1.2 Objectives

The main objective of this work is the design of a machine learning system that, based on the EEG signal from only C3 and C4 electrodes, can classify MI tasks with high accuracy, robustness to trial and inter-subject signal variations, and reasonable processing time, for application to a functional and low-cost BCI system for neurorehabilitation.

The secondary objectives are:

- To study the algorithms of each functional component in a BCI design in order to select the best alternatives to integrate a BCI system that combines high accuracy, robustness to trial and inter-subject signal variations and reasonable processing time.

- To validate the proposed model on benchmark datasets, as well as on an original dataset obtained from a low-cost commercial headset.

- To design and build an orthosis as a particular case of a neurorehabilitation device, with the purpose of integrating the proposed model to a functional BCI system.

1.3 Work description and contributions

To accomplish the proposed objective, five main stages were performed:

- Bibliographic research of BCI functional blocks;

- Study of signal processing and machine learning techniques applied to BCI;

- Development of the BCI model;

- Validation of the BCI model;

- Testing of the complete BCI system;

Firstly, a bibliographic research about the structure of a BCI model (covering every design step), its application to neurorehabilitation systems, and the neurophysiology of the brain electrical signal was conducted.

In the second phase, the most common preprocessing methods applied to EEG data for signal analysis were studied, focusing on one of the most suitable for EEG: Wavelet Packet Decomposition (WPD). This was followed by the analysis of features from different nature (e.g. statistical, energy, phase features) that can be extracted from the data in order to offer diversity on our data description. Then a survey of the Machine Learning techniques for feature selection and classification was carried out. This study included Genetic Algorithms, Correlation based Feature Selection, ReliefF, Neural Networks, Support Vector Machines, Ensemble of Classifiers, and others.

Based on the previous bibliographic study, the main characteristics and algorithms for each functional block of the BCI system were defined. With the objective of designing a model that manages to provide good data representation and generalization performance, the EEG signal was preprocessed through the WPD, a Genetic Algorithm was chosen for feature selection and the classification was performed by an Ensemble of Classifiers.

To validate the developed BCI model, benchmark datasets and a dataset obtained from a commercial EEG headset (Emotiv EPOC+) were employed as case studies. The performance analysis of the BCI system over the benchmark datasets was compared with other works in the literature.

Finally, the BCI model was tested on an implemented neurorehabilitation BCI system composed by an active hand orthosis designed for hand grasping rehabilitation.

Therefore, the main contributions of this Dissertation are:

- The definition of a BCI structure based on the following stages: preprocessing, feature extraction, feature selection, and classification.

- The comparison of state of the art feature selection methods (Correlation based Feature Selection, ReliefF, Minimum Redundancy Maximum Relevance, Consistency, C4.5 and Genetic Algorithms) employed with EEG data based on the accuracy obtained from five classifiers (Support Vector Machine, Linear Discriminant Analysis, K-Nearest Network, Radial Basis Network and Probabilistic Neural Network) on a benchmark dataset.

- The proposal of an ensemble model composed of many different classifiers and state of the art fusion techniques, resulting in a model that provides high accuracy and reasonable computation time.

- The design and construction of a functional neurorehabilitation device composed of a low-cost EEG acquisition headset, the proposed BCI system, and a hand orthosis.

1.4 Master Dissertation

The reminder of this document is organized into 5 additional chapters. Chapter 2 presents the characteristics of the brain electrical signal and mental tasks used in BCI models, giving a more detailed description of the EEG, as the most common non-invasive BCI signal. Additionally, Chapter 2 explains the different functional blocks of a BCI model, detailing the theory behind the principal techniques and focusing on the ones that the characteristic of the signal indicates as the best option for our BCI system. Finally, the chapter introduces different models of MI based BCI found in the literature.

In Chapter 3, the proposed BCI system is described. Firstly, the general structure of the proposed model (preprocessing, feature extraction, feature selection, and classification) is reported, followed by a detailed description of the algorithms employed for each component of the system.

Chapter 4 is dedicated to the two case studies, focusing on the performance of the proposed model. The first part analyzes the characteristics of the benchmark datasets. Others sections are devoted to the analysis of the system performance during variations of the algorithms employed in the model.

Chapter 5 describes firstly the dataset obtained as part of this work and has the purpose of applying the proposed model on a real application. In this chapter, the implementation of the neurorehabilitation device is addressed, firstly describing the design of the orthosis, passing to the description of the MI experiments and EEG acquisition and finally interpreting the results obtained.

Finally, in Chapter 6 the conclusions of the work and the possible lines to follow on future works are discussed.

2. EEG based Brain Computer and Machine Interfaces System

A BCI system can be divided into four main stages: preprocessing, feature extraction, feature selection, and classification, as shown in Figure 1. Each of these stages performs a critical task for the success of the system application and will be explained in details in the next sections, deepening in the theoretical basis of some of the most commonly employed methods found in the BCI and EEG signal processing literature.

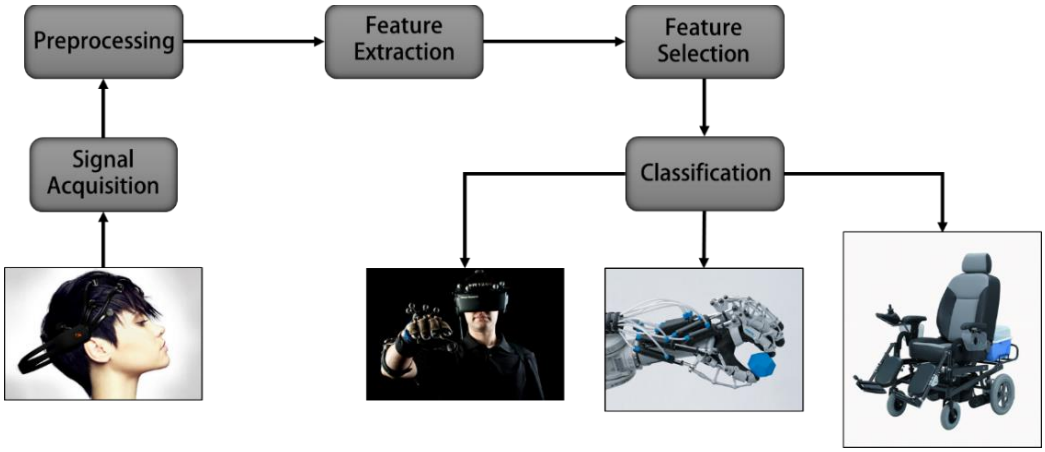

Figure 1. Block diagram of a BCI system

2.1 Brain electrical signal

The brain is composed of 10^{10} to 10^{11} neurons that connect with others by their axons and dendrites [37]. Dendrites represent the primary destination for synaptic inputs from other neurons and have the function of receiving and integrating the information [38]. The number of inputs could go from 1 to 10^5, corresponding to the complexity of the dendritic arbor. The synaptic contacts are made through the presynaptic terminal and the postsynaptic specialization. These components use the secretion of molecules from the presynaptic to the postsynaptic to communicate. The axon is the unit that conducts the signal from the synapses of the neuronal dendrites to the next site of synaptic integration [38].

This unit can be as short as hundreds of μm, or as long as one meter. The action potential is the electrical event that carries the signal over those distances. It could be explained as a self-regenerating wave of electrical activity that goes from the initiation point, on the cell body, to the end of the axon, where the synaptic contacts occur. The process of passing the information coded as action potentials to the next cell is known as synaptic transmission.

Neurons are organized in ensembles of neural circuits, which can process specific information that offers the foundation of behavior, perception, and sensation. These circuits represent the information as a pattern of action potentials, named neural code. When these neural circuits (or parts of them) are activated, local current flows are produced [39]. This constitutes the electrical activity of the brain and depending on their spatial and temporal behavior, different processes taking place in this organ are represented.

2.1.1 Signal acquisition

As stated in the introductory chapter, the signal acquisition is classified as either invasive or non-invasive, depending on the nature of the techniques employed. The most common non-invasive method for BCI applications is the EEG. This technique is of relatively low cost, when compared to other non-invasive techniques, as well as portable, which permits its daily use [18].

EEG recording employs electrodes situated over the scalp, usually according to the international standard 10-20 system [40] in order to measure the combination of brain waves determined by the neuronal activity in each electrode position. The type of electrodes is defined by their different characteristics in: disposable electrodes, reusable disc electrodes, headbands and electrode caps, saline-based electrodes, needle electrodes and dry electrodes [39]. The measure is the difference of the variation over time of the signal acquired by the active electrode and a reference electrode [41]. Figure 2 illustrates the 10-20 electrode location.

The acquired signals, obtained by the electrodes, are the passed through an amplification block. In this block, the signal is amplified from the μV values to ranges where the signal can be successfully digitalized. Those amplifiers need to provide amplification selective to the physiological signal, guaranteeing

protection from damages through voltage and current surges and reject noise and interference signals [39]. Also in this unit, analog filters are attached to reduce low frequency interferences and assure that the signal is band limited. Then the signal is digitized using Analog to Digital (A/D) Converters and transmitted to the recording device (usually a computer).

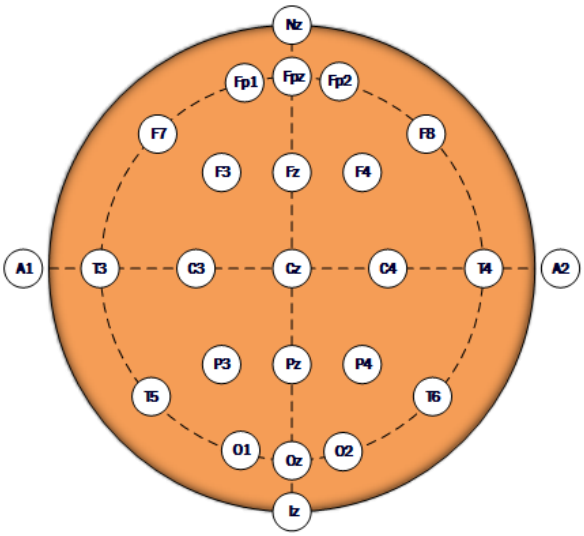

Figure 2. 10-20 Electrodes Placement System

For BCI applications, usually the signal is acquired from a multichannel electrode cap. The number of channels on these caps ranges from 16 to 256, allowing a better spatial resolution with the density increment but, on the other side, elevating the processing power required to process the signal. Figure 3 shows a multichannel cap with the acquisition system commercially available offered by the ant neuro company known as eego mylab.

Figure 3. EEG cap with recording system eego mylab

2.1.2 EEG characteristics

EEG is a complex signal due to the complexity of its source. In general, EEG is described as a stochastic process that presents great similarities with noise [42]. Its properties can be listed as:

- Noisy and pseudo-stochastic. The amplitude of 10-300 μV makes it affected by noises and artifacts. This reflects on an elevated grade of nonstationarity and randomness.

- Time-varying and nonstationary. EEG varies depending on the physiological states. These variations include sinusoidal, spikes or polyspikes, and spindles or polyspindles.

- High nonlinearity. EEG is a nonlinear process whose nonlinearity is also time, state and site dependent.

2.1.3 Mental tasks paradigms

The acquired brain electrical activity through the EEG recording can be interpreted as control commands as part of a BCI system. Such control commands are normally obtained from brain signals that, due to the source of the performed task, people can learn to modulate. The brain signals currently used are based on the following paradigms: VEPs, SCPs, P300 evoked potentials and sensorimotor rhythms [18].

VEPs are the variations in the brain activity taken place in the visual cortex when a sensory stimulation of the visual field is received [43]. In a VEPs based BCI, feature analysis allows identifying the target of the user gaze by analyzing the VEP pattern. This is possible since each target is coded as a unique stimulus sequence. The modulation of this response is not difficult, since the amplitude of the obtained pattern is larger when the stimulus is closer to the central visual field [44]. BCI systems that employ VEPs are classified, according to the stimulus sequence modulation, as: time modulated VEP BCIs, frequency modulated VEP BCIs, and pseudorandom code modulated VEP BCIs.

SCPs are low voltage shifts of the brain activity that can last from 300 ms to several seconds and can oscillate between negative and positive polarizations [45]. This activity is encountered in the EEG with frequencies below 1 Hz [46]. Negative SCPs reflect a decrease in the threshold for the excitation of underlying neuronal structures, leading to increasing neuronal activity. Positive SCPs, on the contrary, are related to a reduction of cortical excitation of the underlying neuronal structures, which indicates a decrease in the activity [47]. The voluntary production of this positive and negative shifts can be trained by employing a thought translation device [48]. This device makes use of visual and auditory marks so that the user produces negative or positive shifts relative to a baseline.

P300 evoked potentials are positive peaks in the EEG that take place when a subject detects an informative task-relevant stimulus [49]. It owes the name to the characteristic that the peak occurs about 300 ms after attending the stimuli [50]. This stimulation is defined as an oddball stimulus among repetitive stimulus and it has been prove that the less probable the stimulus, the higher the peak amplitude of the response [51]. This paradigm does not need of user training, in change, the user performance can be reduced by his familiarization with the infrequent stimulus (which produces P300 amplitude reduction). For BCI applications, visual and auditory stimuli have been applied [52].

Sensorimotor rhythms are the neurophysiological rhythmic activities recorded over the sensorimotor cortex modulated by actual movement, motor intention or motor imagery [29]. The major components bands are usually defined as the mu band (8-13 Hz) and beta band (13-30 Hz) [53]. The amplitudes of these rhythms vary when brain activity is related to any motor task and similar behavior can be obtained by motor intention or motor imagery (without any real motor output). The two amplitude modulations manifested by this phenomenon are defined as a power suppression in the low-frequency components known, as event-related desynchronization (ERD) and an amplitude enhancement known as event-related synchronization (ERS). Through motor imagery (imagination of the movement of any motor part of the body) those changes in the cerebral activity can be successfully modulated [54].

Controlling of sensorimotor rhythms has demonstrated to be not an easy task, due to troubles with the motor imagery paradigm [18]. People tend to imagine visual images of related real movements, which produce patterns that are

different from motor imagery. As a solution, training is of great significance. For this kind of training, the subject is asked to perform a determined motor imagery task, followed by the extraction of the sensorimotor rhythms and their classification by a comparison with a determined reference. Depending on the results of the comparison, a visual or auditory feedback is provided to the subject.

2.1.4 Neurophysiology of Motor Imagery tasks

The MI tasks that can modulate the synchronization or desynchronization of the sensorimotor rhythms are correlated with the neural activity and cortical mechanism of the brain. The analysis of these physiological sources began with the neuroimaging studies conducted after the introduction of the MI concept by M. Jeannerod [55], which demonstrated that MI tasks activate several cortical and subcortical regions that are overlapped with those for movement execution [28].

Further studies confirmed that those kinesthetic imageries induce a somatotopically organized activity in the Primary Motor Cortex, the Premotor and Supplementary Motor Areas, the Posterior Parietal Cortex, and the Prefrontal Cortex [54] [28]. Specifically, the Premotor and Supplementary Motor Areas were defined as the fundamental structures during planning and preparation stages of motor control, with a better predictive value for the Supplementary Motor Areas for imagery tasks. The Primary Motor Cortex also provided a significant activation, but in a lower level than the actual movement. The representation of the main activated areas of the brain for MI tasks is shown in Figure 4.

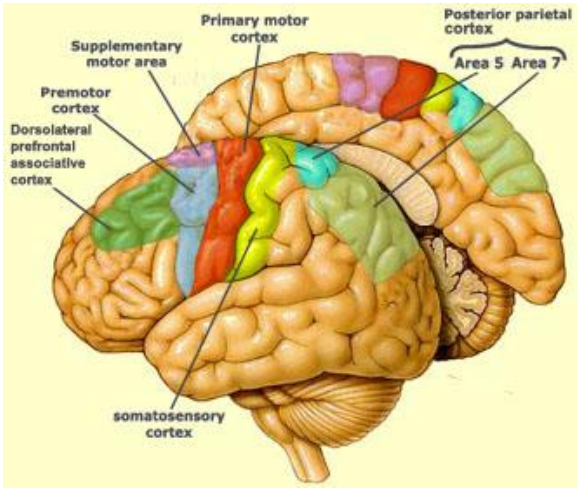

Figure 4. Main brain areas activated for MI tasks

Studies employing EEG confirmed that the most relevant changes of the signals during MI tasks are identified over the Primary Sensorimotor Cortex [29] [54]. This is the reason for the usual selection of the electrodes C3 and C4 (located over the sensorimotor cortex) for the acquisition of the EEG signals for this tasks [56], even when in some cases the peak of the activity is not situated in the exact position of those electrodes, but in their surrounding [54].

2.2 Preprocessing

The preprocessing stage has, as main purposes, the removal of artifacts and other undesired signals present in the EEG, as well as to improve the Signal to Noise Ratio (SNR), which results in a global optimization of the differentiation of the signal classes [4]. In addition, the signal is prepared for the analysis and feature extraction process, either in the time domain or by its transformation to the frequency domain. There is a large number of methods that are used to complete these tasks, the most commons are: frequency filtering (FIR filters) [57]; referencing (common reference, average reference and current source density) [58], which is the technique to select the reference brain voltage for the comparison with the measurement of the electrodes; spatial filters (common average referencing, Surface Laplacian, Principal Component Analysis, Independent Component Analysis, Common Spatial Patterns, and others) [57] [58] [59]; and time / frequency analysis (Autoregressive Modeling, Fast Fourier Transform and Wavelet Transform) [18] [57] [59].

In the next sub-sections, it will be explained firstly the artifacts that can be present in the EEG measurement and it will be mentioned techniques that can eliminate or reduce their effect. This will be followed by a detailed description of the time / frequency analysis used for the BCI model proposed in this work, the WPD as extension of the Wavelet Transform (WT) technique.

2.2.1 Artifacts Elimination

Artifacts are undesirable signals that produce alterations on the magnitude of the measurement, affecting the overall signal of interest. Those artifacts can be

avoided by preventing their inclusion during the signal recording, or by removing them after the signal is obtained [16]. These spurious components generally cause a decrement of the BCI system performance, affecting the signal class recognition.

Artifacts can be classified according to their origin as physiological or technical artifacts [18]. The main sources of physiological artifacts are eye movement and blinks (ocular artifacts), the muscular artifacts (produced by swallowing, talking, walking or other muscular contractions), and cardiac activity. Ocular artifacts (EOG) are fundamentally received by the frontal electrodes, but can influence other ones [16]. These alterations have higher amplitude when produced by blinking, and lower magnitude for eye movement (related to higher frequency interference). For the muscular artifacts, the degree and type of muscle contraction determine the shape and intensity of the interference [16]. These artifacts generate perturbations on all classic frequency bands of the EEG and present less repetition than other artifacts, making it harder to detect and correct. Heart activity is the electrical activity of the heart and introduces a low amplitude rhythmic component to the EEG. Lastly, technical artifacts are related to power-line noises and changes in the electrode impedances.

To eliminate those artifacts, classical filtering techniques (low pass, band pass, and high pass) or shielding (for technical artifacts) are used. Filtering is only helpful when the signal and the undesired components do not share the same frequencies. For these cases, a number of diverse techniques have been developed, such as adaptive filtering, Wiener filtering, Bayes filtering, regression, EOG correction, blind source separation, wavelet transform method, empirical mode decomposition and non-linear mode decomposition [16].

2.2.2 Wavelet Transform

Wavelet Transform is a method commonly employed for biological signal processing due to its ability to represent the signal information with an excellent time-frequency resolution [60]. The randomness, time varying, non-stationarity and transient components of the EEG signal indicate the WT as a better method than other traditional techniques, such as Fourier Transform or Autoregressive Model [61] [62] [63]. Wavelet Transforms allows the signal to be simultaneously

represented in time and frequency domains. Additionally, it separates the frequency bands that carry the information without losing temporal resolution.

The WT is a technique that transforms a time signal using fixed building blocks named mother wavelet. The procedure produces scaled and shifted versions of this waveform of limited duration and zero average value over the temporal domain [64]. WT can be categorized into continuous wavelet transform (CWT) and discrete wavelet transform (DWT). The CWT applies the continuous variation of the scaling and translation factors, resulting in high processing effort at the calculation of wavelet coefficients for each possible factor. This fact makes the DWT more popular for BCI applications.

DWT provides a non-redundant, highly efficient wavelet representation employing a simple recursive filtering scheme. Therefore, the signal can be reconstructed by an inverse filtering operation without losing any information. For a signal x(t), the wavelet decomposition can be represented as:

$$x(t) = \sum_{k=-\infty}^{+\infty} C_{N,k}\,\theta(2^{-N}t - k) + \sum_{j=1}^{N}\sum_{k=-\infty}^{+\infty} d_{j,k}\,2^{\frac{-j}{2}}\varphi\left(2^{-j}t - k\right) \quad (1)$$

Where $C_{N,k}$ are the representation of the approximation coefficients at level N and $d_{j,k}$ are the representation of the detail coefficients at level j, with j between 1 and N. $\varphi(t)$ is the wavelet function and $\theta(t)$ is a scaling function [65].

These functions are related to high and low pass filters, giving a multiresolution approach to the decomposition of the signal into frequency sub-bands with the mentioned successive filtering steps, losing half of the samples in each step by the Nyquist theorem. The resultant implementation follows an octave-band tree structure, sequentially separating it in approximated version (low-frequency part) and residual details (high-frequency part) and iterating the process at each step only on the low-pass branch of the tree [66].

2.2.2.1 Wavelet Packet Decomposition

WPD is an extension of the DWT that maintains the same filtering scheme but allows the details functions to be further split into subsequent sub-bands [60].

The result is a complete Wavelet Packet Tree (as can be observed in Figure 5), producing, for n levels of decomposition, 2^n different sets of coefficients, as opposed to the n+1 sets from the DWT. However, the down sampling operation guarantees the same overall number of coefficients. As result, this method offers the possibility of adaptation to the frequency components of the signal. Wavelet analysis can employ a great number of families of mother wavelets, the most used ones being the Daubechies, Symlets, Morlet, Haar and Coiflets. In the Figure 5, S represents the signal, with A and D being the Approximation and Detail coefficients respectively.

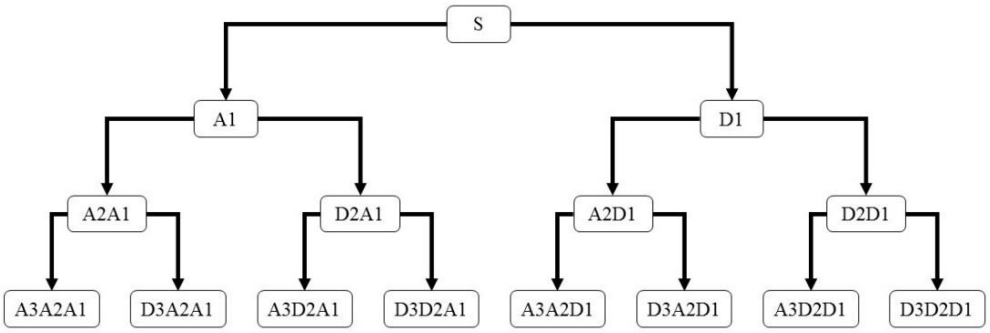

Figure 5. Third Level Wavelet Packet Decomposition

2.3 Feature Extraction

Feature Extraction is the process of converting the data into useful attributes or features that offer a good representation of the data. It can be defined as the key step in the data analysis process, greatly determining the success of the machine learning system [67]. The principal objectives are to reduce the dimensionality and to increase the discrimination capacity of the dataset. Those features are extracted based on different algorithms and principles, ranging from statistical measures and energy distributions to spatial parameters and fractal coefficients. As the WPD represent a suitable preprocessing technique for EEG, the focus of the next sections is on the typical features extracted from this transformation.

2.3.1 Statistical and Morphological Features

Those types of features are statistical or parametrical measures of the values of a series that characterize its behavior. The following statistical features can be cited: average, maximum, minimum, standard deviation, median, variance, kurtosis, skewness and entropy of the values for each frequency sub-band [62] [68] [69] [70] [71]. In the group of the morphological parameters, the most common are: area, signal slope, range, peak to peak time window, the number of zero crossings, slope sign alterations, integrated EEG, root mean square and the simple square integral of the values of each frequency sub-band [70] [71] [72]. Finally, another commonly used attribute is the ratio between magnitudes of different sub-bands as the average of the absolute values [68] [69].

2.3.2 Energy Features

Time and frequency representations of the energy distribution or power spectrum of the wavelet coefficients in each frequency sub-band have proved their relevance in the description of the behavior of the decomposed EEG signal, as well as in detecting the similarity between segments of the signal. Some of the application examples that can be found in the literature are [62] [68] [69] [73] [74].

2.3.3 Phase Features

To evaluate the synchronization of two signals, two main approaches have demonstrated to be relevant for EEG analysis: coherence and phase locking value (PLV) [75] [76] [77]. The coherence is derived from the cross-spectrum of two time series signals, which can be calculated using the Fourier Transform. The PLV determines the synchrony between the phase of two signals and can be obtained employing the Hilbert transform, which allows the computation of the instantaneous phase of the signals [78]. These features do not consider the amplitude of the signals, therefore, constitutes a different approach in the signal description.

2.4 Feature Selection

Feature Selection is one of the most relevant processes in a Machine Learning System [79]. It guarantees the reduction of the number of features, removing irrelevant, redundant and noisy ones, and improves the results for the specific application (it augments the velocity of the algorithm, enhances the learning performance and increases the model interpretability) [80]. This technique can be defined as a process that chooses the best subset of features according to a certain criterion.

As mentioned in the Introduction, EEG data are correlated, noisy, and their quality is affected by subject's degree of concentration during the recordings. In addition, the dimensionality of the data after the feature extraction is usually high (several features extracted from a number of channels) and the quantity of samples is small (due to the nature of the acquisition trials), which is known as the "curse of dimensionality" [81]. This two characteristics of the EEG data represent a challenge for any machine learning system, and the application of some feature selection technique is usually mandatory [82].

The feature selection algorithms can be grouped into three main categories: filter, wrapper and embedded models [80]. The filter approach (Figure 6) evaluates the individual features or subsets of features based on predefined parameters, independently of the learning algorithm. In the case of the feature subset evaluation, filters use a search algorithm to generate the subsets for their posterior evaluation. The wrapper approach (Figure 7), on the other hand, requires a learning algorithm, using its accuracy to evaluate the subsets of features selected by a search algorithm. Figures 6 and 7 show the block diagrams of these models. The last category is the embedded approach, whose algorithms incorporate the feature selection as part of the machine learning training process, evaluating its utility for optimizing the objective function of the learning algorithm. Examples of this last category are the C4.5 [83] and the sparse logistic regression [84]. For the filter approach, the Correlation based Feature Selection (CFS) [85], ReliefF [86], Minimum Redundancy Maximum Relevance (mRmR) [87], Consistency [88], and Koller's [89] can be cited. Some of the algorithms that can be included as wrapper approach are Genetic Algorithm (GA) [90], Particle Swarm

Optimization (PSO) [91], and Ant Colony Optimization (ACO) [92], among others.

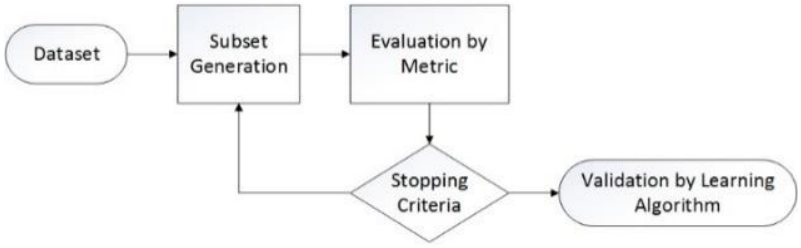

Figure 6. Filter approach for subset of features

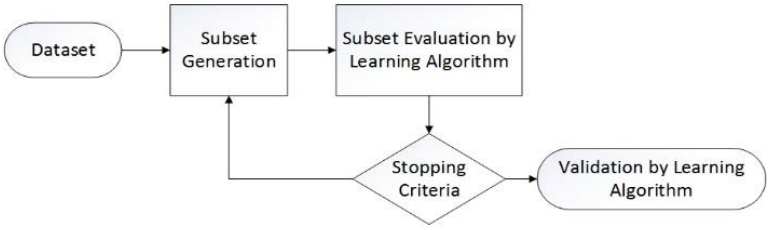

Figure 7. Wrapper approach

Concerning BCI applications, there is little information in the literature of which is the most adequate method. In [82], the author empirically compares five feature selection methods (Correlation based Feature Selection, Information Gain, ReliefF, Consistency and 1RR) on benchmark BCI data (MI tasks dataset) with 60 channels and a sampling rate of 250 Hz. As a result, Consistency was the method that selected the smallest subset of features, however, Correlation based Feature Selection and Information Gain were defined as the best methods employing a pair-wise accuracy comparison (wins-draws-losses ranking). In [81], a general survey of feature selection methods and machine learning algorithms applied to BCI data is presented, without comparing their performance. In [93] two feature selection models are compared on a MI task dataset: Principal Component Analysis (PCA) and a wrapper approach with forward selection, concluding that the wrapper approach outperforms the PCA considering the accuracy of the model. Therefore, there is no consensus yet on the best feature selection approach for motor imagery tasks.

The next sub-sections describe four state-of-the-art algorithms as examples of filter methods (CFS, ReliefF, mRmR and Consistency), one for embedded

methods (C4.5), and the application of biologically inspired optimization methods on a wrapper approach.

2.4.1 Correlation based Feature Selection

CFS [85] evaluates subsets of features based on the hypotheses that a good subset is the one that contains features highly correlated with the output classes and not correlated between them. The validation criterion is the utility of the features for the prediction of the class and their correlation with the others features. The merit of a subset is defined as:

$$Merit_S = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \quad (2)$$

Where $Merit_S$ is the heuristic merit of a subset S that contains k features, $\overline{r_{cf}}$ is the mean of the correlation class-feature and $\overline{r_{ff}}$ is the mean of the correlation inter-feature. The numerator could be interpreted as how predictive is the feature subset and the denominator as how much redundancy exist in the features that compose the subset. Therefore, the method identifies irrelevant and redundant features. This algorithm depends on a search algorithm that generates the subset of features, as Best First Search [94], Forward Selection [95], Backward Elimination [96], among others.

2.4.2 ReliefF

Relief [97] is an algorithm inspired by instance-based learning with the objective of distinguishing the statistically relevant features through the validation of how good the values of a feature could differentiate between instances that are close to one another. This method selects an instance X of the dataset and finds, using the Euclidean distance, the closest instances belonging to each class: Near Hit (NH) to the same class and Near Miss (NM) to the other. Then it compares the value of each feature in the instance X with the correspondent feature in NH and NM, updating the relevance weight for each feature. This process is repeated for

every instance of a randomly selected group. The features chosen for the final subset are the ones with a relevance superior to a predefined threshold.

ReliefF [86] is an improved version of Relief to allow handling multiclass problems and incomplete data. In this case, the algorithm searches more than an NM and NH using k-Nearest Neighbor (k-NN), extending the search of these values for multiple classes.

2.4.3 Minimum Redundancy Maximum Relevance

The minimum redundancy - maximum relevance algorithm [87] is based on a similar heuristic to the CFS algorithm. In this case, the mutual information is the measure employed to validate the relevance of the features, generating a ranking where the features are ordered by its mutual information with the class and with the other features. The most relevant feature is the one which has the maximum mutual information with the class and minimum mutual information with the others features. This is accomplished by the maximization of the following expression:

$$\max(V_I/W_I) = \max\left(\frac{\frac{1}{n_f}\sum I(c,f)}{\frac{1}{{n_f}^2}\sum I(f_1,f_2)}\right) \quad (3)$$

Where V_I and W_I represent maximum relevance and the minimum redundancy conditions respectively, the n_f is the number of features in the set, $I(c,f)$ is the mutual information between a class and a feature and $I(f_1,f_2)$ is the mutual information between two features.

After the ranking procedure, the heuristic followed in [87] suggest the formation of subsets conformed by different numbers of features ordered by the ranking. The learning algorithm will then validate these groups, giving the number of features in the ranking order that obtains the best classification accuracy.

2.4.4 Consistency

Consistency [88] is a metric for evaluating feature subsets based on the inconsistency rate. The principles of the inconsistency rate are: two instances are inconsistent if their only difference is their class; for all equal instances the inconsistency count is the number of instances minus the biggest number of instances belonging to one class, and the inconsistency rate is the sum of all the inconsistency counts divided by the total number of instances.

The algorithm randomly creates a feature subset S. If the number of features (C) is less than the number of the best subset, then the data with the attributes conforming the subset S is evaluated by the inconsistency criteria. If its inconsistency rate is less than that of the best subset, then S becomes the best subset. The algorithm stops when the inconsistency rate of the best subset is higher than a determined threshold. Therefore, the algorithm selects the subset whose inconsistency rate is not higher than the threshold and has the lower number of features.

2.4.5 C4.5

C4.5 [83] is an algorithm to construct decision trees based on the divide and conquer strategy. The development of the tree has two steps: construction and pruning of the tree. The construction begins with the root, using the feature that best differentiates the instances in their classes. Then, based on the number of values that the feature takes, the tree is divided into that number of branches. Based on the information gain ratio of each feature, the algorithm continues the process selecting features for the nodes. The pruning phase helps to avoid overfitting. In this phase, the node is eliminated only if the accuracy of the pruned tree is not worse than the one of the unpruned tree.

This decision tree can be seen as a ranking of the features based on their information gain ratio. This metric is defined as the division of the information gain by the intrinsic information of each feature. In feature selection, the features that compose the final tree are the subset that will be validated by the learning algorithm.

2.4.6 Biologically inspired optimization algorithms

The biologically inspired optimization algorithms are complex methods for problem-solving based on nature and biological activities. Their heuristic basically performs the iterative improvement of a population of possible solutions for a fitness function.

For feature selection, these algorithms are applied on a wrapper approach with the fitness function defined as the accuracy of the learning model. The best solution determines the subset of features that offer a higher accuracy. Some commonly applied methods on this type of applications are: GA, PSO, ACO, among others.

2.5 Classification Models

Classification is the last step of the BCI model. In this stage, the dataset composed of the selected features by trials (instances of the dataset) is used to train and test the classification model. The most commonly employed classification models in BCI environments are discussed in the following sections.

2.5.1 Support Vector Machine

A Support Vector Machine (SVM) [98] is a classifier that employs a separating hyperplane that maximizes the margins between the data classes mapped into a space of higher dimension by the "kernel trick". This maximization leads to an increase of the generalization capacity of the model, and the use of different kernels helps to create nonlinear decision boundaries. In addition, the SVM is known by its immunity to overfitting and the "curse of dimensionality".

2.5.2 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) [99] is based on the assumption that each probability density functions of the class can be addressed as a normal density and have the same covariance. The objective of LDA is to determine a

hyperplane that separates the data into their classes. This hyperplane is obtained by the search of the projection that maximizes the distance between the mean vector of each class and minimize the interclass variance. A new data is classified determining the highest probability density function.

2.5.3 K-Nearest Neighbor

k-NN [100] is one of the simplest classification methods. This technique uses the k closest samples to the new data instance to decide, by a majority principle, its class. This method can produce nonlinear decision boundaries with a high value of k and enough training samples. The k nearest neighbors are determined employing a distance metric as Euclidean, or correlation coefficients as Pearson.

2.5.4 Artificial Neural Networks

Artificial Neural Networks (ANNs) [101] are learning models inspired by the brain neurons and properties like their parallelism and redundancy. The ANNs are composed of simple processing units called neurons and their directed weighted connections. Those elements can be distributed in different topologies, as feedforward networks, recurrent networks, and completely linked networks [102]. The networks for a supervised learning, in dependence of the defined training algorithm, present the input patterns to the net and compare the response to a target output, and the difference determines how the weights will be updated. This is an iterative process that ends when the response of the networks converges to the desired output. In general, the ANNs commonly used for BCI models are the Multi-Layer Perceptrons (MLPs) [103], Radial Basis Function Networks (RBFs) [104], Probabilistic Neural Networks (PNNs) [105], Convolutional Neural Networks [106], among others.

2.6 Ensemble of Classifiers

An ensemble of classifiers can be defined as a group of classifiers that combine their individual outputs in some way that derive a consensus opinion in order to classify new examples [107]. This style of cooperation between classifiers has roots on the principle that different classifier designs offer complementary information about the patterns to classify, which can represent an improvement of the performance of the BCI model [108].

The design of a successful ensemble of classifiers depends in great measure on two parameters: the accuracy of the components of the ensemble and their diversity [109]. A classifier can be labeled as accurate if its accuracy value is above the random guessing for new examples. The diversity is assured if the errors of the members of the ensemble are uncorrelated.

In general, building an ensemble of classifiers presents two main tasks: the creation of the individual classifiers composing the ensemble and the combination of their outputs. Those tasks can be addressed using multiple approaches [110]. The creation of the individuals classifiers present several aspects of interest. First is the definition of characteristics as the selection of the base classifier or to use different ones, the number of them, and the type of training (at the same time or incrementally). In addition, it is important for the training stage the processing of the training dataset, which can be manipulated either in the instances presentation and the feature space for its presentation to each classifier. After the creation of the classifiers, it is necessary to decide the mode of combination of their outputs. The different combinations that can be obtained from those parameters led to the creation of successful algorithms such as Random Forest [111], Bagging [112], Adaboost [113], Random Subspace Ensembles [114], Error Correcting Output Codes [107], and others. In the Figure 8 can be seen a general model of an ensemble of classifiers, where C_1 to C_n denotes the classifiers members of the model.

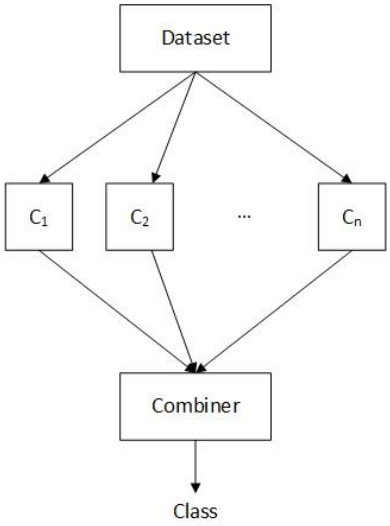

Figure 8. General ensemble model

Based on three aspects of the ensemble modelling, three main categories can be defined [110]. The first aspect relates to the type of combination, being divided into classifier fusion and classifier selection. In classifier fusion, each component of the ensemble is trained on the whole feature space. On the other hand, in classifier selection, each component is trained in parts of the feature space, with the objective of being an expert in the patterns of that simplified task. Therefore, for the fusion approach, a combination of the classifiers output is required, such as average, product or majority voting. On the contrary, for the selection approach a classifier is selected for the labelling of one input sample.

Another aspect involves the decision whether the ensemble fusion method is trainable or non-trainable, depending on the necessity of additional training for the ensemble fusion after training each classifier. Usually, non-trainable methods employ statistics as mean, median, maximum, minimum or product as the fusion operator. In the trainable case, the algorithms apply combination rules as weighted average or stacked generalization, which establish their parameters through an additional learning process.

The last category separates ensembles into static or dynamic, taking into consideration how the instances (input patterns) interact with the conformation of the ensemble [34]. The static ensembles are the ones that employ the complete set of classifiers for all instances. A dynamic ensemble, on the other hand, dynamically selects subsets of classifiers for each instance.

2.6.1 Fusion Methods

In a scheme where the classifiers cooperate to decide the final label, the fusion method that unifies their decisions has a relevant role. The type of fusion method is directly related to the classifier's output. There are four main categories of outputs [115]: class labels, ranked class labels, numerical support for the classes, and the one known as oracle.

The class label offers a label without information about the certainty of the decision. The ranked class label is defined as a subset of class labels ordered by their probability [116]. In the numerical support category, the classifiers offer a vector containing output values that represent the support given to the hypothesis that the sample belongs to each class. The oracle defines the systems that employ the knowledge that a classifier is wrong or not, disregarding the class label that has been assigned [117].

Combination methods for the class label outputs possess an intuitive nature or are based on a probability framework, such as the Majority Vote [118], Weighted Majority Vote [119], Naïve-Bayes Combiner [110] and the multinomial methods or Behavior Knowledge Space (BKS) [120]. The continuous valued outputs employ mathematical operators and more complex trainable models, as average, maximum, minimum, median, product, fuzzy integrals [121] [122], or the use of a classifier as a combiner (meta-classifier) [123].

2.7 BCI/BMI Applications

Since the early years of BCI research, the focus of the applications was in medical assistance. However, in the last two decades the number of research groups in this subject grew exponentially (from eight groups to more than one hundred) [18] [124]. This growth, combined with a number of technological advances, allowed the expansion of BCI applications, reaching even the consumer market [29] [125] [126].

The medical assistive applications can be divided into two general groups, depending on their final purpose: the ones dedicated to communication and control and the ones dedicated to motor substitution or recovery. BCI offers the possibility to severely disabled individuals to communicate with other people and

their environment. One of the most successful mechanisms for this purpose is the BCI driven spelling devices [127] [128] [129] [130]. Their operation is based on a virtual keyboard shown on a screen that allows the subject, through the control signal, to select a letter and compose words. In this type of devices, the control signals are diverse, including SCPs, P300, detection of eye blinks, and MI combined with cognitive tasks. In this category, web browsers control is also included [131] [132], that employs a heuristic over SCPs or P300 to allow navigate the Internet. In addition, there are environmental control applications that facilitate the control of domestic devices as lights, TV, lights, room temperature, and others, offering independence to the patient. Those user designed interfaces were presented in [133], and the different control modalities were matched with the loss of motor abilities that each user presented.

The other general application is for the substitution or recovery of motor functions. Motor functions, as grasping, can be restored in some grade by the application of functional electrical stimulation (FES) [134], also termed as neuroprostheses. BCI can be employed as control signal of a FES, as demonstrated in [135] using MI tasks (foot movements). These initial applications for the recovery of motor functions have been expanded to elbow and arm, adding other robotic devices as hand orthosis and even virtual reality for after stroke rehabilitation [136] [137] [138] [139]. Another paradigm used for restoration of grasp movement are the Steady State Visually Evoked Potentials (SSVEP) [140]. Another area where BCI represent motor substitution is in assistive mobility. This can be accomplished by BCI-controlled wheelchairs [141] or mentally driving robotic devices as telepresence mobile robots or quadcopters [142] [143]. These applications rely on the asynchronous identification of MI combined with cognitive tasks, or also the P300 recognition.

Outside the medical purposes, BCI also presents relevant applications in the entertainment field, which opens the door to the BCI utilization for non-disabled people. The first of them is in the gaming area, where the number of BCI games [144] [145] [146] has grown exponentially. The paradigms included goes from MI and SSVEP to P300. In addition, BCIs can indicate the emotional status, boredom, and frustration of the user, which can be used in the game design, helping to maintain the user's interest and affect him positively. Virtual Reality constitutes another application that is receiving attention due to its promising quality as a

feedback provider to BCI systems. MI, SSVEP, and P300 were used in systems that explore virtual spaces and interacts with the virtual environment [147] [148]. Finally, the employment of BCI for photo or music browsing, music composition, and painting [149] [150] [151] can also be mentioned.

An important result of the advances in BCI research and its relevance in the entertainment field lead to the development of low-cost commercial BCI systems [152] [153]. Headsets developed by companies as Emotiv [154], Neurosky [155], Neuroelectronics [156], Cognionics [157], OpenBCI [158] and Macrotellect [159] offer to the general public the possibility of training their neural brain waves with different acquisition quality, electrodes quantity, and final applications. Those applications cover from PC games to relaxation or concentration focus, allowing their use for research purposes.

Other fields that employ the neurofeedback offered by the BCIs are cognitive performance [160], speech skills, affection [161], biometrics [162] [163], detection of covert behavior [164], drowsiness detection [165], defense applications [166], and treatment of mental disorders (epilepsy, attention deficit, schizophrenia, depression, and others) [167] [168] [169] [170] [139]. Neuromarketing is a new field that applies neuroscience methods to marketing research [171] [172]. The applications go from tests of product concepts to advertising campaigns, making use of brain signals to learn the real preferences of the users. This opportunity is already offered by some companies as Neurofocus [173], Neuro-Insight [174] and Forebrain [175].

3. A Robust Motor Intention Prediction Ensemble Model

3.1 General Model

As defined in Chapter 2, the main functional blocks of a BCI system are: preprocessing, feature extraction, feature selection, and classification. For each block, there is a great number of techniques and algorithms that can be applied, which makes the selection of the appropriate method for each stage a challenging task. This selection has to consider the fundamental characteristics of the desired system, such as, in this Dissertation, robustness to trial and inter-subject signal variations, high accuracy, and the lowest processing time possible. Those conditions usually lead to a subset of possible models for each stage, which are then assessed and the one that provides the best performance in the desired characteristics is selected. The following sections describe the models selected (and the reasons for their selection) from the vast BCI literature for each functional block of the proposed Robust Motor Intention Prediction Ensemble Model (RMIPE) illustrated in Figure 1 located in Chapter 2.

3.2 Preprocessing and Feature Extraction

The EEG is characterized by its randomness, abrupt variations in time, non-stationarity and the presence of transient components. Those characteristics theoretically signalize the WT as more appropriate to represent the signal than other traditional techniques such as Fourier Transform or Autoregressive Model (each one applied to EEG with good results) [61] [62] [63]. Fourier Transform, besides not being very suitable for non-stationary signals [62] and transient signals [61], represents the signal in short time duration blocks, which can limit the spectral resolution of the representation [61]. In addition, Autoregressive Model is incapable of correctly capturing the transient components of the signal [63].

Of the three techniques that apply the WT (CWT, DWT, and WPD – see Section 2.2.2), WPD was the one selected for preprocessing the signal. This

selection is based on the high computational cost of the CWT (its functional principle produces an infinite number of coefficients, generating redundancy [18], [176], [177]), and on the capacity of the WPD to decompose both details and approximation coefficients depending on the decomposition level (the DWT only decomposes the approximation coefficients, which are the low-frequency components). As a result, the signal can be represented in frequency bands of equal width, which is very helpful for the EEG representation.

There are two main attributes for the successful application of the WPD: the mother wavelet and the number of levels of the decomposition. The number of levels defines the amount of sequential filtering steps suffered by the signal, allowing the delimitation of the frequency bands width.

The mother wavelet chosen was the Daubechies of order 4 (db4), which is usually employed on EEG signals [62] [178]. Moreover, in [179], after a comparison of seven different mother wavelets, Daubechies was defined as the most suitable wavelet family for extracting more discriminative features from the imaginary EEG signals.

The definition of the number of levels depends on the major frequency bands components for the control task of the EEG system (motor imagery in this case). As explained in Chapter 2, the major components bands of the ERD and ERS are the mu band (8-13 Hz) and beta band (13-30 Hz). Although these frequencies present the majority of the signal information for motor imagery, the best discriminative components from the frequency bands vary significantly due to the non-stationary nature of the EEG [180]. Some works in the literature indicate better results using broader frequency bands, including delta (from 0.5 to 4Hz), theta (4-7Hz), and gamma (30-40Hz) [181] [182]. Additionally, good results have been reported with only the delta band contribution [183] [184] [185] [186]. Therefore, the application of a broader frequency scope can contribute to a better response of the BCI system to the variations of the EEG signal. For a sampling frequency of 128Hz, a decomposition in four level was implemented. This procedure firstly divides the sampling frequency in half, starting the bandwidth of the analysis with 64Hz. Then, for each subsequent level, divides the correspondent frequency ranges in half, ending in the fourth level with sub-bands of 4Hz from 0 to 64Hz. For the proposed model the following eight sub-bands

were selected: 0-4Hz, 4-8Hz, 8-12Hz, 12-16Hz, 16-20Hz, 20-24Hz, 24-28Hz and 28-32Hz.

Usually, after applying the WPD, the wavelet coefficients are used as features to represent the signal to be analyzed. However, for this BCI system, the signals of each frequency band were reconstructed and then used on a feature extraction process. This approach is based on previous unpublished works of the author that indicated a better performance with the reconstructed signals than the wavelet coefficients. An example of the obtained EEG signal waveforms in the time domain, for each frequency band, is presented in Figure 9. This sample was obtained from a trial of the subject S4 from the dataset IIIb of the BCI competition III.

3.2.1 Feature extraction for the RMIPE model

To improve the discrimination capacity of the BCI/BMI system, some features are extracted from the reconstructed signals. These features combine statistical, power and phase properties of the signals and have already been employed for EEG processing. There is no evidence in the literature of using the subset of features proposed as a group for this application before. The complete feature vector is composed of the parameters described in Table 1, where N is the number of samples, S_i is the signal in the time domain of sub-band i, μ is the mean and C_i are the wavelet coefficients. $\Delta\theta$ represents the difference between the phase information of the signal of the channels C_3 and C_4. The instantaneous phase of the signals was obtained using the Hilbert transform, computing the phase angle of the decomposed values.

The feature proposed for this work is based on other works in the literature that used the ratio of other metrics with good results [68] [187]. The calculation of those parameters for each frequency sub-band produces a final feature vector composed of 134 features:

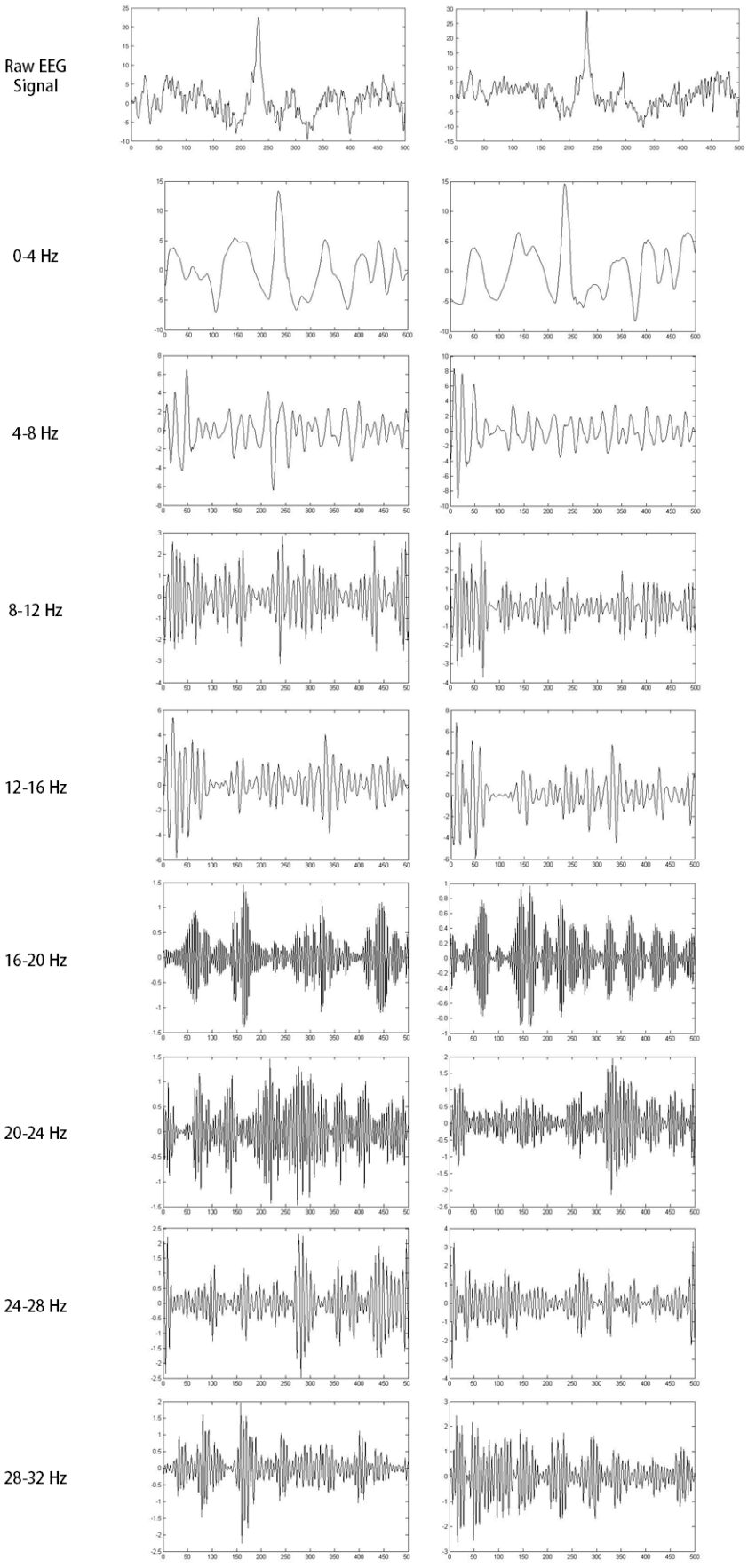

Figure 9. Example of the reconstructed signals of electrodes C3 (left) and C4 (right) for each frequency sub-band

- 7 features for 8 frequency sub-bands for 2 channels (112 values)
- 1 feature that calculates the difference between the phase of the two channels for 8 frequency bands (8 values)
- 1 feature that calculates the ratio between adjacent sub-bands for 2 channels (14 values)

Table 1. Features for the RMIPE model

Name	Formula	Reference
Mean of the absolute values of each sub-band	$F_1 = \frac{1}{N}\sum_{n=1}^{N}\lvert S_i(n)\rvert$	[187]
Average amplitude change of each sub-band	$F_2 = \frac{1}{N}\sum_{n=1}^{N}\lvert S_i(n+1) - S_i(n)\rvert$	[70]
Standard deviation of each sub-band	$F_3 = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}\lvert S_i(n) - \mu\rvert^2}$	[69]
Variance of each sub-band	$F_4 = \frac{1}{N-1}\sum_{n=1}^{N}\lvert S_i(n) - \mu\rvert^2$	[188]
Energy of each sub-band	$F_5 = \frac{1}{N}\sum_{n=1}^{N}(S_i(n))^2$	[62]
Entropy of the coefficients of each sub-band	$F_6 = -\sum_{i} {C_i}^2 \log({C_i}^2)$	[69]
Phase Locking Values of each sub-band	$F_7 = \frac{1}{N}\sum_{n=1}^{N}\exp(j\Delta\theta(n))$	[189]
Root mean square of each sub-band	$F_8 = \sqrt{\frac{1}{N}\sum_{n=1}^{N} S_i(n)^2}$	[70]
Ratio of the energy of adjacent sub-bands	$F_9 = \frac{F_5(S_{i+1})}{F_5(S_i)}$	Proposed for this work

3.3 Feature Selection

A great number of feature selection techniques has been applied on EEG data, including filter, wrapper and embedded approaches [81] [82] [190] [191] [192] [193] [194]. However, as discussed in Chapter 2, there is little information in the literature of which is the most adequate method or at least a comparison of their performance with different classifiers.

Therefore, this work selected a set of six feature selection methods, separately used in previous EEG data [20] [82] [190] [195] [196] [197], in order to evaluate and compare their performance [194]. Those methods are: four filter approaches (CFS, ReliefF, mRmR and Consistency), one embedded approach (C4.5), and one wrapper approach (GA). The accuracy evaluation was accomplished over a subject of a benchmark dataset (subject S4 of the dataset IIIb -MI of the BCI Competition III [198]). Every trial consists of 7 seconds of EEG recording of the electrodes C3 and C4 of left or right hand imagery movement, sampled with 125Hz and filtered with a Notch Filter between 0.5 and 30 Hz. The number of instances of the balanced dataset is 540.

Matlab 2013a and Weka 3.7.13 software tools have been used to develop the algorithms (ReliefF, mRmR and GA in Matlab and CFS, Consistency and C4.5 in Weka). The CFS and Consistency methods used Best First with forward selection and backward elimination as the search algorithms. For applying the ReliefF algorithm, the final heuristic of the method was altered. As stated in Section 2.4.2, the result of the ReliefF algorithm is a rank of all the features, and the feature selection is done defining a threshold that signalizes the lower value of relevance for the features of the final subset. This process is challenging, given that the search for the appropriate threshold may be difficult. The selection heuristic proposed in [87] for the mRmR algorithm presents a different approach. Based on the ranking procedure, subsets are generated composed of different features. For example, for a three features group, the first three features are selected, for a five features group, the first five features, and so on. Those predefined groups are then tested with the objective of selecting the subset with the best accuracy. This generates better results that the threshold limitation, being a more detailed analysis, the reason why it is the heuristic applied on the ReliefF approach proposed for this work. Additionally, the number of nearest neighbors

for the search of NHs and NMs has also been varied. The GA was configured with tournament selection, scattered crossover and Gaussian mutation; the population size and the number of generations were also varied in the experiments.

To confirm the benefit of each feature selection algorithm for the MI data, all of them have been compared by the performance obtained with five different classifiers commonly used with EEG data: PNN [199], SVM [200], RBF [104], LDA [201] and k-NN [202]. These algorithms were selected due to their low computational cost and good performance, mixing linear and non-linear approaches. All the classifiers were implemented in Matlab 2013a. The next table shows the parameters for the configuration of each classifier.

Table 2. Configuration of the parameters for each classifier

Classifier	Parameters	Range of the Values	Steps
PNN	spread of the activation function	0.1-100	0.01
SVM	kernel	linear, quadratic, rbf	
	solution method	quadratic programming, least square, and sequential minimal optimization	
	soft margin (C)	0.1-50	0.1
	sigma (for the rbf kernel)	0.1-50	0.1
RBF	spread of the activation function	0.1-50	0.1
LDA	discriminant function	linear, mahalanobis distance	
k-NN	distance metric	Euclidean distance, cosine, correlation	
	number of neighbors	0-200	1

To evaluate and compare the classification algorithms, the following performance measure has been applied:

$$P = 100 * \frac{C_c}{C_t} \tag{4}$$

Where P is the performance given in percentage, C_c is the number of correct classifications and C_t the total number of patterns. For the classification process, the dataset was divided into a training set (70% of the dataset) and a test set (30%)

and normalized between 1 and 0. The best results obtained are seen on the following tables. Table 3 presents the performance in terms of classification accuracy and Table 4 describes the number of features selected by each algorithm (from a total number of 134 original features). The acronyms CFS-FS and CFS-BE correspond to CFS with Best First with Forward Selection and Backward Elimination respectively. Also, the Con-FS and Con BE are the acronyms for Consistency with Best First with Forward Selection and Backward Elimination.

Table 3. Performance (%) of every method for every classifier

	PNN	SVM-l	SVM-q	SVM-rbf	RBF	LDA-l	LDA-m	kNN-cor	kNN-cos	kNN-eucl
CFS-FS	77.36	79.87	78.62	**81.76**	77.99	76.10	78.62	**81.76**	81.13	79.87
CFS-BE	78.62	**88.68**	77.36	82.39	84.91	77.36	50.31	83.01	80.50	81.13
ReliefF	81.13	**88.05**	84.28	86.16	87.42	86.16	83.65	86.16	84.91	84.28
Con-FS	75.47	78.62	**81.76**	81.13	76.10	76.73	72.33	80.50	79.87	78.61
Con-BE	79.25	81.13	**83.02**	**83.02**	79.25	80.50	70.44	80.50	81.76	81.76
mRmR	82.39	**88.05**	84.91	86.16	81.13	84.91	77.99	86.16	84.28	83.65
C4.5	81.13	86.79	85.54	85.54	80.50	**87.42**	77.36	85.55	84.91	84.28
GA	88.05	91.82	91.20	90.57	89.31	**93.71**	**93.71**	93.08	91.20	92.45

Table 4. Number of features selected by every method for every classifier

	PNN	SVM-l	SVM-q	SVM-rbf	RBF	LDA-l	LDA-m	kNN-cor	kNN-cos	kNN-eucl
CFS-FS	11	11	11	11	11	11	11	11	11	11
CFS-BE	99	99	99	99	99	99	99	99	99	99
ReliefF	46	98	41	47	97	71	24	46	34	40
Con-FS	18	18	18	18	18	18	18	18	18	18
Con-BE	17	17	17	17	17	17	17	17	17	17
mRmR	20	98	40	66	66	66	44	96	90	28
C4.5	40	42	40	40	42	40	9	36	41	42
GA	61	66	55	49	80	60	55	62	65	59

The results indicate that all feature selection algorithms were able to reduce the number of features with a relatively good performance, being Consistency and CFS the ones with smaller subsets (not including CFS with backpropagation). However, due to their severe reduction in the number of features for the subsets, the resultant classification performance of these methods is poorer in comparison with the ones obtained by other algorithms. As it can be seen from Tables 3 and 4, good accuracy is obtained with a number of features between 40 and 60, which is less than half of the original number of extracted features.

The highest accuracy results were produced by the wrapper approach - GA, boosting the performance of every classifier with smaller subsets of features than the best results of CFS, Relief, and mRmR. Due to the nature of the wrapper approach, the results for GA are a logic outcome. As explained in Chapter 2, the wrappers optimize the performance of the classifier, while the filters optimize their ranking metrics, without considering the classifier performance. For MI data, if the objective is to obtain the best performance of the system, then GA guarantees a successful system, being remarkable that even with a classifier that is not suitable for this kind of data (in this case LDA with mahalanobis distance) the GA method obtained the best result.

Due to the results provided by this performance analysis, the feature selection based on GA has been selected for the RMIPE model.

3.4 Classification

A great variety of classifiers has been applied to EEG data, including Naïve Bayes [203], k-NNs [202], SVMs [200], MLPs [103], RBFs [104], Extreme Learning Machines (ELM) [77] and Deep Neural Networks [204]. However, despite all the efforts in search of adequate methods and classifiers, it is still impossible to conclude that one algorithm is superior for this type of biological data, limiting the conclusions to only one dataset or another. Moreover, for a real BCI system, there is not a consensus about an algorithm that combines high performance, low processing time and robustness to the signals variations introduced in previous sections with superiority over the others.

It is well known that the cooperation of classifiers from different designs can offer additional information about the data that has to be classified, leading to a better performance of the classification model and generating a better global model. As stated in [34], [104], [205], [206] and [207], fusion of classifiers has promising results applied to BCI data. In addition, ensembles can represent a good solution to the “curse of dimensionality” and be robust to the variance of EEG data both in the time domain and inter-subject (dataset), and even robust to the presence of artifacts. Those characteristics make them really interesting for their application in a BCI system.

In Chapter 2, the main parameters that provide a good ensemble design were presented, as defined in [109]: the accuracy of the individual components of the ensemble and their diversity. In this BCI system design, the diversity of the ensemble is assured by the different learning principles of its components. More specifically, the proposed ensemble is composed of different models of classifiers, which can improve the chance that the errors of the members are uncorrelated [123]. The accuracy is also provided, since the classifiers that form part of the ensemble have already been used with EEG data, with positive results. The selected classifiers are the following: PNN, SVM with linear kernel, SVM with quadratic kernel, LDA linear, RBF, k-NN with the Euclidean distance, k-NN with the Mahalanobis distance, k-NN with correlation, k-NN with cosine (even defined as an inconvenient metric because it does not satisfy the triangle accuracy, presented good accuracies in previous works), and MLP. The resultant ensemble can be seen in Figure 10. By assuring accuracy and diversity conditions, the proposed model should outperform the performance of each individual. Each classifier was tuned separately and independently through an exhaustive search, in order to both obtain the best classification performance and minimize the training time. The proposed ensemble model is a static one. Static ensembles present lower computational cost than dynamic ones, which is an important characteristic for BCI applications.

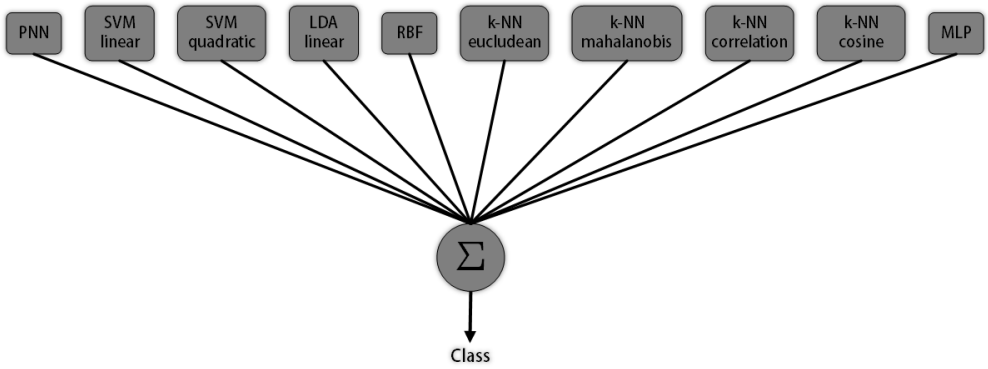

Figure 10. Ensemble composition

The last important aspect of the ensemble modeling is the definition of the fusion method. In this work, trainable and non-trainable approaches were assessed. The feature selection method is applied individually to each classifier,

with the objective of obtaining the representation of the signal that produces the best performance. So, each classifier is trained over the whole feature space.

Four methods already applied in conjunction with BCI or EEG data were selected: one majority voting approach, four weighted majority voting approaches, Naïve Bayes combination, and an MLP as a meta-classifier [207] [208] [209] [210]. In the majority voting approach, the final class is decided simply by the number of votes received (each classifier gives only one vote). In case of a tie, the winning class is chosen randomly.

The weighted majority voting approaches compute a weighted average based on the relevance of each classifier, as shown in the next expression:

$$C_f = \max_{j=1,\dots,c} \sum_{i=1}^{L} w_i D_{i,j} \tag{5}$$

Where C_f is the final class, c is the number of classes, L is the number of classifiers that voted for the class j, w_i is the weight of the classifier i, and $D_{i,j}$ is the classifier i that voted for the class j.

As mentioned, four different weighted majority methods have been implemented:

1) The weights are defined as the classifier's performance;
2) The weights are also defined as the classifier's performance but a binary GA is integrated in order to select the classifiers that will take part in the final decision;
3) The weights are defined by a GA with real representation;
4) In this case, two GAs are applied: one that specifies the classifier's relevance (weights) and another that selects the classifiers that will be considered in the fusion process. Three modes are implemented: weights between 0 and 1; weights between 0 and 1 but summing 1; and weights between -1 and 1.

The implementation of the Naïve Bayes combiner method on a dataset Z of cardinality N is based on the formula:

$$C_f = \max_{j=1,\ldots,c} \left[\frac{1}{N_j^{L-1}} \prod_{i=1}^{L} cm^i(j, s_i) \right] \quad (15)$$

Where C_f is the final class, c is the number of classes, L is the number of classifiers that compose the ensemble, N_j is the number of elements of the dataset from class j, cm^i is the confusion matrix of the classifier i, and its entry (j,s_i) is the number of elements of the dataset belonging to class j and were classified by the classifier i as class s. This is, the support of the fusion method for a class A given the proposed classes that the individual classifiers predicted for an instance, depends on the multiplication of the number of instances that each classifier classified as the proposed class being of the class A, divided by the number of instances belonging to the class A. This calculation is performed for each class of the problem, and the maximum of the support values is the final class given to the instance.

The meta-classifier approach is also known as stacking approach. In the proposed model, an MLP is trained using, as input vector, the outputs of each ensemble member, and the desired class as its output, for the complete dataset. The structure of the stacking model for N classifiers is illustrated in Figure 11.

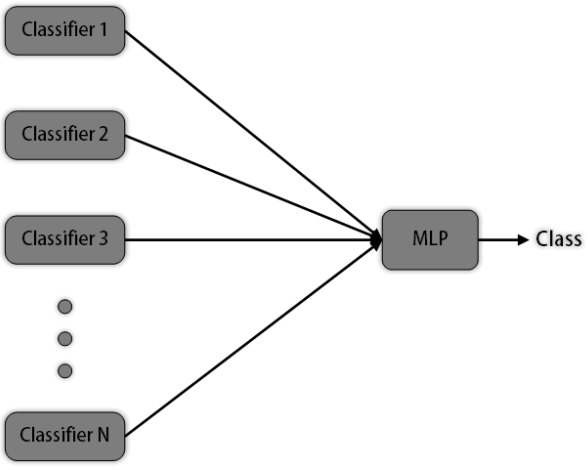

Figure 11. Model of a meta-classifier for fusion of Multiple Classifiers

Based on all the chosen methods, for each component of the proposed system, the final model of RMIPE can be seen in Figure 12.

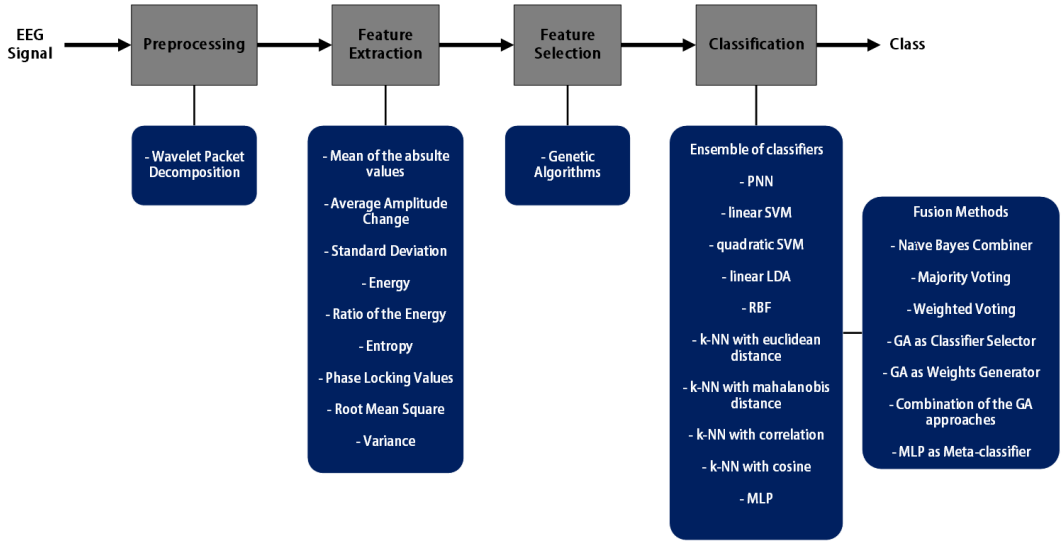

Figure 12. RMIPE Model

4. Case Study I: Classification Performance

This first case study has as main objective the validation of the proposed model. The RMIPE is applied to different databases, freely available online, to assess the best fusion method for the ensemble of classifiers. The best fusion method is then compared to the results of other state-of-the-art methods. Finally, the influence of some of the functional blocks and their algorithms is evaluated.

4.1 Databases

The selected databases for this case study were all chosen from challenges of BCI competitions. Specifically, the databases are: dataset III from BCI competition 2003 (BCI competition II) [211], datasets IIIa [212] and IIIb [198] from the BCI competition III and dataset IIa from the BCI competition IV [213].

The dataset III from BCI competition 2003 was provided by the Department of Medical Informatics, Institute for Biomedical Engineering, Graz University of Technology. The subject was a 25 years old woman, and the recordings were performed during a feedback session to control a bar by the imagery movement of the left or right hand (two classes). The dataset is composed of 280 trials of 9 seconds length (balanced between the two classes: left and right hand imagery movement). The experiment was designed with 2 seconds of relaxation, followed by an acoustic stimulus indicating the beginning of the trial and a cross displayed on the screen during one second. Then, an arrow in the screen indicating left or right signalizes which hand is necessary to move. A scheme of the acquisition experiment can be seen in Figure 13. The EEG signals over three bipolar channels over the C3, C4 and Cz positions were recorded using a G.tec amplifier and Ag/AgCl electrodes. The signals were sampled with 128Hz and filtered between 0.5 and 30Hz.

The dataset IIIa from the BCI competition III was a contribution of the Laboratory of Brain-Computer Interfaces of the Graz University of Technology. The dataset presented 360 trials of a multiclass motor imagery experiment (left hand, right hand, foot, and tongue), being balanced among the four classes. In the

recording procedure, a Neuroscan amplifier of 64 channels was employed. The sampling frequency of the signal was of 250Hz and it was used a filter between 1 and 50Hz with a Notch filter. Each trial for each subject (three in total) started with an acoustic stimulus after two seconds of relaxation followed by a displayed fixation cross. After one second the cue appeared in the form of an arrow pointing left, right, up or down; and correspondently the subject performed the motor imagery task until the beginning of the seventh second.

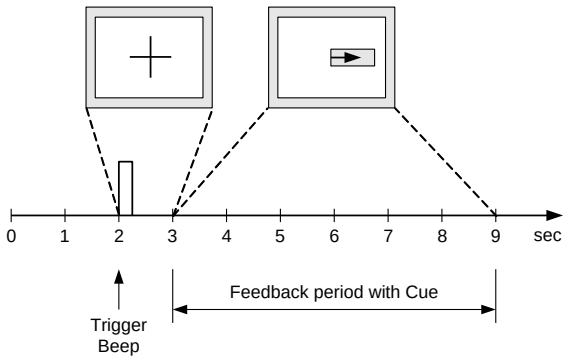

Figure 13. Acquisition experiment for the trials of the dataset III from BCI competition 2003

The database proposed as problem IIIb for the third BCI competition was composed by subjects S4, X11, and O3. The subjects S4 and X11 carried out 1080 trials of a basket paradigm feedback and O3 performed 640 of a virtual reality feedback for a two class motor imagery task (right and left hand), resulting in a balanced dataset between the two classes. In both cases, the approach is similar to the previous ones and their time description can be seen in Figure 14 and Figure 15. The recordings were acquired using the C3 and C4 channels with a bipolar amplifier from G.tec and the EEG signal was sampled with 125Hz and filtered between 0.5 and 30Hz with the Notch filter on. The paradigm for the feedback of the experiment of Figure 15 is described on [214]. In it, the green ball is displayed in a middle of two baskets (one red and one green) after 3 seconds. After one more second, the ball start to fall and the user have to maintain it in the desired region (right or left basket) through the imagination of left or right hand movement.

The dataset IIa from the BCI competition IV is composed of 9 subjects doing 4 motor imagery tasks (left hand, right hand, feet, and tongue). The total

number of trials is 576, balanced among the four classes. The paradigm is the same of the dataset IIIa from the BCI competition III, but with a different time scheme, as is shown in Figure 16. The recording process used 20 Ag/AgCl electrodes (channels), the frequency sampling of the signal was 250Hz and a bandpass filter, between 0.5 and 100Hz, was applied. An additional 50Hz notch filter was enabled to suppress line noise. Intentionally, the data were left with EOG artifacts, and three EOG channels were provided for the application of artifact removal algorithms.

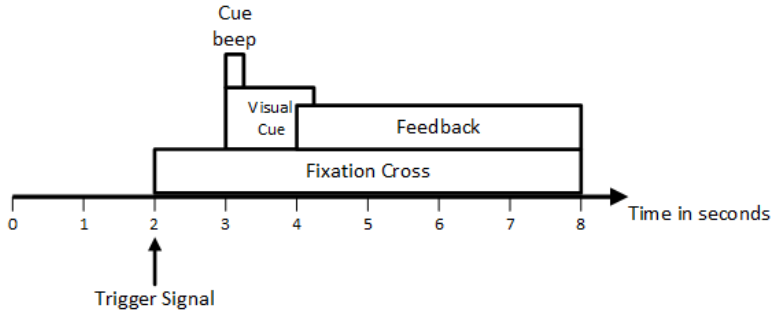

Figure 14. Acquisition experiment for subject O3 from the dataset IIIb of the BCI competition III

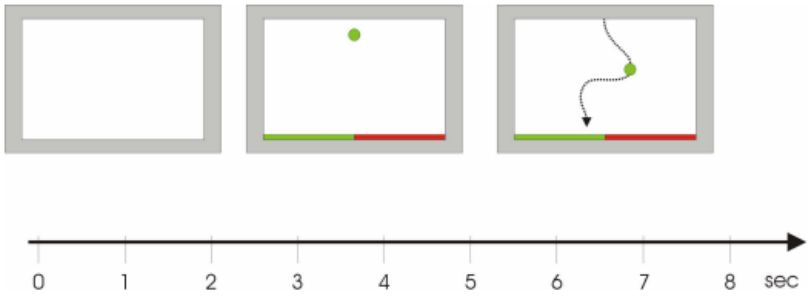

Figure 15. Acquisition experiment for subjects S4 and X11 from the dataset IIIb of the BCI competition III

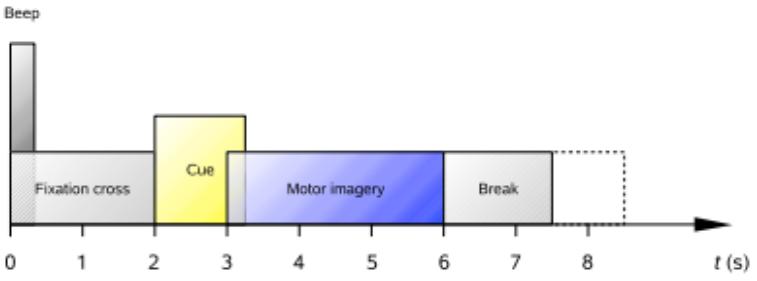

Figure 16. Time scheme for the dataset IIa from the BCI competition IV

4.1.1 Benchmark

In order to validate the results of the proposed model on the datasets described in the previous section, a comparison with other BCI/BMI models found in the literature was performed. The selected publications included results obtained using the subject of the dataset III from BCI competition 2003 and the three subjects from the dataset IIIb of the third BCI competition. The comparison also includes the results of the teams involved in the problem of the third BCI competition. In the tests performed in this work, the data is maintained at the original rates of each competition, with the objective of realizing the comparison on the same terms described in each work. The number of trials for training and testing is shown in the next table.

Table 5. Training and testing trials for the subjects of the datasets used for the Benchmark comparison

	S4	X11	O3	A2
Number of trials for training	540	540	160	140
Number of trials for testing	540	540	160	140

Twenty models were selected, among those that have also used these data, to realize a comparison with the obtained results. These models are described in Table 6:

Table 6. Benchmark models

Name	Preprocessing and Features	Classifier model
Lemm, Schäfer and Curio [30]	Complex Morlet Wavelets	Probabilistic classification method based on the Gaussian distributions and the Bayes error of misclassifications
Burmeister, Reichl, and Mikut [215]	alpha and beta frequency bands selected using the multivariate analysis of variance (MANOVA)	SVM
Pei and Bin [216]	FFT with Hanning window	Fisher Discriminant Analysis
Parini [217]	PSD and AR parameters	LDA
Coyle, Prasad, and McGinnity [218]	Prediction Neural Networks and Short Time Fourier transform (STFT)	LDA

Yao, Yin, and Liao [30]	Scales of Morlet-Wavelet and wavelet-based bandpower	SVM
Tavakolian and Rezaei [30]	AAR parameters	Bayesian Network Classifier (BNC)
Lotte [219]	Bandpower	Fuzzy Inference System (FIS), MLP, SVM and a Linear Classifier
Brodu, Lotte and Lécuyer [31]	Multifractal cumulants, Predictive complexity, bandpower	LDA
Wu and Ge [220]	Common Spatial Pattern	Biomimetic Pattern Recognition (BPR), SVM and LDA
Ahangi [34]	Statistical information of the coefficients of a DWT	AdaBoost, Bagging, Behavioral Knowledge Space, Decision Template, Majority Voting and Weighted Majority Voting, among others
Omar, Wassim and Mohammed [221]	Welch PSD	LDA and Quadratic Discriminant Analysis (QDA)
Djemili, Bouroba and Korb [222]	Welch PSD, standard PSD and logarithm of bandpowers	SVM and LDA
Sampanna and Mitaim [35]	Coefficients of a DWT combined with AR Model coefficients	Ensemble of SVMs
Chen, Fang and Zheng [223]	Phase space features	LDA
Bashashati [32]	Bandpower calculated by a filtering approach and the bandpower calculated by the Morlet wavelet approach	Boosting and Random Forest, among others
Lin, Guo and Huang [224]	AR coefficients	SVMC (controlled by the soft margin), SVM controlled by the number of support vectors, SVM with polynomial kernel function and SVM with radial basis function
Tan, Sa and Yu [225]	Bandpower	ELM, LDA, and SVM
Bashashati, Ward and Bashashati [36]	Same as [32]	Ensemble composed by Linear Regression classifiers with Majority Voting, Averaging, Stacking and Maximum
Tan [226]	Average and logarithm of the bandpower	ELM probabilistic model, ELM, SVM and LDA

4.2 Performance Analysis without Feature Selection Algorithm

The first performance analysis is about the individual results of each classifier that composes the Ensemble, without the application of the feature selection algorithm. This approach represents the baseline performance, allowing the evaluation of the impact of feature selection methods and the Ensemble model on the performance of the BCI system, when they are applied. As part of the preprocessing stage, for each trial of the EEG signals from channels C3 and C4, the time frame of the idle state was eliminated, conforming a block of the time of real motor imagery tasks. In addition, a process for removing noisy or erroneous trials was included. After this process, the number of trials of subjects S4, X11, O3 was reduced to 531, 538, 159 for training and 533, 540, 157 for testing, respectively, of the previously mentioned two classes of the datasets (left and right hand imagery movement). The parameters configuration for each classifier can be seen in Table 7. An exhaustive search was performed using all possible values in the range, for all parameters, to obtain a good relation between performance and processing time, concluding in the described ranges. All algorithms have been implemented using the Matlab software.

It is worth to mention that due to the random initialization of the weights of the MLP, three neural networks are created. The best performance on the validation set is the one that is selected as the best classifier. The results obtained for the test sets are shown in the next table. Appendix 1 presents the correspondent Confusion Matrix for each classifier (class 0 represents the left hand and class 1 represent the right hand).

The first important aspect that must be highlighted is that the MLP outperformed in a large margin all the other classifiers for each subject. Moreover, the mean of the results for each classifier shows that MLP is superior in a 9.03 percentage points to the next best classifier, which is the linear SVM. The two SVM approaches presented relatively good classification performances. In general, other classifiers performance are over 70.00%, except the k-NN with Mahalanobis distance that performed very poorly with a mean of 60.78%.

Table 7. Configuration Parameters of each classifier

Classifier	Parameter	Value		
		Min	Max	Steps
PNN	Spread	0.1	1.2	0.1
SVM linear kernel (SVM-l)	Soft margin	0.1	1	0.1
	Solution method	Least Square		
SVM quadratic kernel (SVM-q)	Soft margin	0.1		
	Solution method	Quadratic Programming		
LDA (LDA-l)	No parameter			
RBF	Spread	1.6	1.8	0.04
		2.1	2.3	0.04
KNN euclidean (KNN-e)	Number of neighbors	20	50	1
KNN mahalanobis (KNN-m)		27	31	1
KNN cosine (KNN-cos)		95	105	1
KNN correlation (KNN-cor)		80	100	1
MLP	Number of neurons in the hidden layer	8	10	1
	Training function	trainlm		

Table 8. Results for individual classifiers without feature selection

	S4	X11	O3	A2	Mean
PNN	69.42±0.00	68.52±0.00	66.88±0.00	80.71±0.00	71.38±6.31
SVM-l	77.49±0.00	73.70±0.00	75.80±0.00	79.29±0.00	76.57±2.39
SVM-q	77.11±0.00	73.15±0.00	72.61±0.00	77.86±0.00	75.18±2.69
LDA-l	74.11±0.00	71.11±0.00	67.52±0.00	65.00±0.00	69.44±4.00
RBF	71.48±0.00	68.33±0.00	72.61±0.00	72.86±0.00	71.32±2.08
KNN-e	70.92±0.00	70.00±0.00	64.97±0.00	81.43±0.00	71.83±6.91
KNN-m	68.11±0.00	64.26±0.00	58.60±0.00	52.14±0.00	60.78±6.96
KNN-cos	70.92±0.00	67.59±0.00	58.60±0.00	80.00±0.00	69.28±8.84
KNN-cor	72.05±0.00	68.70±0.00	60.51±0.00	79.29±0.00	70.14±7.79
MLP	**85.55±0.55**	**78.70±1.62**	**85.99±1.88**	**92.14±1.48**	**85.60±5.49**

The observation of the Confusion Matrices for each classifier (see Appendix 1) reveals that, for the subject X11, there is a tendency to misclassify the left hand, obtaining a balanced response only for the cases of PNN, k-NN with cosine and with Mahalanobis distance. On the contrary, the subject O3 presented difficulty to correctly predict the right hand class, with eight of the classifiers

showing this limitation. Subjects S4 and A2 did not present signs of a certain tendency. Particularly analyzing the classifiers, RBF presented a higher rate of misclassifications for the left hand and the k-NN with Mahalanobis distance for the right one. The two SVM approaches were the ones that demonstrated a tendency to obtain balanced results.

4.3 Performance Analysis with Feature Selection and Single Classifiers

The addition of the feature selection stage, based on a GA (see Section 3.3), to eliminate irrelevant features provides the results illustrated in Table 9. The number of generations and population size of the GA were tuned in order to reduce the training processing time without causing a significant negative effect on the performances. Appendix 2 provides a compilation of the obtained Confusion Matrices from the classification processes.

Table 9. Performance of the individual classifiers with GA as feature selection algorithm

	S4	X11	O3	A2	Mean
PNN	70.54±2.27	70.56±2.88	68.15±1.91	80.00±2.58	72.31±5.25
SVM-l	78.42±2.07	74.26±2.51	73.89±2.87	81.43±2.51	77.00±3.60
SVM-q	79.55±1.60	72.59±2.53	75.16±2.57	80.71±2.85	77.00±3.79
LDA-l	75.24±1.60	71.48±2.78	69.43±3.84	66.43±2.14	70.65±3.70
RBF	72.05±2.32	69.26±2.15	73.25±2.24	77.86±2.89	73.11±3.58
KNN-e	72.42±1.89	70.37±2.43	66.88±1.94	82.14±2.18	72.95±6.54
KNN-m	73.17±2.18	67.04±2.04	57.96±2.57	72.86±2.51	67.76±7.11
KNN-cos	73.92±2.45	67.22±2.22	66.24±2.24	80.71±1.89	72.02±6.72
KNN-cor	73.92±2.63	69.07±2.13	70.06±3.21	80.00±2.14	73.26±4.96
MLP	**86.30±1.04**	**81.85±1.11**	**90.45±1.27**	**94.29±1.09**	**88.22±5.36**

The results presented in Table 9 confirm that the feature selection stage improves the individual performances. In general, all classifiers had their performance enhanced by at least 1 percentage point. The most benefited classifier is the k-NN with Mahalanobis distance, increasing its poor average performance by almost 7 percentage points. Other classifiers that received a boost in their accuracies were the k-NN with correlation, k-NN with cosine, and MLP,

with a 3.1288, 2.7472, and 2.625 percentage points of increment in the mean performance, respectively.

A comparison of the obtained results for the SVM with linear and quadratic kernel, with and without feature selection, illustrates that the improvements were not as significant as the values for other classifiers. This is an evidence of the robustness of the SVMs to the "curse of dimensionality" problem.

The analysis of the Confusion Matrices from Appendix 2 shows that, in general, subjects X11 and O3 tended to misclassify the right hand class, with 6 classifiers with this limitation. On the contrary, S4 presented difficulties with the classification of the left hand class, also 6 with classifiers. Subject A2 did not reveal any tendency. Focusing on the classifiers, more balanced results can be observed for the MLP, SVM with linear kernel and k-NN with correlation, while RBF was the only classifier with a tendency for misclassifications, in this case for the right hand.

4.3.1 Features Analysis

An interesting issue is to examine which features are the most relevant to the good discrimination by the classifiers of the MI tasks performed by the subject. In order to implement this analysis, the most selected features (more than three times for five runs of the GA), for each classifier, are presented for each subject in the correspondent tables of Appendix 3.

The first case to analyze is the one for the features that can be defined as more relevant (selected by seven classifiers or more). Table 10 shows the number of classifiers that select the most relevant features for at least one subject and present a total of selections bigger than 20 (except the energy of C3 for sub-band 24-28 Hz that is relevant for the analysis and obtained a total of 19).

Table 10. Features defined as more relevant for at least one subject

	S4	X11	O3	A2	Total
Avg abs C3 sb0-4	10	3	6	5	**24**
Avg abs C3 sb8-12	8	7	4	1	**20**
Avg abs C3 sb12-16	4	4	7	5	**20**
Avg abs C3 sb16-20	2	6	4	8	**20**

Avg abs C3 sb28-32	8	3	4	5	**20**
Avg abs C4 sb12-16	5	4	7	5	**21**
Avg abs C4 sb28-32	2	7	7	4	**20**
Avg ampl C3 sb0-4	9	6	2	3	**20**
Avg ampl C3 sb12-16	5	6	6	7	**24**
Avg ampl C3 sb28-32	6	5	7	6	**24**
Avg ampl C4 sb0-4	7	5	4	5	**21**
Avg ampl C4 sb12-16	6	5	9	5	**25**
Avg ampl C4 sb16-20	6	7	4	8	**25**
Avg ampl C4 sb28-32	4	6	7	6	**23**
Std dev C3 sb8-12	4	4	8	6	**22**
Std dev C3 sb20-24	6	5	5	7	**23**
Std dev C4 sb0-4	9	4	4	4	**21**
Std dev C4 sb4-8	5	8	6	3	**22**
Std dev C4 sb16-20	5	7	6	6	**24**
Energy C3 sb8-12	2	7	5	6	**20**
Energy C3 sb16-20	8	5	5	3	**21**
Energy C3 sb24-28	8	2	1	8	**19**
Energy C3 sb28-32	9	2	3	7	**21**
Energy C4 sb24-28	7	4	4	5	**20**
Energy C4 sb28-32	1	8	3	8	**20**
Rt En C3 sb4-8 and sb8-12	7	4	6	7	**24**
Rt En C3 sb20-24 and sb24-28	4	9	4	5	**22**
Rt En C3 sb24-28 and sb28-32	2	3	7	8	**20**
Entropy C3 sb0-4	7	5	3	5	**20**
Entropy C3 sb4-8	5	7	4	7	**23**
Entropy C3 sb20-24	7	9	6	5	**27**
Entropy C3 sb24-28	4	5	7	8	**24**
Entropy C3 sb28-32	7	7	7	5	**26**
Entropy C4 sb20-24	1	8	5	6	**20**
Entropy C4 sb28-32	5	4	7	4	**20**
PLV C3 and C4 sb20-24	5	8	4	4	**21**
Root mean sq C3 sb8-12	7	5	5	4	**21**
Root mean sq C3 sb12-16	8	5	6	5	**24**
Root mean sq C3 sb28-32	4	7	6	7	**24**
Root mean sq C4 sb4-8	7	4	4	6	**21**
Root mean sq C4 sb16-20	5	8	4	3	**20**
Variance C3 sb16-20	5	7	3	5	**20**
Variance C3 sb20-24	5	3	6	7	**21**
Variance C3 sb24-28	5	3	7	6	**21**

Variance C3 sb28-32	10	3	7	5	**25**
Variance C4 sb8-12	5	6	7	5	**23**
Variance C4 sb16-20	5	7	8	4	**24**

In general, there are only fourteen features (from 134) that are repeatedly selected for more than seven classifiers for at least two of the subjects (highlighted in the table in gray). This is an evidence of the necessity of an independent feature selection process for each subject, since there is not a great number of common features that generates a good classification for the majority of the subjects. Particularly, the only feature selected as relevant for three of the subjects (S4, X11 and O3) is the entropy for the sub-band 28 to 32 Hz of C3 channel. An overview of those fourteen features demonstrates the variability for each subject of the relevance of the features. As an example, the energy for channel C3 of sub-bands 24-28 Hz and 28-32 Hz was relevant for subjects S4 and A2, and irrelevant for subjects X11 and O3.

Five of the features presented the highest number of selections when summed the results of all the classifiers (twenty-five or more). Those features (with red borders in the table) are: average amplitude change from C4 and sub-bands 12-16 Hz and 16-20Hz, entropy for C3 and sub-bands 20-24 Hz and 28-32 Hz, and variance for C3 and sub-band 28-32 Hz. In general, all the features were selected as relevant for at least one subject, with the feature named mean of the absolute values being represented in the table eleven times between combination of channel and sub-band as the highest value. The PLV was the one with the smallest value (3). The other important aspect is the evaluation of the proposed feature, which was selected as relevant for at least one of the subjects for five sub-bands, all for the channel C3.

Table 11 portrays the results for the features that were selected by a lower number of classifiers (three or less) for each subject, with a reduction to the features that presented a total of selection of less than 20 (in order to decrease the dimension of the table and narrow the analysis). In this case, there are 35 features (highlighted in the table) selected as less relevant at least for two of the subjects, with eight that were selected by 3 classifiers or less for three subjects. Those were: for channel C4 the average amplitude change for 4-8 Hz, the standard deviation for 12-16 Hz, the energy for 4-8 Hz and 8-12 Hz, the ratio of the energy

between sub-bands 0-4 Hz and 4-8Hz, and the variance for 0-4 Hz and 24-28 Hz; and the PLV between channels C3 and C4 for sub-band 24-28 Hz. In general, the majority of the features were selected at least one time for a subject (98 features). The feature that presented the lowest number of selections was the PLV between channels C3 and C4 for sub-band 16-20 Hz with only ten selections between all the subjects, followed by the variance of C4 for sub-bands 0-4 Hz and 24-28 Hz.

Table 11. Features defined as less relevant for at least one subject

	S4	X11	O3	A2	Total
Avg abs C3 sb4-8	2	5	2	8	**17**
Avg abs C3 sb20-24	5	3	7	3	**18**
Avg abs C3 sb24-28	5	3	4	3	**15**
Avg abs C4 sb0-4	6	4	4	2	**16**
Avg abs C4 sb4-8	6	3	2	5	**16**
Avg abs C4 sb8-12	1	1	5	5	**12**
Avg abs C4 sb16-20	3	7	2	6	**18**
Avg abs C4 sb20-24	7	4	5	3	**19**
Avg abs C4 sb24-28	1	4	5	2	**12**
Avg ampl C3 sb4-8	5	1	4	5	**15**
Avg ampl C3 sb8-12	5	4	3	5	**17**
Avg ampl C3 sb20-24	6	2	5	6	**19**
Avg ampl C4 sb4-8	3	2	3	4	**12**
Avg ampl C4 sb8-12	3	4	7	5	**19**
Avg ampl C4 sb24-28	2	6	4	4	**16**
Std dev C3 sb0-4	4	2	2	4	**12**
Std dev C3 sb12-16	7	3	4	5	**19**
Std dev C3 sb16-20	5	4	7	1	**17**
Std dev C3 sb24-28	6	4	5	2	**17**
Std dev C3 sb28-32	5	2	7	4	**18**
Std dev C4 sb8-12	2	2	7	2	**13**
Std dev C4 sb12-16	2	3	3	5	**13**
Std dev C4 sb20-24	5	6	5	2	**18**
Std dev C4 sb24-28	1	5	3	4	**13**
Std dev C4 sb28-32	0	7	4	4	**15**
Energy C3 sb0-4	2	6	4	5	**17**
Energy C3 sb4-8	2	6	6	5	**19**
Energy C3 sb12-16	3	6	2	7	**18**
Energy C3 sb24-28	8	2	1	8	**19**
Energy C4 sb0-4	2	3	5	5	**15**

Energy C4 sb4-8	3	5	2	3	**13**
Energy C4 sb8-12	2	2	8	3	**15**
Energy C4 sb16-20	4	4	5	3	**16**
Energy C4 sb20-24	5	5	6	2	**18**
Rt En C3 sb0-4 and sb4-8	4	3	7	4	**18**
Rt En C3 sb12-16 and sb16-20	2	7	4	4	**17**
Rt En C3 sb16-20 and sb20-24	4	5	4	1	**14**
Rt En C4 sb0-4 and sb4-8	2	6	2	3	**13**
Rt En C4 sb4-8 and sb8-12	6	3	4	4	**17**
Rt En C4 sb12-16 and sb16-20	3	3	5	5	**16**
Rt En C4 sb16-20 and sb20-24	3	5	1	5	**14**
Rt En C4 sb20-24 and sb24-28	4	4	2	4	**14**
Rt En C4 sb24-28 and sb28-32	3	6	1	6	**16**
Entropy C3 sb8-12	4	2	6	4	**16**
Entropy C3 sb12-16	6	1	4	5	**16**
Entropy C3 sb16-20	4	5	2	3	**14**
Entropy C4 sb4-8	5	4	6	2	**17**
Entropy C4 sb8-12	4	2	3	5	**14**
Entropy C4 sb16-20	4	4	3	4	**15**
Entropy C4 sb24-28	5	6	5	3	**19**
PLV C3 and C4 sb0-4	8	2	1	6	**17**
PLV C3 and C4 sb4-8	2	5	4	3	**14**
PLV C3 and C4 sb12-16	4	3	4	3	**14**
PLV C3 and C4 sb16-20	1	3	1	5	**10**
PLV C3 and C4 sb24-28	6	2	3	3	**14**
PLV C3 and C4 sb28-32	6	7	3	3	**19**
Root mean sq C3 sb0-4	6	3	4	5	**18**
Root mean sq C3 sb4-8	7	3	5	4	**19**
Root mean sq C3 sb16-20	1	5	6	4	**16**
Root mean sq C3 sb20-24	6	1	6	4	**17**
Root mean sq C3 sb24-28	5	3	7	4	**19**
Root mean sq C4 sb0-4	6	3	5	4	**18**
Root mean sq C4 sb8-12	5	2	6	3	**16**
Root mean sq C4 sb12-16	4	4	6	1	**15**
Root mean sq C4 sb20-24	2	5	6	6	**19**
Root mean sq C4 sb24-28	3	4	4	7	**18**
Root mean sq C4 sb28-32	0	8	4	6	**18**
Variance C3 sb0-4	3	3	5	4	**15**
Variance C3 sb4-8	5	3	3	7	**18**
Variance C3 sb8-12	0	6	5	3	**14**
Variance C4 sb0-4	3	5	1	2	**11**

Variance C4 sb12-16	5	3	6	4	**18**
Variance C4 sb20-24	2	6	4	4	**16**
Variance C4 sb24-28	4	3	2	2	**11**
Variance C4 sb28-32	1	5	2	5	**13**

In order to analyze the relevance of the sub-bands selected for the proposed model, Table 12 reorganize the data from the features that can be defined as more relevant (selected by seven classifiers or more), displaying the number of relevant features defined for each sub-band. The results can be observed in Table 12.

Table 12. Relevant features divided in frequency bands

		S4	X11	O3	A2	Total
0-4 Hz	**Avg abs C3**	10	3	6	5	24
	Avg ampl C3	9	6	2	3	20
	Avg ampl C4	7	5	4	5	21
	Std dev C4	9	4	4	4	21
	Rt En C3	4	3	7	4	18
	Entropy C3	7	5	3	5	20
	PLV C3 and C4	8	2	1	6	17
4-8 Hz	**Avg abs C3**	2	5	2	8	17
	Std dev C4	5	8	6	3	22
	Rt En C3	4	3	7	4	18
	Entropy C3	5	7	4	7	23
	Root mean sq C3	7	3	5	4	19
	Root mean sq C4	7	4	4	6	21
	Variance C3	5	3	3	7	18
8-12 Hz	**Avg abs C3**	8	7	4	1	20
	Avg ampl C4	3	4	7	5	19
	Std dev C3	4	4	8	6	22
	Std dev C4	2	2	7	2	13
	Energy C3	2	7	5	6	20
	Energy C4	2	2	8	3	15
	Rt En C3	7	4	6	7	24
	Root mean sq C3	7	5	5	4	21
	Variance C4	5	6	7	5	23
12-16 Hz	**Avg abs C3**	4	4	7	5	20
	Avg abs C4	5	4	7	5	21
	Avg ampl C3	5	6	6	7	24
	Avg ampl C4	6	5	9	5	25

	Std dev C3	7	3	4	5	19
	Energy C3	3	6	2	7	18
	Rt En C3	2	7	4	4	17
	Root mean sq C3	8	5	6	5	24
16-20 Hz	**Avg abs C3**	2	6	4	8	20
	Avg abs C4	3	7	2	6	18
	Avg ampl C4	6	7	4	8	25
	Std dev C3	5	4	7	1	17
	Std dev C4	5	7	6	6	24
	Energy C3	8	5	5	3	21
	Rt En C3	2	7	4	4	17
	Root mean sq C4	5	8	4	3	20
	Variance C3	5	7	3	5	20
	Variance C4	5	7	8	4	24
20-24 Hz	**Avg abs C3**	5	3	7	3	18
	Avg abs C4	7	4	5	3	19
	Std dev C3	6	5	5	7	23
	Rt En C3	4	9	4	5	22
	Entropy C3	7	9	6	5	27
	Entropy C4	1	8	5	6	20
	PLV C3 and C4	5	8	4	4	21
	Variance C3	5	3	6	7	21
24-28 Hz	**Energy C3**	8	2	1	8	19
	Energy C4	7	4	4	5	20
	Rt En C3	4	9	4	5	22
	Entropy C3	4	5	7	8	24
	Root mean sq C3	5	3	7	4	19
	Root mean sq C4	3	4	4	7	18
	Variance C3	5	3	7	6	21
28-32 Hz	**Avg abs C3**	8	3	4	5	20
	Avg abs C4	2	7	7	4	20
	Avg ampl C3	6	5	7	6	24
	Avg ampl C4	4	6	7	6	23
	Std dev C3	5	2	7	4	18
	Std dev C4	0	7	4	4	15
	Energy C3	9	2	3	7	21
	Energy C4	1	8	3	8	20
	Rt En C3	2	3	7	8	20
	Entropy C3	7	7	7	5	26
	Entropy C4	5	4	7	4	20

	PLV C3 and C4	6	7	3	3	19
	Root mean sq C3	4	7	6	7	24
	Root mean sq C4	0	8	4	6	18
	Variance C3	10	3	7	5	25

In general, all the sub-bands possess several number of features among the most relevant for at least one of the subject, being the ones with a lower number of features the 0-4 Hz, 4-8 Hz and 24-28 Hz sub-bands (with 7 features) and the one with the higher number of features the 28-32 Hz sub-band (with 15 features). The obtained results are evidence that the frequency scope selected represent in a diverse and correct manner the EEG signal produced by the MI tasks, with any sub-band presenting a low number of selected features (including the ones that are not commonly employed for the representation of this MI tasks). In particular, the observation of the number of features for each channel demonstrates that channel C3 provides almost double the number of selected features than C4 (43 and 25, respectively).

4.4 Performance Analysis with Feature Selection and Ensemble of Classifiers

This next experiment is carried out with the Ensemble approach and different fusion techniques. The main objective is to verify the fusion technique that offers the best accuracy. Subjects K3 from the dataset IIIa of the BCI competition III and A01 from the dataset IIa from the BCI competition IV are added to the analysis.

Subject K3 is added with the intention of assisting in the assessment of the response of the proposed model to EEG signals characterized by higher frequency sampling and lower number of trials. Being the main objective of this work the analysis of MI with the already tested two classes (left and right hand) and the electrodes C3 and C4, the other electrodes and classes mentioned in the description of the dataset (Section 4.1) were not included in the analysis. The resulting number of training samples is 90, as well as 90 for testing purposes.

The A01 subject is added with the intent of analyzing the robustness of the system to artifacts in the signals (in this case EOG). The electrodes and classes are

the same as above. The number of training trials is 144, with 143 testing trials (after removal of noisy trials during the preprocessing stage). The resultant classification accuracies are shown in Table 13. Eleven different approaches combining the result of the classifiers with fusion methods were evaluated:

- Naïve Bayes Combiner;
- Majority Voting and Weighted Voting applied on the data without feature selection;
- Majority Voting and Weighted Voting applied on the data with feature selection;
- GA for classifier selection, in which the results of the selected classifiers are combined using the Weighted Voting technique;
- Three approaches of GA for the definitions of the weights of each classifier in the Weighted Voting: one with the weights between 0 and 1, other with the same condition but summing 1, and another with weights between 1 and -1 summing 1;
- A combination of GA for classifier selection and another GA defining the weights for the Weighted Voting method that combines the results of the selected classifiers;
- MLP as meta-classifier.

Two combinations of a reduced set of classifiers chosen randomly and one with all the classifiers, were applied on the Majority Voting and Weighted Voting approaches, in order to observe the effect of reducing the number of classifiers on the classification performance. In the simple Weighted Majority Voting combiner, the weights were defined as the performances obtained for each individual classifier. The application of the MLP as a meta-classifier was accomplished testing 10 neural networks for each different configuration of 5 to 20 neurons in

the hidden layer. The obtained Confusion Matrices, for each method, are presented in Appendix 4.

The best fusion method is the MLP as a meta-classifier, with a difference of 9.9741% over the second best one (GA for classifier selection). The accuracy of this model is really high, reaching over the 90% for almost all subjects (underperforming just in X11). In addition, for the A01 subject that presented EOG artifacts, when almost all the other fusion methods presented accuracies below 80%, this method reached a value as high as 90.9091%. In comparison with the individual classifiers results, it can be seen that the MLP as fusion method outperforms the performances of each individual classifier for the tested subjects (S4, X11, O3, and A2), which validates our approach to the ensemble model application.

Table 13. Performances of the Ensemble of Classifier in % for the different fusion algorithms

	Classifier	S4	X11	O3	A2	K3	A01	Mean
Na Bayes		78.24±0.00	70.56±0.00	71.34±0.00	81.43±0.00	92.22±0.00	61.54±0.00	75.89±10.56
ority ting		77.67±0.82	74.07±0.71	72.61±1.10	81.43±0.00	92.22±0.00	63.64±0.00	76.94±9.57
ghted ting		78.80±0.00	74.07±0.00	77.71±0.00	81.43±0.00	91.11±0.00	63.64±0.00	77.79±9.01
FS ajority ting	SVM,LDA, KNN,PNN, RBF,MLP	79.93±0.71	74.26±0.28	75.16±1.10	81.43±0.42	92.22±0.64	65.73±1.06	78.12±8.83
	SVM,LDA, SVMQ,MLP	78.99±0.68	74.44±0.39	80.26±0.97	82.86±1.09	92.22±0.64	65.04±1.46	78.97±9.02
		81.05±0.99	75.37±0.98	82.80±1.47	83.57±1.49	91.11±0.64	65.73±1.45	79.94±8.60
FS Weighted Voting	SVM,LDA, KNN,PNN, RBF,MLP	80.11±0.00	75.00±0.00	75.16±0.00	81.43±0.00	93.33±0.00	65.04±0.00	78.35±9.33
	SVM,LDA, SVMQ,MLP	81.24±0.00	76.67±0.00	80.89±0.00	82.86±0.00	92.22±0.00	65.73±0.00	79.94±8.66
		81.24±0.00	75.74±0.00	85.98±0.00	85.71±0.00	91.11±0.00	66.43±0.00	81.04±8.82
GA sel clas		80.68±0.32	76.30±0.56	79.62±1.91	87.86±2.51	94.44±1.11	80.42±3.16	83.22±6.67
GA sel weights		80.11±1.70	76.11±1.48	82.80±2.87	87.86±2.51	93.33±1.70	69.93±2.46	81.69±8.33
GA sel weights sum 1 de 0 a 1		81.43±0.61	83.52±0.93	80.89±1.91	58.57±1.8	94.44±1.11	81.82±1.76	80.11±11.72
GA sel weights sum 1 de -1 a 1		84.24±2.63	79.63±2.98	70.15±3.51	83.57±2.86	90.00±1.70	72.03±2.8	79.94±7.63
GAs combined		81.80±0.61	80.19±1.44	73.89±2.05	91.43±1.8	93.33±0.64	76.92±1.76	82.93±7.84
MLP as expert		**90.43±0.78**	**87.59±1.16**	**94.91±0.97**	**96.43±1.09**	**98.89±0.64**	**90.91±1.07**	**93.19±4.25**

The worst combination methods were the Naïve Bayes combiner and the Majority Voting without feature selection, which presented a mean even lower

that some of the individual classifiers (SVM linear and SVM quadratic). All others techniques presented values above the single classifiers, except for the MLP classifier, which performed much better than other single classifiers.

The GA applied as classifier selector was the second most successful method, with an average performance of 83.2184%. This result highlights the irrelevant contribution of some classifiers (see Table 14) since the weighted methods were not able to totally zero the contribution of some classifiers, generating lower mean classification accuracies. In general, for at least one subject all the classifiers were selected. The classifiers with a least relevant contribution (only one selection) were the PNN, SVM with linear kernel and SVM with quadratic kernel. On the contrary, there were classifiers with a great contribution to the final performance of the models. Those were the RBF and the MLP, both selected for the final ensemble in all subjects.

A particular case is the weight definition by the GA that summed 1 in the [0,1] range. This method presented promising accuracies for subjects S4, X11 and O3. However, for the A2 subject, its accuracy drastically fell to 58.5714. If the contribution of this subject is eliminated from the average calculation, the results of the method become very promising, surpassing the ones of the GA as classifier selector. The two of them were the other methods that performed fairly well with the artifact dataset (subject A01).

Table 14. Best solution for the GA employed as classifier selector

Classifiers	S4	X11	O3	A2	K3	A01
PNN	0	0	1	0	0	0
SVM-l	0	0	0	0	1	0
SVM-q	0	0	0	1	0	0
LDA-l	0	0	1	1	0	0
RBF	1	1	1	1	1	1
KNN-e	1	0	0	0	1	1
KNN-m	0	1	0	1	0	1
KNN-cos	1	0	1	0	0	1
KNN-cor	0	0	1	1	0	1
MLP	1	1	1	1	1	1

The GA combination approach demonstrated to be convenient for the classification of MI data, being the third highest performance. This is reasonable, giving that the purpose of the method is to eliminate the non-useful classifiers, while weighting the contribution of the useful ones. The performance of this method is not very different from the one of the GA for classifier selection. This can be explained by the weights definition, which can reduce the contribution of some of the selected classifiers to non-significant values, which can decrease the diversity of the ensemble in a sense of decrease the relevance of those classifiers in the final decision.

The number of training examples seems to affect the performance of the model as well. Higher performances were obtained for subjects O3 and A2 that presented a much lower number of trials than S4 and X11. Those results are in accordance with the ones presented in the literature (see details in the next section). This can be an evidence of the variability of the EEG signals between sessions with the same subject. The recordings of subject A2 were conducted in just one session, differently from that the three sessions of O3, S4 and X11. The O3 subject, particularly, presented an error of duplication of trials in the published dataset, reason why the number of trials was reduced to half of the original trials, which probably reduced the number of sessions on the dataset.

The results with the subject with an augmented sampling frequency (250Hz) indicate that a higher sampling frequency provides a better description of the features attaining good accuracies for each fusion method and peaking at 98.8889% for the MLP as meta-classifier algorithm. As mentioned in the previous analyses for the subject A01 that presents EOG artifacts, the majority of methods did not reach even the 70% of performance. However, methods as GA for classifier selection, GA for weights definition with sum 1 and in [-1, 1] range, GA combined and the MLP as meta-classifier were robust to this problem, especially the MLP method with an accuracy as good as 90.9091%.

Observing the results shown in Appendix 4, it can be seen that the models had difficulty in classifying the right hand imagery movement for subjects O3, K3 and A01. In particular for O3, this problem was detected in all the previous analysis with the Confusion Matrices. The other subject that presented a tendency for misclassification a particular class was S4, in this case for the left hand. In the case of the fusion models, five of them (Naïve Bayes, Majority Voting with

Feature Selection for six and 4 classifiers, and Weighted Voting with Feature Selection for all and six classifiers) tended to misclassify the right hand class. Majority Voting and Weighted Voting with Feature Selection, and MLP as meta-classifier, presented the most balanced results.

Another approach to the analysis of the results is the statistical validation, which allows determining the significance of the results and whether the conclusions obtained are supported by the experimentation settings. The Friedman test [227] was selected, with the null hypothesis stating that all the algorithms present the same rank due to their equivalence [228]. Therefore, if the hypothesis is rejected, it means that there are significant differences between the tested algorithms. If this condition is fulfilled, then a post-hoc analysis can be conducted to found whether there is a relevant difference between the defined control algorithm and the others. For this post-hoc analysis the Holm method [229] was selected. This method implements a pairwise comparison between a defined control classifier and other classifiers. The Friedman's ranking results are shown in the next table.

Table 15. Ranks of the algorithms for the Friedman's test

	Classifier	S4	X11	O3	A2	K3	A01	Mean
Naïve Bayes		14	15	14	12	9	15	13.17
Majority Voting		15	13.5	13	12	9	13.5	12.67
Weighted Voting		13	13.5	9	12	13	13.5	12.33
FS Majority Voting	SVM,LDA,KNN, PNN, RBF,MLP	11	12	10.5	12	9	9	10.58
	SVM,LDA, SVMQ,MLP	12	11	7	8.5	9	11.5	9.83
		7	9	3.5	6.5	13	9	8
FS Weighted Voting	SVM,LDA,KNN, PNN,RBF,MLP	9.5	10	10.5	12	5	11.5	9.75
	SVM,LDA, SVMQ,MLP	5.5	5	5.5	8.5	9	9	7.08
		5.5	8	2	5	13	7	6.75
GA sel clas		8	6	8	3.5	2.5	3	5.17
GA sel weights		9.5	7	3.5	3.5	5	6	5.75
GA sel weights sum 1 de 0 a 1		4	2	5.5	15	2.5	2	5.17
GA sel weights sum 1 de -1 a 1		2	4	15	6.5	15	5	7.92
GAs combined		3	3	12	2	5	4	4.83
MLP as expert		1	1	1	1	1	1	1

The observation of the mean rankings confirms that the best algorithm is the MLP as meta-classifier, followed by the combination of GAs, the GA as classifier selector and the weight definition by the GA that summed 1 with limits in the range of 0 to 1 (this last two sharing the spot for third better rank). Table 15 also confirms that the worst algorithms are Naïve Bayes, Majority Voting without feature selection and Weighted Voting. The two ensembles that did not apply feature selection methods were included among the worst results (Majority and Weighted Voting), which reiterate the importance of the feature selection stage for the model.

The Friedman statistic is distributed then according to the approximation (χ_F^2) [227] with k-1 degrees of freedom (with k being the number of tested algorithms)[1]. In this analysis, the obtained value of χ_F^2 is 48.8708. For a p<0.001, the critical value for values of k*N bigger than 30 [230] (with N being the number of subjects) or k and N bigger than 5 or 13 respectively [231] can be found in the Table of Chi-Square Distribution [227], which in this case is 36.12. As the value of the chi-square approximation of the Friedman statistic is greater than the critical value then the null hypothesis of equal ranks between all models can be rejected.

The MLP as meta-classifier was selected as the control method of the Holm post-hoc technique. This analysis organizes the algorithms in decreasing order and calculates a z value for all non-control algorithms. With the obtained z values the corresponding p-values are found using the normal distribution table (Table A1 in [227]). Next, each p-value is compared to an adjusted critical value calculated by the Holms's expression: $\left(p_H = \frac{\alpha}{k-i}\right)$. Then, if the p-value is lower than the adjusted critical value, the null hypothesis is rejected. The level of significance (α) was defined as 0.1, one of the commonly selected value for this test [232] [233] [234]. The obtained results for α=0.1 are illustrated in Table 16 (with the adjusted critical value displayed as the Holm column).

As it can be observed from Table 16, the MLP as meta-classifier significantly outperformed the first ten algorithms, being statistically similar to the GA as weighted selector, the GA as classifier selector, the weight definition

[1] The calculated value of the chi squared approximation is compared to the value of the chi-square distribution, and if it is greater, the null hypothesis is rejected.

by the GA that summed 1 with limits in the range of 0 to 1, and the GA combination.

Table 16. Results of the Holm procedure for comparison with a control algorithm

	Classifiers	$Z = (R_i - R_j)/SE$	p-value	Holm
Naïve Bayes		4.7121	< 0.0001	0.0071
Majority Voting		4.5185	< 0.0001	0.0077
Weighted Voting		4.3894	< 0.0001	0.0083
FS Majority Voting	SVM,LDA,KNN,PNN, RBF,MLP	3.7116	0.0001	0.0091
FS Majority Voting	SVM,LDA,SVMQ,MLP	3.4211	0.0003	0.01
FS Weighted Voting	SVM,LDA,KNN,PNN, RBF,MLP	3.3889	0.0004	0.0111
FS Majority Voting		2.7111	0.0034	0.0125
GA sel weights sum 1 de -1 a 1		2.6788	0.0037	0.0143
FS Weighted Voting	SVM,LDA,SVMQ,MLP	2.3561	0.0092	0.0167
FS Weighted Voting		2.227	0.0130	0.02
GA sel weights		1.8397	0.0329	0.025
GA sel clas		1.6137	0.0533	0.0333
GA sel weights sum 1 de 0 a 1		1.6137	0.0533	0.05
GAs combined		1.4846	0.0688	0.1

4.5 Benchmark comparison

With the purpose of validating the results obtained in this work, this section presents the comparison of the proposed model with the results of other teams that participated in the BCI competition III, as well as with recent works found in the literature. For this analysis, the subjects were limited to the ones with a lower sampling frequency (similar to the data that will be presented in the next chapter) and artifacts free (giving a better scope to the comparison). The comparison is presented in Table 17.

Table 17. Benchmark performances

	A2	O3	S4	X11	Classifier	Training (%)	Test (%)
Proposed method	**96.43**	**94.91**	**90.43**	**87.59**	Ensemble fused by MLP	50	50
Lemm et. al [30]		**89.31**	**88.52**	83.33	Probabilistic Model	50	50
Burmeister et. al [30]		85.53	77.04	77.78	SVM	50	50

Pei et. al [30]		86.79	82.41	83.52	Fisher Discriminant Analysis	50	50
Parini et. al [30]		76.10	75.56	75.93	LDA	50	50
Coyle et. al [30]		88.05	78.52	81.30	LDA	50	50
Yao et. al [30]		**89.31**	86.48	74.81	SVM	50	50
Tavakolian et. al [30]		65.72	61.48	71.30	BNS	50	50
Lotte [219]		86.70	74.70	75.70	FIS	50	50
		86.80	75.90	75.40	MLP	50	50
		86.60	75.50	74.60	SVM	50	50
		84.10	71.80	72.70	Linear Classifier	50	50
Brodu et. al [31]	80.70		81.70	80.90	LDA	50	50
Wu et. al [220]			78.70		BPR	33.33	67.77
			87.30		Adaptative BPR	33.33	67.77
	85.70				BPR	50	50
	80.70				SVM	50	50
	79.30				LDA	50	50
Ahangi et. al [34]	84.28				KNN	50	50
	68.75				Naive Bayesian	50	50
	74.00				MLP	50	50
	87.86				LDA	50	50
	88.57				SVM	50	50
	73.64				Mean	50	50
	69.57				Max	50	50
	84.35				Median	50	50
	89.14				Adaboost	50	50
	89.56				BKS	50	50
	88.85				Bagging	50	50
	88.57				Decision Template	50	50
	87.35				Majority Voting	50	50
	90.00				Weighting Majority Voting	50	50
Omar et. al [221]		**89.38**			LDA	50	50
		92.50			QDA	50	50
Djemili et. al [222]	69.30	79.10	75.50	78.90	LDA	50	50
	68.00	78.00	80.00	85.00	SVM	50	50
Sampanna et. al [35]	**92.93**				Ensemble of SVMs	50	50
Chen et. al [223]	**90.71**	84.78	70.16	72.25	LDA + AFA	50	50
	90.71	85.53	76.95	73.18	LDA + AFAPS	50	50
Bashashati. et al [32]		80.50	77.96	78.15	Boosting	50	50
		82.39	83.89	78.15	Logistic	50	50
		79.25	79.26	77.78	Random Forest	50	50
		81.76	83.52	77.78	SVM	50	50
		81.13	81.11	76.48	LDA	50	50
		79.25	72.41	74.81	QDA	50	50
		83.65	82.22	76.67	MLP	50	50
Lin et. al [224]	70.67		79.31	82.89	C-SVM	50	50
	71.09		78.70	81.53	nu-SVM	50	50
	68.32		80.37	79.71	SVM-polynomial	50	50
	70.25		78.80	80.29	SVM-radial basis function	50	50
Tan et. al [225]	82.10	70.10			LDA	50	50
	83.50	74.10			SVM	50	50
	85.00	79.10			ELM	50	50

Bashashati et. al [36]		86.10	84.20	**86.46**	MLR	50	50
		83.77	83.30	83.89	Voting	50	50
		83.84	83.35	**84.98**	Averaging	50	50
		84.28	83.15	80.74	Max	50	50
Tan. et al [226]	84.20	77.10			LDA	50	50
	85.30	82.25			SVM	50	50
	84.90	79.00			ELM	50	50
	79.30	85.90			Weight-PM	50	50
	86.40	85.75			ELM-PM	50	50

The proposed model presented the best classification accuracy for all subjects. Table 17 also highlights, for each subject, the other two or three results that are closer to the performance of our method. It is worth emphasizing the inter-subject robustness of the proposed method, being the highest value of all previous works. The works that applied multiple classifier strategies were [32], [34], [35], and [36], and their results suggest that any form of application of ensemble of classifiers can offer good classification accuracies, being the highest results for two of the subjects: the Ensemble of SVMs proposed by [35] for A2 and the multi-response linear regression (MLR) of [36] for X11.

4.6 Ensemble optimization

Although the proposed model (the ten classifiers fused by the MLP as meta-classifier) offered the best classification performance, it presents a high computational cost in the training process. In order to reduce this cost a backward elimination of the classifiers that form the multiple classifier system was implemented, with the main measure defined as the mean of the performances obtained with subjects S4, X11, O3, and A2. Table 18 describes the best results for each ensemble configuration, after testing all the possibilities in a backward elimination way, depending on the number of classifiers.

The first inference that can be realized from the results presented in Table 18 is that the number of classifiers is important for the accuracy of the system. This highlights the importance of the diversity of the ensemble design. If a reduction of the computational cost is required, then, depending on the level of the accepted reduction in accuracy, 5 classifier ensembles can be selected, with a mean performance of 89.8309, or even a drastic reduction with 3 classifiers with 89.7455% of correct classifications.

Table 18. Mean performances obtained for the best combinations of classifiers

Ensemble of Classifiers	Mean
MLP, SVM, SVMQ, LDA, RBF, KNN, KNN-COS, KNN-M, PNN	91.23
MLP, SVM, LDA, RBF, KNN, KNN-COS, KNN-M, PNN	90.89
MLP, SVM, RBF, KNN, KNN-COS, KNN-M, PNN	90.37
MLP, SVM, RBF, KNN-COS, KNN-M, PNN	90.13
MLP, SVM, KNN-COS, KNN-M, PNN	89.83
MLP, SVM, KNN-COS, PNN	89.45
MLP, SVM, KNN-COS	89.75
MLP, SVM, PNN	89.75

Another way to reduce the computational cost of the model is to focus on the factor that represents the heaviest load on that cost: the GA as feature selector. Considering the importance of the diversity for the accuracy of the model, a modified version of the proposed model can be devised, in which the final ensemble is composed of classifiers with feature selection and others without it. The following analysis applies a Forward Selection technique (sequentially testing each possibility and selecting the one that obtained the best performance) to evaluate which combination of classifiers gives a better result depending on the number of classifiers with feature selection. The ensemble is always composed by the ten original classifiers. The analysis assumed that only half of the classifiers can use feature selection, since more than that will not provide a significant reduction on the final computational cost. The results can be observed in the Table 19, with the classifiers with feature selection in the first column.

Table 19. Best combination of classifiers with feature selection and classifiers without feature selection

	S4	O3	X11	A2	Mean
Without FS	86.68	92.36	81.85	93.57	88.62
MLP	88.37	97.45	84.26	97.14	91.81
MLP+SVMQ	88.74	84.82	96.82	97.14	91.88
MLP+SVMQ+KNNCOS	89.87	85.37	97.45	97.14	**92.46**
MLP+SVMQ+KNNCOS+KNN	89.68	85.00	97.45	96.43	92.14
MLP+SVMQ+KNNCOS+KNN+PNN	89.31	85.56	96.82	96.43	92.03

As it can be seen, the best result is obtained with only three classifiers with feature selection: the MLP, the SVM with quadratic kernel and the k-NN with cosine metric. This confirms that this approach can obtain a good mean performance, maintaining the diversity of the ensemble and reducing in a great scale the training time (see Table 20). Another important conclusion is the difference in the mean performance of the ensemble without any feature selection and the ensemble with one classifier with feature selection (the MLP), which is a little over 2 percentage points. This fact confirms that although the diversity presents great relevance, the feature selection stage is fundamental to improve the accuracy of the model.

The gradual reduction of the number of generations and population size without a drastic change of the obtained performance can produce a significant variation on the computational cost. Therefore, the next analysis addresses the problem for each classifier individually, comparing the results of the GA for subjects S4, X11, O3, and A2 for each classifier. The purpose of the analysis is to decide until which number of solutions it is safe to reduce the computing of the algorithm without a significant reduction in the performance.

In general, from the information of the next figures (Figures 17 to 26) it can be inferred that for every classifier a reduction on the number of evaluations can be obtained without significant decrease in accuracy. Significantly, for the PNN, SVM, SVM with quadratic kernel and k-NN the search for solutions can be reduced in more than a half, with convergences for each of the subjects for 640,360, 600 and 750 respectively. For the group conformed by the LDA, k-NN with mahalanobis, correlation and cosine metrics, the reduction can be between 39.65 percentage points (7000 for the LDA) and 43.86 percentage points (1000 for the k-NN with cosine metric). A smaller proportion occurs with RBF and MLP, with 35 (480) and 33.34 (700) percentage points respectively. An interesting result is the one obtained by the SVM for subjects O3 and A2, with no evidence of evolution. This can be an evidence of the robustness of the system to the “curse of dimensionality”, given that for those subjects that have lower number of instances, a variation on the number of features does not produce a change in their accuracy.

Finally, Table 20 presents the comparison of the training time for the main tested BCI models. It is important to notice that the duration of the training

depends on the number of instances (trials) being employed. Two quantities are illustrated: 538 from the X11 subject and 140 from the A2 subject (the highest and the lowest values). The results are described in the next table, where FS denotes Feature Selection and the Best Combination models correspond to the combination of only algorithms with feature selection (as illustrated in Table 18), with three variations of the number of classifiers (seven, five, and three). To have a reference of the classifiers present in each combination see Tables 18 and 19. The tests were conducted using Matlab 2016b, on a computer (PC) with a processor Intel Core i5-750 (2000MHz) and 8Gb of DDR3 RAM Memory.

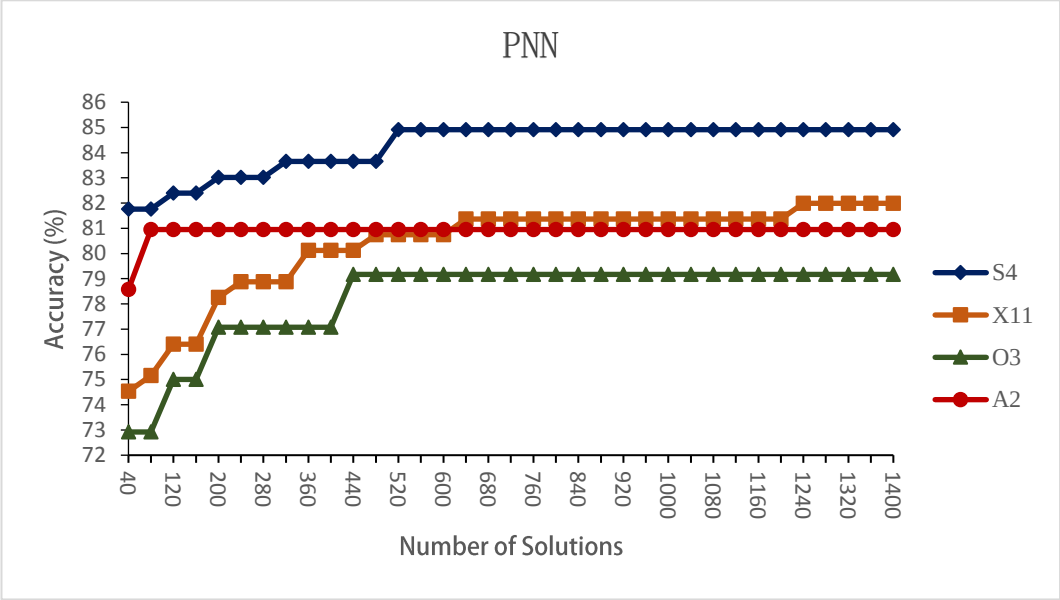

Figure 17. Accuracy of the GA for feature selection for the PNN classifier

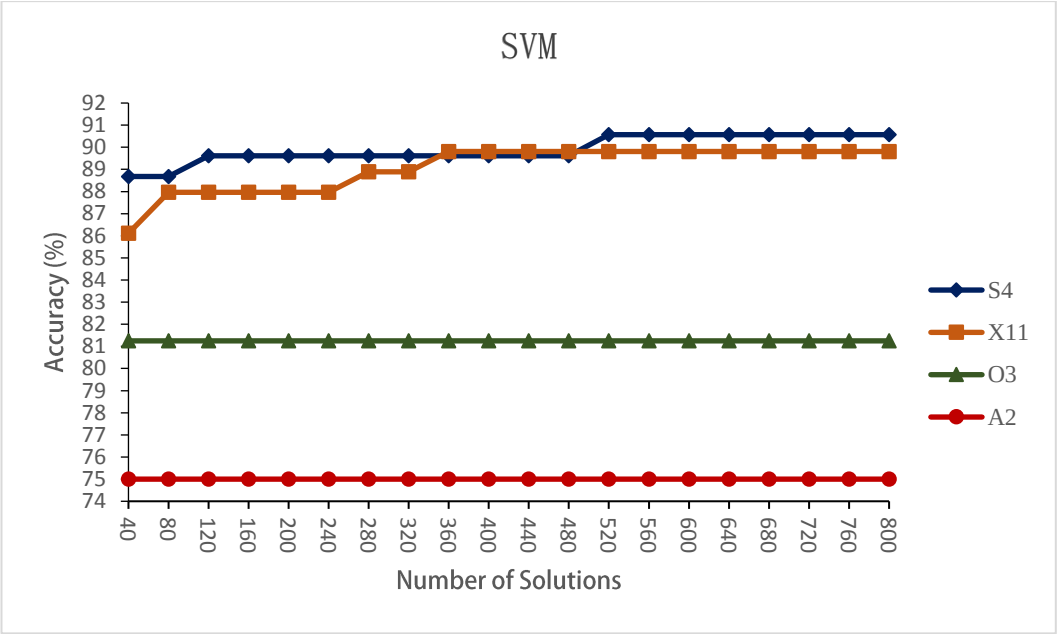

Figure 18. Accuracy of the GA for feature selection for the SVM classifier

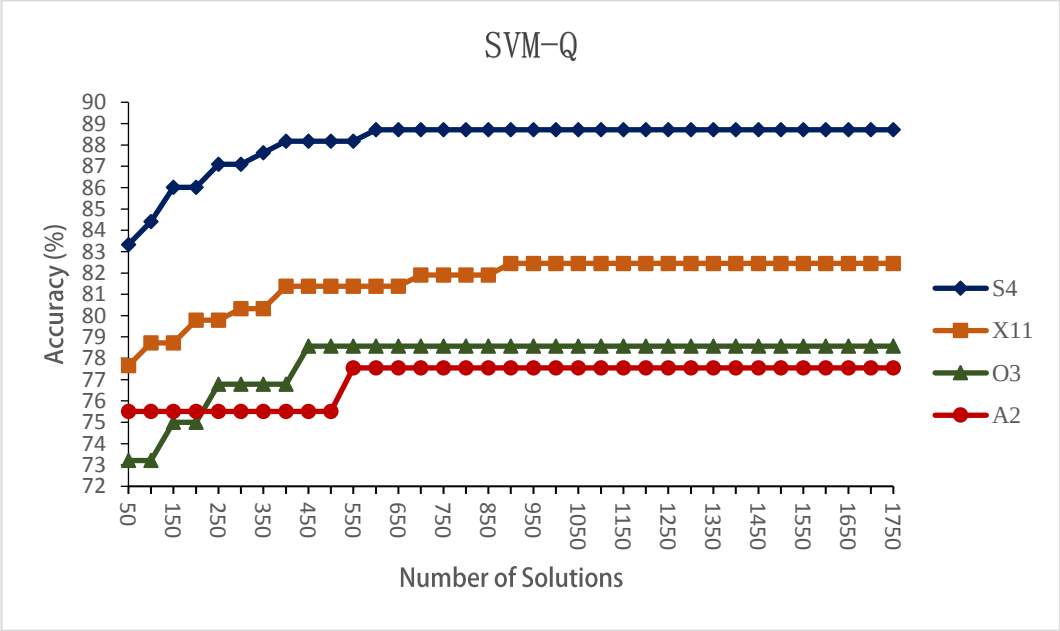

Figure 19. Accuracy of the GA for feature selection for the SVM classifier with quadratic kernel

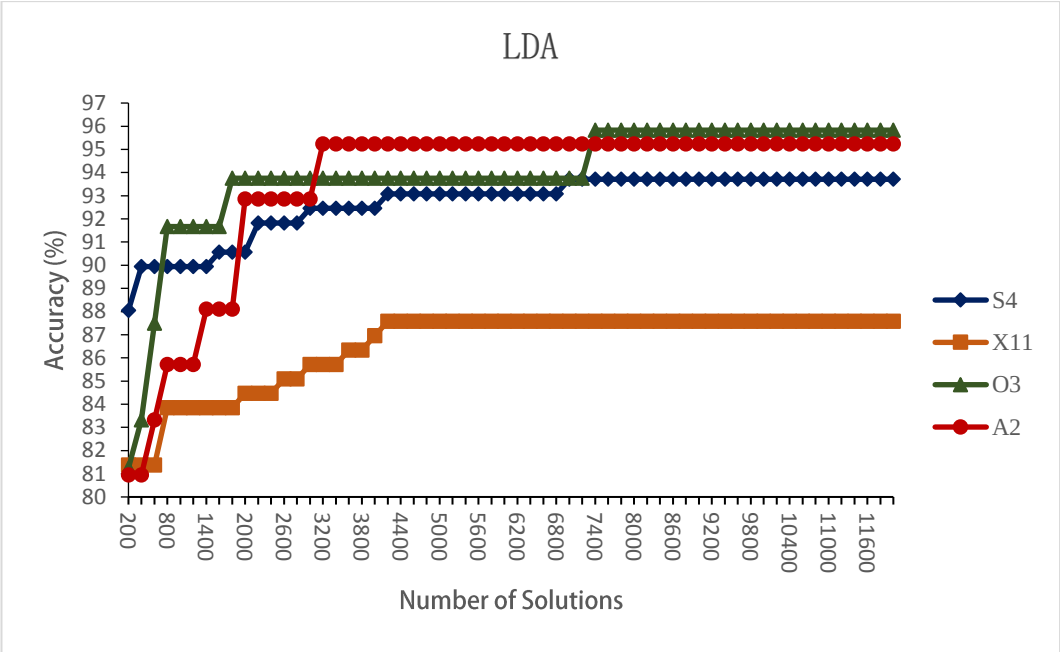

Figure 20. Accuracy of the GA for feature selection for the LDA classifier

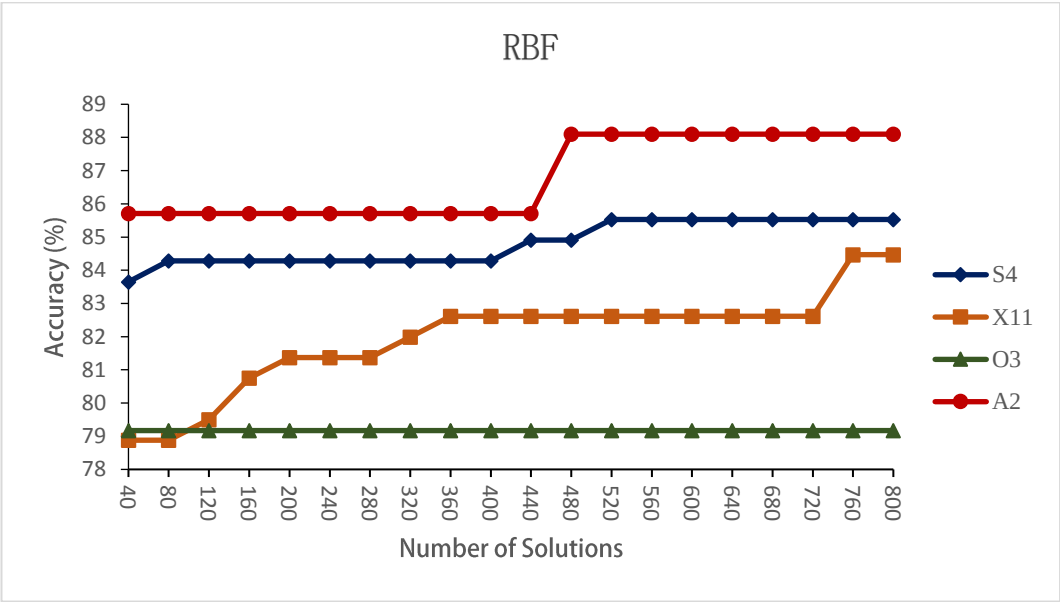

Figure 21. Accuracy of the GA for feature selection for the RBF classifier

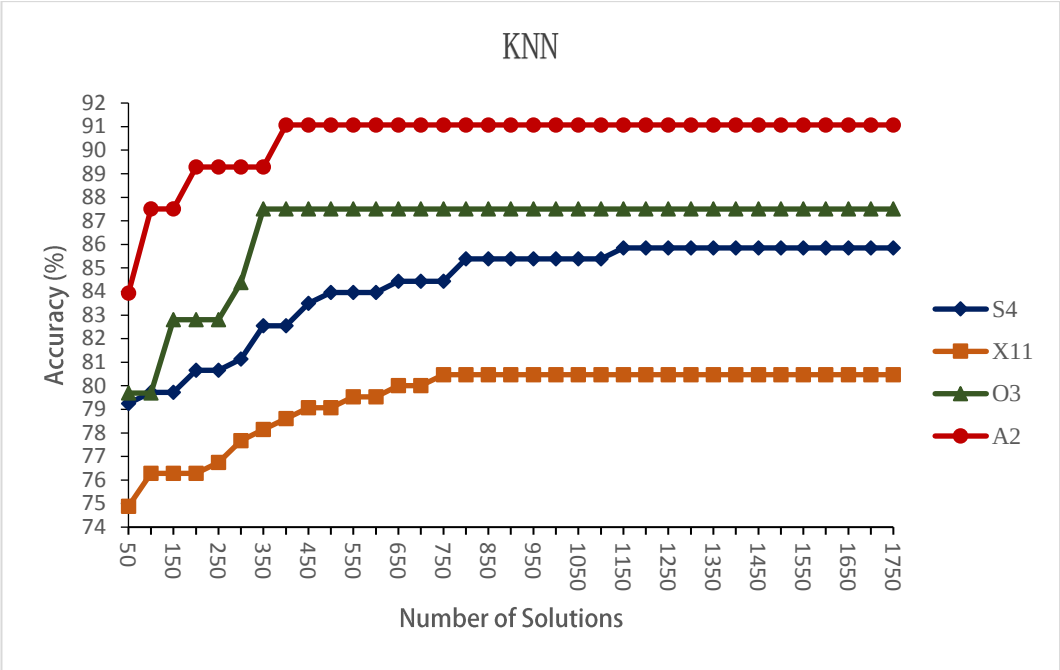

Figure 22. Accuracy of the GA for feature selection for the k-NN classifier

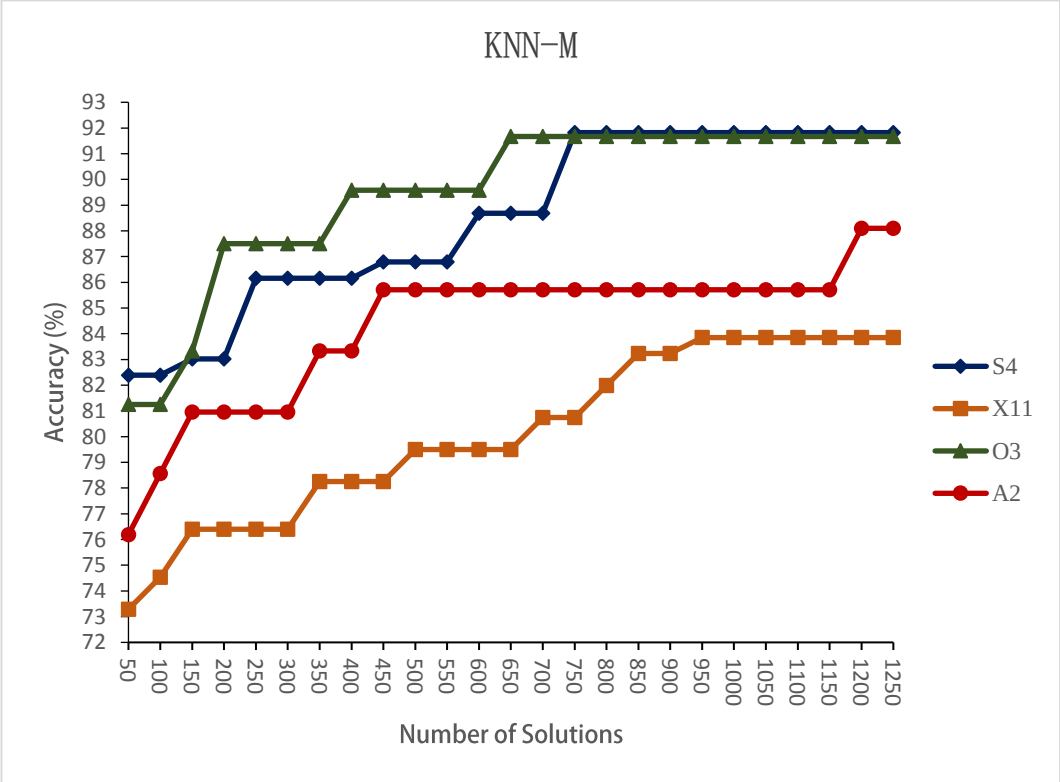

Figure 23. Accuracy of the GA for feature selection for the k-NN classifier with mahalanobis distance

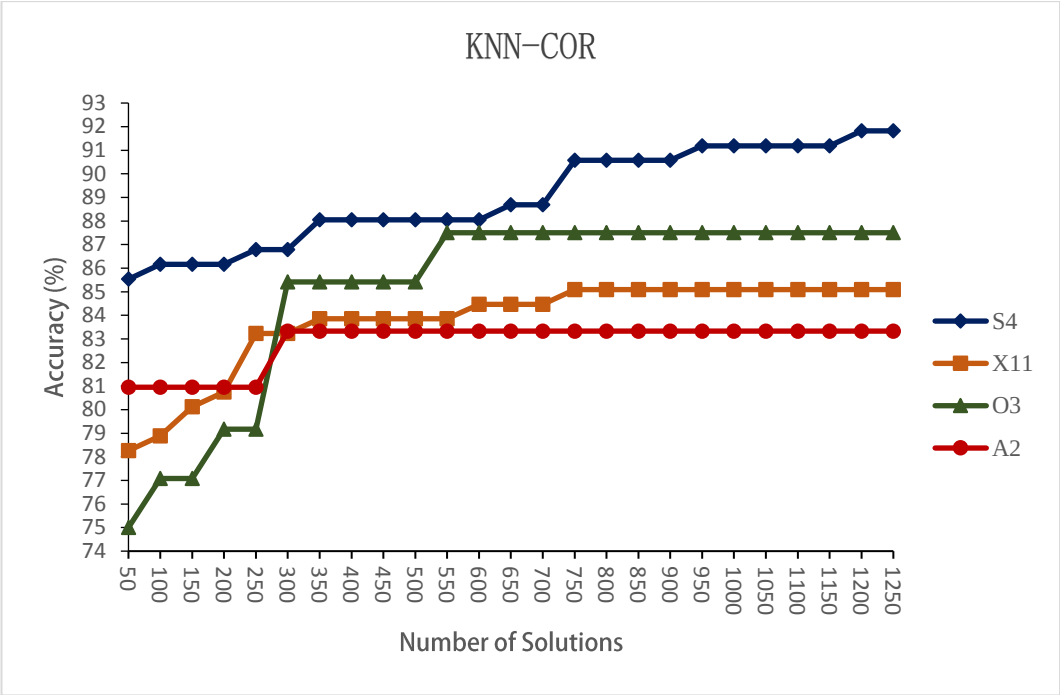

Figure 24. Accuracy of the GA for feature selection for the k-NN classifier with correlation

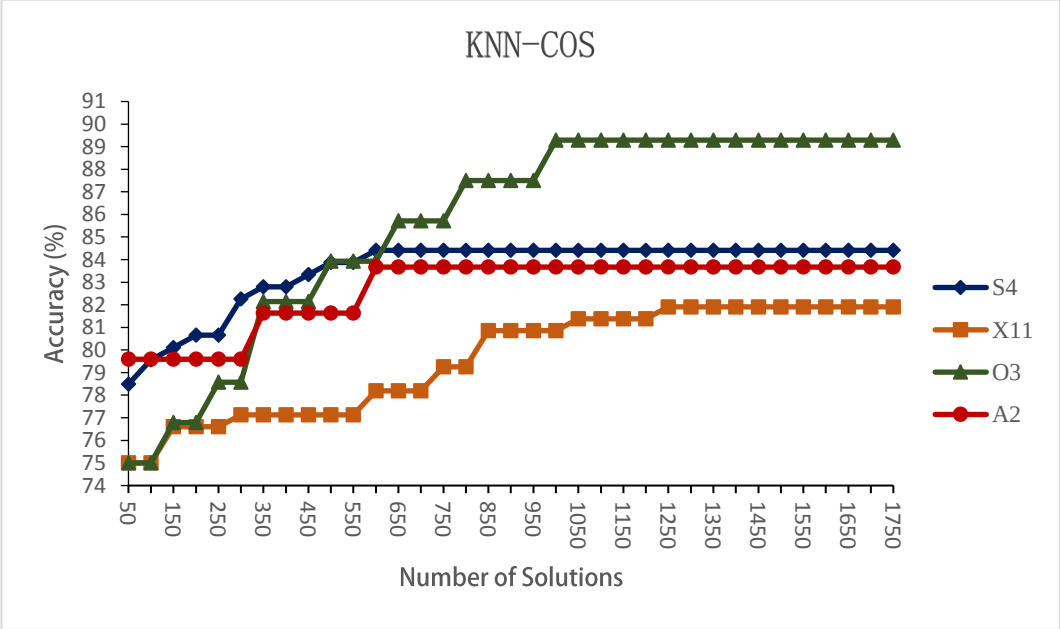

Figure 25. Accuracy of the GA for feature selection for the k-NN classifier with cosine

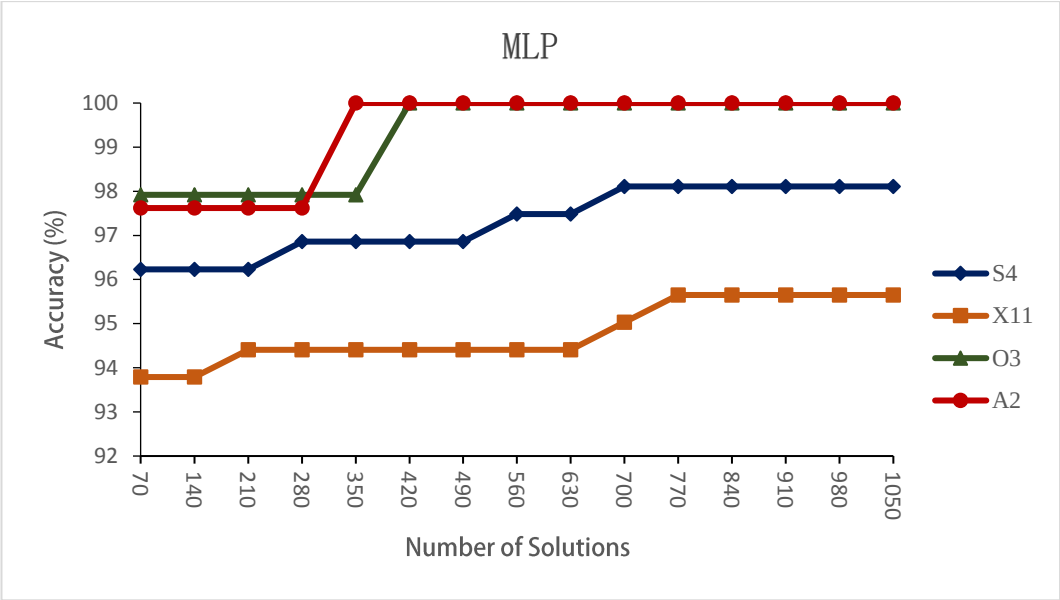

Figure 26. Accuracy of the GA for feature selection for the MLP classifier

Table 20. Training time of different models

Model	Instances	Time (s)			
		Preprocessing and Feature Extraction	Feature Selection	Classification	Total
Full Model	538	57,59	7305,23	71,82	7434,64
	140	15,71	2706,73	45,82	2768,26
Without FS	538	57,59		92,38	149,97
	140	15,71		57,13	72,84
Best Combination (7 classifiers)	538	57,59	6178,67	67,88	6304,14
	140	15,71	2435,71	49,32	2500,74
Best combination (5 classifiers)	538	57,59	4213,76	62,0808	4333,43
	140	15,71	1647,45	45,8618	1709,02
Best combination (3 classifiers)	538	57,59	2636,98	51,1385	2745,71
	140	15,71	981,87	44,0262	1041,61
Combination of classifiers with FS and without FS (5 classifiers)	538	57,59	3780,95	81,583	3920,12
	140	15,71	1634,98	54,028	1704,72
Combination of classifiers with FS and without FS (3 classifiers)	538	57,59	2055,84	84,2317	2197,66
	140	15,71	721,99	54,4293	792,13
Combination of classifiers with FS and without FS (MLP)	538	57,59	1195,02	82,3357	1334,95
	140	15,71	557,6	53,4701	626,78
Reduced GA Solutions	538	57,59	3895,16	73,08	4025,83
	140	15,71	1624,66	44,97	1685,34

The full model, as expected, requires a lot of processing time, a little over two hours for the 538 trials case. On the other hand, the model without feature selection is the fastest, with very low values (around 2.5 minutes). Interestingly, the GA with reduced number of solutions reduced almost to a half the processing time, matching the training time of other approaches, such as the best combination with 7 and 5 classifiers, and the combination of classifiers (some with feature selection) for 5 classifiers. The most successful model in classification accuracy, the combination of classifiers (some with feature selection) for 3 classifiers, presented a good value for the processing cost, with only 1334,95 seconds for 538 instances and 626,78 seconds for 140 instances (22,3 and 10.45 minutes, respectively).

5. Case Study II: Orthosis for hand grasping neurorehabilitation

The second case study is a real application of the proposed RMIPE model on a BCI with neurorehabilitation purposes. The signal is acquired through the Emotiv EPOC+ headset [154] and the output device is an orthosis designed for hand grasping assistance.

5.1 Orthosis design

The conceptual design of the orthosis was developed considering the lack of actual commercial dynamic orthosis with low cost and a modern design and operation. Another important topic considered is the lightweight that the orthosis should have. Some of the commercial orthotic systems available today are shown in Table 21. Figure 27 shows the illustration of the orthotic systems of Table 21.

Table 21. Commercial orthotic devices

Device	Price ($)	Actuator
Becker Talon [235]	354.92	No
Bunnell Splint for MP and wrist extension [236]	266.66	No
Saebo Flex [237]	1240.99	No
Saebo Glove [238]	312.78	No
Jaeco PowerGrip [239]	1500	Linear Actuator
Gloreha Pro 2 Professional System [240]	17000	Hydraulic System

Figure 27. Example of commercial orthosis. Upper left corner: Becker talon, Upper center: Bunnell Splint, Upper right corner: Saebo Flex, Lower left corner: Saebo Glove, Lower center: Jaeco "PowerGrip", and Lower right corner: Gloreha Pro 2

The final concept of the proposed device is described in Figure 28. This approach is based on a 3D printed device that combines the low weight characteristic with the mechanical components that can offer the necessary movement and a modern design, providing also a low fabrication cost.

Figure 28. Concept of the hand grasping orthosis

The design of the orthosis prototypes was accomplished on the SolidWorks software. The first prototype is only composed of a skeleton, which should be integrated with Velcro fabric to make it fit to the hand, but it is adjustable to any user. The orthosis has two superior pieces: one that assures the coupling of the motor and another that is linked to the motor and generates the movement of the attached fingers to close the hand. The resultant design is shown in Figure 29.

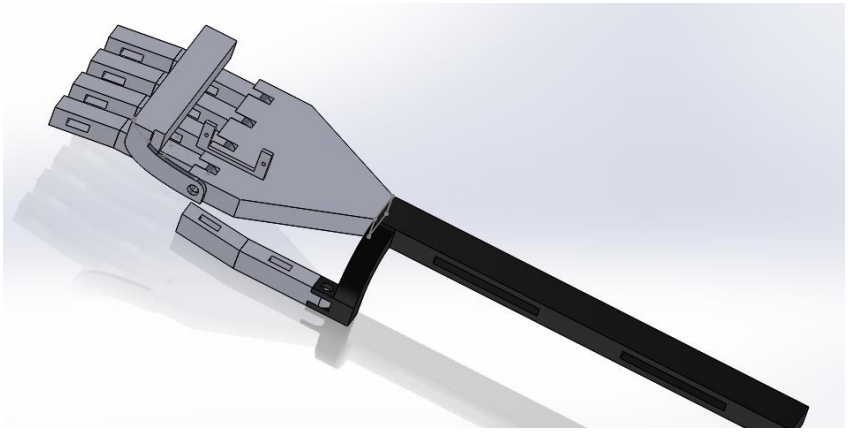

Figure 29. First Prototype of the hand orthosis

This prototype presented some design misconceptions, including the connection to the thumb. This motivated the development of a second prototype, with independent and separated pieces for the fingers, as well as the redesign of the body of the hand. The thumb connection was also improved, offering the correct angle for the position of the finger. The arm piece was substituted by less bulky pieces situated on the sides of the arm, which give support to several motors that augmented the strength of the system. The resultant design can be seen in Figure 30 and the final product in Figure 31.

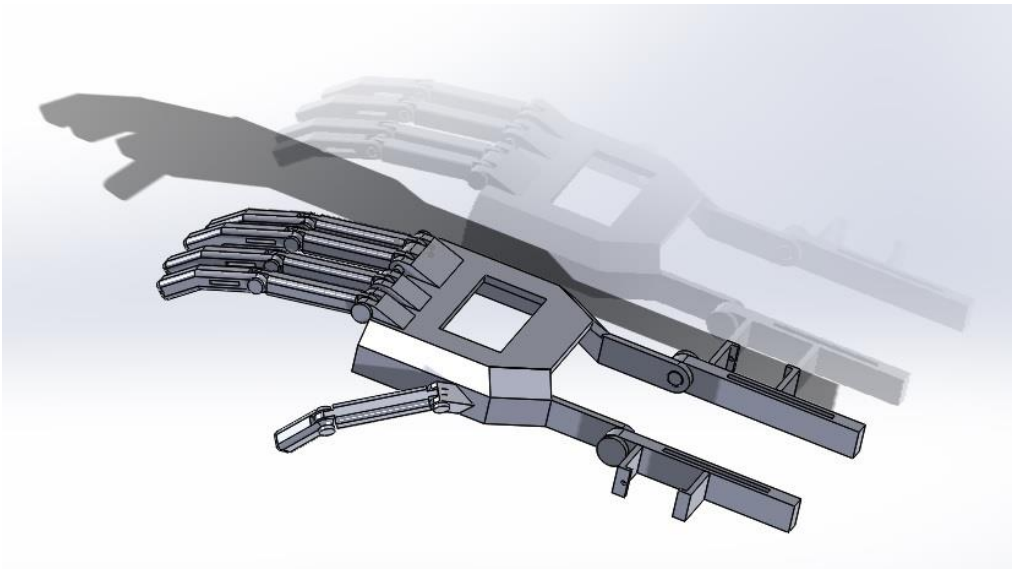

Figure 30. Second orthosis prototype: ExoClaw

Figure 31. ExoClaw printed in 3D

5.2 Emotiv EEG signal and mental tasks

The Emotiv EPOC+ (see Figure 32) is a commercial headset for the acquisition of scientific contextual EEG with applications on BCIs. The wireless headset features 14 channels (See blue circles in Figure 33) through saline based wet sensors, which make the system easier to setup for the consumer use. The sampling rate is of 128Hz and the signal is bandwidth filtered between 0.2 and 43Hz, with Notch filters at 50 and 60Hz. The device includes a software that provides access to the raw EEG, saved to files in binary EEGLAB format. The Emotiv was first introduced with a gaming purpose, but presented a fast expansion to other areas, reaching the research teams as a useful acquisition system with a good number of channels and a much lower cost that the professional ones. Real applications in research are described in [241], [242],[243], [244], [245] and [246].

Figure 32. Emotiv EPOC+ headset

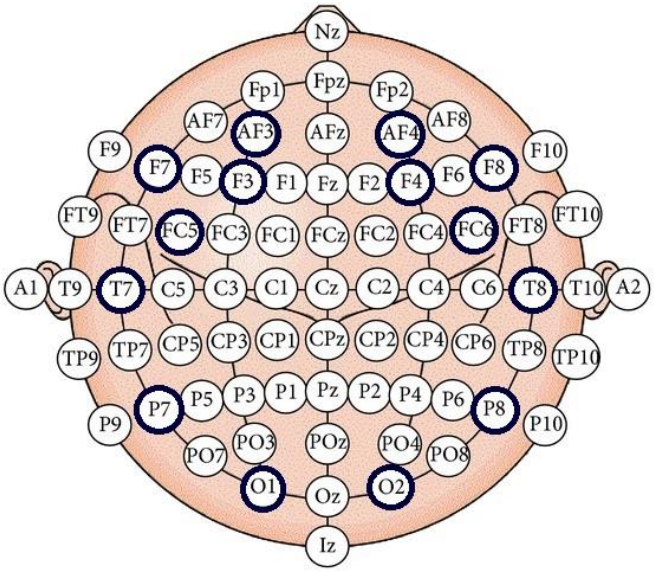

Figure 33. Position of the electrodes of the Emotiv EPOC+ headset

Due to the channel distribution of the headset, in this work the headset had to be moved backward, placing the frontal electrodes (the ones in the AF3, F3, AF4 and F4) in the C3, C4, FC3 and FC4 positions. The acquisition test procedure was designed as the one of the BCI competition IIa, with two seconds of idle status followed by a fixation cross. In time = 3s the audio stimuli is activated as the visual cue (in this case an arrow pointing left or right), indicating in which direction the hand should be moved imagery. The interval between each trial was defined randomly between 7 and 14s. The trials lasted for 6 seconds. Figure 34 illustrate the trial time scheme.

The experiments were conducted in a chair with armrest, with the subjects in a relaxed position in front of the display. The test program was developed in Matlab and features a trial time and number of trials defined by the test conductor. The program saves in a file the timestamps of each start trigger that will be used to delimitate the trial EEG signal read from the file produced by the Emotiv Xavier TestBench software. The only channels included in the research were C3 and C4, the remaining 12 were obviated. For this presented research, a male subject of 27 years old was selected, without any historical of physiological or mental diseases.

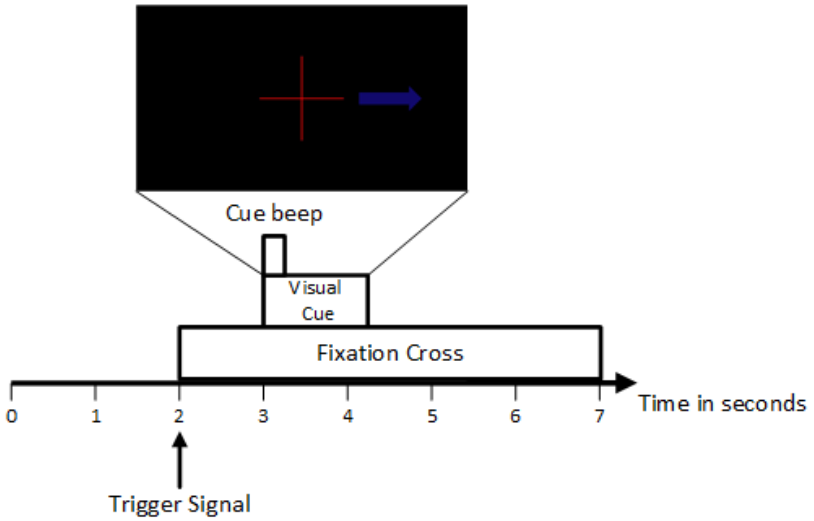

Figure 34. Time scheme of the EEG signal acquisition

5.3 Performance of the system

The dataset was obtained from the selected user (subject X) in three sessions, recorded on the same day, with a temporal separation of several minutes. Each session was composed of 50 trials, resulting in a total of 150 trials. This number of trials was divided in 100 for training and 50 for testing. For this particular dataset, during the preprocessing stage, a different behavior was presented from the datasets employed in the previous Chapter. In this case, the energy distribution of the frequency bands evidenced that the energy levels of the frequency band from 0 to 4Hz surpassed the ones from the others, in a proportion of thousands. This can be caused by the absence on our acquisition and preprocessing system of any type of artifact or noise elimination. As a consequence, for the validation of the proposed model, that frequency band was eliminated of the analysis.

The accuracies obtained in the tests conducted on the RMIPE system are shown in Table 22. The results demonstrate the challenge that represents the classification of the obtained data, with a performance for the second best fusion method tested on the previous chapter of only 65,625%. The low accuracy can be the result of the small number of trials (150 in this case), being insufficient for a good evolution of the GA algorithm. The obtained accuracies validate the importance of the RMIPE model, with a performance for the full model of MLP as meta-classifier of 93.75%, which is a great improvement compared to the result

of the GA for classifier selection. It can be also highlighted that the diversity demonstrated to be fundamental in this real case, presenting lower accuracy values for the models with a lower quantity of classifiers (GA for classifier selection and the best combination of only 3 classifiers).

Table 22. Performances obtained using the Emotiv EPOC+ as acquisition device

	X
GA for classifier selection	65.63
Full Model	93.75
Best Combination (7 classifiers)	84.38
Best combination (5 classifiers)	81.25
Best combination (3 classifiers)	71.88
Combination of classifiers with FS and without FS (5 classifiers)	**87.50**
Combination of classifiers with FS and without FS (3 classifiers)	**87.50**
Combination of classifiers with FS and without FS (MLP)	81.25

The results also validate the capacity of the Emotiv headset to acquire with relatively good quality the modulations of the EEG signal through the MI tasks, presenting great possibilities for its future application on cheaper BCI system. Also, the proposed RMIPE model, even without application of additional muscular or ocular artifact removal algorithms, was capable of obtaining a high accuracy, which is an evidence of its robustness.

6. Conclusions and Future Work

6.1 Conclusions

This Master Dissertation presented, as main objective, the design of a machine learning system that, based on the EEG signal from only C3 and C4 electrodes, can classify MI tasks with high performance, robustness to trial and inter-subject signal variations, and reasonable processing time, for application to a functional and low-cost BCI system for neurorehabilitation.

To accomplish that objective, a Machine Learning system was developed, named RMIPE (Robust Motor Intention Prediction Ensemble). The signal was preprocessed by a Wavelet Packet Decomposition algorithm that provided a time-frequency analysis, dividing the signal into eight principal frequency bands that carried the information. Those frequency sub-bands suffered a feature extraction process, where statistical, power and phase features were calculated in order to properly and diversely represent the EEG signal. The high dimensionality of the feature vector indicated the advantage of applying a feature selection algorithm to eliminate the redundant features. A validation of common feature selection methods was preformed, comparing its results obtained on ten classifiers. As a result, the Genetic Algorithm was selected as the feature selection algorithm for the proposed model.

The classification stage was characterized by the employment of a multiple classifier system, which a priori is supposed to outperform the performance of the individual classifiers, to be immune to the problem of high dimensionality, and to present robustness to the variance of EEG data, both in the time domain and inter-subject (dataset), even in the presence of artifacts. The diversity of the ensemble of classifiers is offered by the application of classifiers with different learning nature, being employed: PNN, SVM with linear kernel, SVM with quadratic kernel, LDA linear, RBF, k-NN with the Euclidean distance, k-NN with the Mahalanobis distance, k-NN with correlation, k-NN with cosine, and MLP. The fusion method then is selected after an assessment of the results of each method on several datasets.

The datasets employed to validate the RMIPE model were obtained from BCIs competitions held in 2003, 2005 and 2008. In total, 6 subjects were used, including two with higher frequency rate, one of them with EOG artifacts. The analysis of the performances resulted in the selection of an MLP as meta-classifier, a method with great performance (mean accuracy of 93.1925% including subjects A2, S4, X11 and O3 from dataset III of BCI competition II and dataset IIIb from BCI competition III) and robustness to the previously explained variations and artifacts (90.9091% of accuracy without any method for artifact removal for subject A01 from dataset IIIa from the BCI competition IV). The superiority of the MLP as meta-classifier was confirmed by the Friedman statistical test, with post-hoc Holm method conducted with the objective to validate the statistical relevance of the obtained results. It was also concluded that a higher frequency sampling offered better results, representing the signal in a more descriptive form that assisted the model classification. In addition, the feature selection stage demonstrated to be of significant importance, providing the subset of features that offers a better classification accuracy.

After the complete model definition, the classification accuracy of the proposed model was compared to some state-of-the-art researches in the literature. The proposed method culminated being more suited for the application, outperforming all others reported results. One of the main characteristics of the model for the obtainment of superior results is the employment of an Ensemble composed of different classifiers, which can diversify the errors of the individual components and in that way to improve the final classification. Other positive characteristic is the chosen fusion method: the MLP is an excellent non-linear method for defining the best relationships between the output of the individual classifiers, which facilitates the differentiation of the classes. Also, the combination of features from a diverse nature and an algorithm for selecting the ones that offered a better representation of the data for each classifier can be declared as of great importance for the improvement of the classification accuracy. A conclusion that can be obtained from the works in the literature is that the application of an ensemble of classifiers to EEG signal classification for BCI systems presented a high classification accuracy, which makes them superior to a single classifier method.

The proposed RMIPE model presented a high computational cost, which motivated the proposal of two approaches for the optimization of the ensemble. Firstly, the outcome of the reduction of the number of classifiers that composed the ensemble was assessed. The obtained results highlighted the importance of the diversity for the model, which led to the proposal of the other approach: maintain the number of classifiers but varying the ones that employed feature selection. As a result, the model with a higher accuracy was obtained for only three classifiers with feature selection: the MLP, the SVM with quadratic kernel and the k-NN with cosine metric. In addition, an analysis was performed of the optimal number of solutions for each individual classifier, in order to reduce its training time. Moreover, it was also illustrated the training time of some of the proposed models.

Finally, a real application of the proposed model was proposed. The designed neurorehabilitation BCI system was developed based on 3D modeled orthosis for the rehabilitation of the hand grasping function. The RMIPE was applied to a EEG dataset obtained with a commercial EEG headset of the Emotiv Company, the EPOC+. The acquisition experiment was developed and the orthosis was designed, being printed on a 3D printer. The final application presented great results, demonstrating the usefulness and robustness of our approach to a BCI system.

6.2 Future Works

The following possibilities of future works can be mentioned:

- Extension of the model for multiclass classification, including feet movement, which can be of assistance for other types of rehabilitation devices;

- Optimization of the processing time mainly in the training phase, which can be done through the application of a Quantum-inspired GA instead of a standard GA;

- Final development of a fully functional orthosis glove more faithful to the required concepts, in order to generate a final product;

- Extension of the experiments to other subjects and a higher number of trials;

- Assessment of the effect of the training process with trials of various sessions dilated in time, and also trials from other subjects on the proposed model;

- Analysis of the RMIPE model in online testing;

- Definition and execution of trials in clinics with the proposed neurorehabilitation system with the objective of acknowledging its short and long term benefits.

References

[1] C. G. Gross, “Early History of Neuroscience,” in *Encyclopedia of Neuroscience*, 1987, pp. 843–847.

[2] “The BRAIN Initiative | The White House.” [Online]. Available: https://www.whitehouse.gov/BRAIN. [Accessed: 10-Dec-2016].

[3] “The Human Brain Project - Human Brain Project.” [Online]. Available: https://www.humanbrainproject.eu/. [Accessed: 10-Dec-2016].

[4] B. Graimann, B. Allison, and G. Pfurtscheller, “Brain–Computer Interfaces: A Gentle Introduction,” in *Brain-Computer Interfaces*, Springer Berlin Heidelberg, 2009, pp. 1–27.

[5] V. Gandhi, G. Prasad, D. Coyle, L. Behera, and T. M. McGinnity, “EEG-Based Mobile Robot Control Through an Adaptive Brain-Robot Interface,” *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 44, no. 9, pp. 1278–1285, Sep. 2014.

[6] T. M. Rutkowski, H. Mori, T. Kodama, and H. Shinoda, “Airborne Ultrasonic Tactile Display Brain-computer Interface -- A Small Robotic Arm Online Control Study,” in *10th AEARU Workshop on Computer Science and Web Technology*, 2015.

[7] B. B. Longo, A. B. Benevides, J. Castillo, and T. Bastos-Filho, “Using Brain-Computer Interface to control an avatar in a Virtual Reality Environment,” in *5th ISSNIP-IEEE Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC)*, 2014, pp. 1–4.

[8] K. K. Ang and C. Guan, “Brain-Computer Interface for Neurorehabilitation of Upper Limb After Stroke,” in *Proceedings of the IEEE*, 2015, vol. 103, no. 6, pp. 944–953.

[9] R. Rupp, S. C. Kleih, R. Leeb, J. del R. Millan, A. Kübler, and G. R. Müller-Putz, “Brain--Computer Interfaces and Assistive Technology,” in *Brain-Computer-Interfaces in their ethical, social and cultural contexts*, G. Grübler and E. Hildt, Eds. Dordrecht: Springer Netherlands, 2014, pp. 7–38.

[10] S. C. Wriessnegger, D. Hackhofer, and G. R. Muller-Putz, “Classification of unconscious like/dislike decisions: First results towards a novel application for BCI technology,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 2331–2334.

[11] H. Gürkök and A. Nijholt, “Affective brain-computer interfaces for arts,” *Proc. - 2013 Hum. Assoc. Conf. Affect. Comput. Intell. Interact. ACII 2013*, pp. 827–831, 2013.

[12] M. A. Lebedev and M. A. L. Nicolelis, “Brain-machine interfaces: past, present and future,” *Trends Neurosci.*, vol. 29, no. 9, pp. 536–546, 2006.

[13] G. Schalk and J. Mellinger, “Brain Sensors and Signals,” in *A Practical Guide to Brain--Computer Interfacing with BCI2000: General-Purpose Software for Brain--Computer Interface Research, Data Acquisition, Stimulus Presentation, and Brain Monitoring*, London: Springer London, 2010, pp. 9–35.

[14] J. del R. Millán *et al.*, “Non-Invasive Brain-Machine Interaction,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 22, no. 5, pp. 959–972, 2008.

[15] G. Repovš, “Dealing with Noise in EEG Recording and Data Analysis,” *Inform. Medica Slov.*, vol. 15, no. 1, pp. 18–25, 2010.

[16] J. A. Urigüen and B. Garcia-Zapirain, “EEG artifact removal-state-of-the-art and guidelines.,” *J. Neural Eng.*, vol. 12, no. 3, p. 31001, 2015.

[17] B. Burle, L. Spieser, C. Roger, L. Casini, T. Hasbroucq, and F. Vidal, “Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view,” *Int. J. Psychophysiol.*, vol. 97, no. 3, pp. 210–220, 2015.

[18] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review,” *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[19] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, “EEG-based discrimination between imagination of right and left hand movement,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 6, pp. 642–651, 1997.

[20] I. Rejer and K. Lorenz, “Genetic algorithm and forward method for feature selection in EEG feature space,” *J. Theor. Appl. Comput. Sci.*, vol. 7, no. 2, pp. 72–82, 2013.

[21] M. R. Hasan, M. I. Ibrahimy, S. M. A. Motakabber, and S. Shahid, "Classification of Multichannel EEG Signal by Linear Discriminant Analysis," in *Progress in Systems Engineering: Proceedings of the Twenty-Third International Conference on Systems Engineering*, H. Selvaraj, D. Zydek, and G. Chmaj, Eds. Cham: Springer International Publishing, 2015, pp. 279–282.

[22] H. Zeng and A. Song, "Optimizing Single-Trial EEG Classification by Stationary Matrix Logistic Regression in Brain-Computer Interface," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 27, no. 11, pp. 2301–2313, 2016.

[23] I. H. Robertson and J. M. Murre, "Rehabilitation of brain damage: brain plasticity and principles of guided recovery.," *Psychol. Bull.*, vol. 125, no. 5, pp. 544–75, 1999.

[24] L. E. H. van Dokkum, T. Ward, and I. Laffont, "Brain computer interfaces for neurorehabilitation-its current status as a rehabilitation strategy post-stroke," *Ann. Phys. Rehabil. Med.*, vol. 58, no. 1, pp. 3–8, 2015.

[25] "Stroke | World Heart Federation." [Online]. Available: http://www.world-heart-federation.org/cardiovascular-health/stroke/. [Accessed: 11-Dec-2016].

[26] J. Xu, A. M. Haith, and J. W. Krakauer, "Motor Control of the Hand Before and After Stroke," in *Clinical Systems Neuroscience*, Tokyo: Springer Japan, 2015, pp. 271–289.

[27] J. S. Knutson, M. Y. Harley, T. Z. Hisel, and J. Chae, "Improving hand function in stroke survivors: a pilot study of contralaterally controlled functional electric stimulation in chronic hemiplegia.," *Arch. Phys. Med. Rehabil.*, vol. 88, no. 4, pp. 513–20, Apr. 2007.

[28] T. Hanakawa, "Organizing motor imageries," *Neurosci. Res.*, vol. 104, pp. 56–63, 2016.

[29] H. Yuan and B. He, "Brain-Computer Interfaces Using Sensorimotor Rhythms: Current State and Future Perspectives," *IEEE Trans Biomed Eng.*, vol. 61, no. 5, pp. 1425–1435, 2015.

[30] S. Lemm, C. Schäfer, and G. Curio, "Aggregating classification accuracy across time: Application to single trial EEG," *Adv. Neural Inf. Process. Syst.*, pp. 825–832, 2007.

[31] N. Brodu, F. Lotte, and A. Lécuyer, "Exploring Two Novel Features for

EEG-based Brain-Computer Interfaces: Multifractal Cumulants and Predictive Complexity," pp. 1–9, 2010.

[32] H. Bashashati, R. K. Ward, G. E. Birch, and A. Bashashati, "Comparing different classifiers in sensory motor brain computer interfaces," *PLoS One*, vol. 10, no. 6, pp. 1–17, 2015.

[33] D. W. Abbott, "Combining models to improve classifier accuracy and robustness," in *Proceedings of Second International Conference on Information Fusion*, 1999, pp. 289–295.

[34] A. Ahangi, M. Karamnejad, N. Mohammadi, R. Ebrahimpour, and N. Bagheri, "Multiple classifier system for EEG signal classification with application to brain-computer interfaces," *Neural Comput. Appl.*, vol. 23, no. 5, pp. 1319–1327, 2013.

[35] R. Sampanna and S. Mitaim, "Noise benefits in motor imagery classification using ensemble support vector machine," in *2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*, 2014, pp. 53–56.

[36] H. Bashashati, R. K. Ward, and A. Bashashati, "User-customized brain computer interfaces using Bayesian optimization," *J. Neural Eng.*, vol. 13, no. 2, p. 26001, 2016.

[37] J. Malmivuo and R. Plonsey, *Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 1995.

[38] D. Purves *et al.*, *Neuroscience*, 3rd ed. Massachusetts: Sinauer Associates, Inc, 2004.

[39] M. Teplan, "Fundamentals of EEG measurement," *Meas. Sci. Rev.*, vol. 2, no. 2, pp. 1–11, 2002.

[40] H. H. Jasper, "The ten-twenty electrode system of the International Federation," *Electroencephalogr. Clin. Neurophysiol.*, 1958.

[41] G. V. Kondraske, "Neurophysiological measurements," in *Biomedical Engineering and Instrumentation*, J. D. Bronzino, Ed. Boston: PWS Publishing, 1986, pp. 138–179.

[42] N. V. Thakor and S. Tong, "Advances in quantitative electroencephalogram analysis methods," *Annu. Rev. Biomed. Eng.*, vol. 6, pp. 453–495, 2004.

[43] D. Regan, *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*. New York: Elsevier, 1989.

[44] G. Bin, X. Gao, Y. Wang, B. Hong, and S. Gao, "VEP-based brain-computer interfaces: Time, frequency, and code modulations," *IEEE Comput. Intell. Mag.*, vol. 4, no. 4, pp. 22–26, 2009.

[45] U. Strehl, "Slow Cortical Potentials Neurofeedback," *J. Neurother.*, vol. 13, no. 2, pp. 117–126, 2009.

[46] N. Birbaumer, T. Elbert, a G. Canavan, and B. Rockstroh, "Slow potentials of the cerebral cortex and behavior," *Physiol. Rev.*, vol. 70, no. 164, pp. 1–41, 1990.

[47] H. Gevensleben *et al.*, "Neurofeedback of slow cortical potentials: neural mechanisms and feasibility of a placebo-controlled design in healthy adults," *Front. Hum. Neurosci.*, vol. 8, no. December, p. 990, 2014.

[48] T. Hinterberger *et al.*, "Brain-Computer Communication and Slow Cortical Potentials," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1011–1018, 2004.

[49] T. W. Picton, "The P300 wave of the human event-related potential.," *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, vol. 9, no. 4. pp. 456–479, 1992.

[50] U. Hoffmann, J. M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 115–125, 2008.

[51] J. Polich, P. C. Ellerson, and J. Cohen, "P300, stimulus intensity, modality, and probability," *Int. J. Psychophysiol.*, vol. 23, no. 1985, pp. 55–62, 1996.

[52] M. Chang *et al.*, "Comparison of P300 responses in auditory, visual and audiovisual spatial speller BCI paradigms," *Proc. 5th Int. Brain-Computer Interface Meet.*, p. 156, 2013.

[53] G. Pfurtscheller *et al.*, "Graz-Brain-Computer Interface: State of Research," in *Toward Brain-Computer Interfacing*, G. Dornhege, J. del R. Millán, T. Hinterberger, D. J. McFarland, and K.-R. Müller, Eds. Cambridge, Massachusetts: The MIT Press, 2007, p. 507.

[54] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain- computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[55] M. Jeannerod, "The representing brain: Neural correlates of motor intention

and imagery," *Behav. Brain Sci.*, vol. 17, no. 2, p. 187, 1994.

[56] A. Kübler and K.-R. Müller, "An Introduction to Brain-Computer Interfacing," in *Toward Brain-Computer Interfacing*, The MIT Press, 2007, pp. 1–25.

[57] P. S. Hammon and V. R. De Sa, "Preprocessing and meta-classification for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 3, pp. 518–525, 2007.

[58] T. Al-ani and D. Trad, "Signal Processing and Classification Approaches for Brain-Computer Interface," in *Intelligent and Biosensors*, V. S. Somerset, Ed. InTech, 2010, p. 386.

[59] V. Gandhi, *Brain-Computer Interfacing for Assistive Robotics*, 1st ed. Academic Press, 2014.

[60] V. J. Samar, A. Bopardikar, R. Rao, and K. Swartz, "Wavelet Analysis of Neuroelectric Waveforms: A Conceptual Tutorial," *Brain Lang.*, vol. 66, no. 1, pp. 7–60, 1999.

[61] M. Akay, "Wavelets in biomedical engineering," *Ann. Biomed. Eng.*, vol. 23, no. 5, pp. 531–542, Sep. 1995.

[62] W. Ting, Y. Guo-zheng, Y. Bang-hua, and S. Hong, "EEG feature extraction based on wavelet packet decomposition for brain computer interface," *Measurement*, vol. 41, no. 6, pp. 618–625, 2008.

[63] Y. Zhang, B. Liu, X. Ji, and D. Huang, "Classification of EEG Signals Based on Autoregressive Model and Wavelet Packet Decomposition," *Neural Process. Lett.*, pp. 1–14, Jun. 2016.

[64] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[65] F. Ebrahimi, M. Mikaeili, E. Estrada, and H. Nazeran, "Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients.," in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2008.

[66] M. Y. Gokhale and D. K. Khanduja, "Time Domain Signal Analysis Using Wavelet Packet Decomposition Approach," *Int. J. Commun. Netw. Syst.*

Sci., vol. 3, no. 3, pp. 321–329, 2010.

[67] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction Foundations and Applications*. Berlin: Spinger, 2006.

[68] A. Subasi and M. Ismail Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8659–8666, 2010.

[69] R. K. Chaurasiya, N. D. Londhe, and S. Ghosh, "Statistical Wavelet Features, PCA, and SVM Based Approach for EEG Signals Classification," *World Acad. Sci. Eng. Technol. Int. J. Electr. Comput. Energ. Electron. Commun. Eng.*, vol. 9, no. 2, pp. 182–186, 2015.

[70] M. H. Alomari, E. A. Awada, A. Samaha, and K. Alkamha, "Wavelet-Based Feature Extraction for the Analysis of EEG Signals Associated with Imagined Fists and Feet Movements," *Comput. Inf. Sci.*, vol. 7, no. 2, p. 17, Mar. 2014.

[71] P. A. Kharat and S. V. Dudul, "Daubechies wavelet neural network classifier for the diagnosis of epilepsy," *WSEAS Trans. Biol. Biomed.*, vol. 9, no. 4, pp. 103–113, 2012.

[72] V. Abootalebi, M. H. Moradi, and M. A. Khalilzadeh, "A new approach for EEG feature extraction in P300-based lie detection," *Comput. Methods Programs Biomed.*, vol. 94, no. 1, pp. 48–57, 2009.

[73] G. Lisi, T. Noda, and J. Morimoto, "Decoding the ERD/ERS: influence of afferent input induced by a leg assistive robot.," *Front. Syst. Neurosci.*, vol. 8, no. May, p. 85, 2014.

[74] H. U. Amin *et al.*, "Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques," *Australas. Phys. Eng. Sci. Med.*, vol. 38, no. 1, pp. 139–149, Mar. 2015.

[75] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Hum. Brain Mapp.*, vol. 8, no. 4, pp. 194–208, 1999.

[76] D. J. Krusienski, D. J. McFarland, and J. R. Wolpaw, "Value of amplitude, phase, and coherence features for a sensorimotor rhythm-based brain-computer interface.," *Brain Res. Bull.*, vol. 87, no. 1, pp. 130–4, Jan. 2012.

[77] W.-Y. Hsu, "Enhancing the Performance of Motor Imagery EEG Classification Using Phase Features," *Clin. EEG Neurosci.*, vol. 46, no. 2,

pp. 113–118, Apr. 2015.

[78] P. Tass *et al.*, “Detection of n: m phase locking from noisy data: application to magnetoencephalography,” *Phys. Rev. Lett.*, vol. 81, no. 15, p. 3291, 1998.

[79] V. Kumar and S. Minz, “Feature Selection: A literature Review,” *Smart Comput. Rev.*, vol. 4, no. 3, 2014.

[80] H. Liu, H. Motoda, R. Setiono, and Z. Zhao, “Feature Selection : An Ever Evolving Frontier in Data Mining,” *J. Mach. Learn. Res. Work. Conf. Proc. 10 Fourth Work. Featur. Sel. Data Min.*, pp. 4–13, 2010.

[81] K. R. Müller, M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz, “Machine learning techniques for brain-computer interfaces,” *Biomed. Tech*, vol. 49, no. 1, pp. 11–22, 2004.

[82] I. Koprinska, “Feature Selection for Brain-Computer Interfaces,” in *New Frontiers in Applied Data Mining*, Springer Berlin Heidelberg, 2010, pp. 106–117.

[83] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[84] G. C. Cawley, N. L. Talbot, and M. Girolami, “Sparse multinomial logistic regression via bayesian l1 regularisation,” *Adv. Neural Inf. Process. Syst.*, vol. 19, 2007.

[85] M. A. Hall, “Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning,” in *ICML ’00 Proceedings of the Seventeenth International Conference on Machine Learning*, 2000, pp. 359–366.

[86] I. Kononenko, “Estimating attributes: Analysis and extensions of RELIEF,” in *Machine Learning: ECML-94: European Conference on Machine Learning Catania, Italy, April 6--8, 1994 Proceedings*, F. Bergadano and L. De Raedt, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 171–182.

[87] C. Ding and H. Peng, “Minimum redundancy feature selection from microarray gene expression data,” *J. Bioinform. Comput. Biol.*, vol. 3, no. 2, pp. 185–205, 2005.

[88] H. Liu and R. Setiono, “A probabilistic approach to feature selection a filter solution,” in *International conference on machine learning - ICML*, 1996, vol. 96, pp. 319–327.

[89] D. Koller and M. Sahami, “Toward optimal feature selection,” in *Proc. of*

the Thirteenth International Conference on Machine Learning, 1996, pp. 284–292.

[90] M. Mitchell, *An introduction to genetic algorithms*, 5th ed. Cambridge, Massachusetts: A Bradford Book The MIT Press, 1996.

[91] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intell.*, vol. 1, no. 1, pp. 33–57, 2007.

[92] R. N. Khushaba, A. Al-Ani, A. AlSukker, and A. Al-Jumaily, "A Combined Ant Colony and Differential Evolution Feature Selection Algorithm," in *Ant Colony Optimization and Swarm Intelligence: 6th International Conference, ANTS 2008, Brussels, Belgium, September 22-24, 2008. Proceedings*, M. Dorigo, M. Birattari, C. Blum, M. Clerc, T. Stützle, and A. F. T. Winfield, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1–12.

[93] I. Rejer, "EEG Feature Selection for BCI Based on Motor Imaginary Task," *Found. Comput. Decis. Sci.*, vol. 37, no. 4, pp. 283–292, 2012.

[94] J. Pearl, *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. Boston, MA: Addison-Wesley Longman Publishing Co., Inc., 1984.

[95] M. A. Hall and L. A. Smith, "Feature subset selection: A correlation based filter approach," *Prog. Connect. Inf. Syst. Vols 1 2*, pp. 855–858, 1998.

[96] G. John, R. Kohavi, and K. Pfleger, "Irrelevant Features and the Subset Selection Problem," *Mach. Learn. Proc. Elev. Int. Conf.*, pp. 121–129, 1994.

[97] K. Kira and L. Rendell, "A practical approach to feature selection," in *Proceedings of the ninth international workshop on Machine learning*, 1992, pp. 249–256.

[98] B. Schölkopf and A. J. Smola, *Learning with kernels : support vector machines, regularization, optimization, and beyond*. MIT Press, 2002.

[99] S. Mika, G. Ratsch, J. Weston, and B. Scholkopf, "Fisher discriminant analysis with kernels," *Neural Networks Signal Process. IX, 1999. Proc. 1999 IEEE Signal Process. Soc. Work.*, pp. 41–48, 1999.

[100] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[101] R. Rojas, *Neural networks : a systematic introduction*. Springer-Verlag,

1996.

[102] D. Kriesel, *Neural Networks*. 2013.

[103] A. Subasi and E. Erçelebi, "Classification of EEG signals using neural network and logistic regression," *Comput. Methods Programs Biomed.*, vol. 78, no. 2, pp. 87–99, 2005.

[104] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, 2007.

[105] T. Wu, B. Yang, and H. Sun, "EEG Classification Based on Artificial Neural Network in Brain Computer Interface," in *Life System Modeling and Intelligent Computing: International Conference on Life System Modeling and Simulation, LSMS 2010, and International Conference on Intelligent Computing for Sustainable Energy and Environment, ICSEE 2010, Wuxi, China, September 17*, K. Li, X. Li, S. Ma, and G. W. Irwin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 154–162.

[106] R. Manor and A. B. Geva, "Convolutional Neural Network for Multi-Category Rapid Serial Visual Presentation BCI," *Front. Comput. Neurosci.*, vol. 9, no. December, p. 146, 2015.

[107] T. G. Dietterich, "Ensemble Methods in Machine Learning," in *Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21-23, 2000 Proceedings*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 1–15.

[108] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[109] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, 1990.

[110] L. I. Kuncheva, *Combining pattern classifiers : methods and algorithms*. Jonh Wiley & Sons, 2004.

[111] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[112] L. Breiman, "Bagging Predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.

[113] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of

On-Line Learning and an Application to Boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

[114] Tin Kam Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, 1998.

[115] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms: Second Edition*, 2nd ed. Jonh Wiley & Sons, 2014.

[116] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining Multi-label Data," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Boston, MA: Springer US, 2010, pp. 667–685.

[117] L. I. Kuncheva and C. J. Whitaker, "Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy," *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, 2003.

[118] W. H. E. Day, "Consensus methods as tools for data analysis," in *Classification and Related Methods for Data Analysis*, H. H. Bock, Ed. Elsevier Science Publishers B.V., 1988, pp. 317–324.

[119] S. A. Goldman and M. K. Warmuth, "Learning binary relations using weighted majority voting," *Mach. Learn.*, vol. 20, no. 3, pp. 245–271, 1995.

[120] Y. S. Huang and C. Y. Suen, "A method of combining multiple experts for the recognition of unconstrained handwritten numerals," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 1, pp. 90–94, 1995.

[121] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, "Soft combination of neural classifiers: A comparative study," *Pattern Recognit. Lett.*, vol. 20, no. 4, pp. 429–444, 1999.

[122] M. Grabisch, "On equivalence classes of fuzzy connectives-the case of fuzzy integrals," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 1, pp. 96–109, 1995.

[123] L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, vol. 33, no. 1, pp. 1–39, 2010.

[124] J. E. Huggins *et al.*, "Workshops of the Sixth International Brain-Computer Interface Meeting: brain-computer interfaces past, present, and future," *Brain-Computer Interfaces*, vol. 2621, no. May, pp. 1–34, 2017.

[125] M. Lebedev, "Brain-machine interfaces: an overview," *Transl. Neurosci.*, vol. 5, no. 1, pp. 99–110, 2014.

[126] D. J. McFarland and T. M. Vaughan, “BCI in practice,” *Prog. Brain Res.*, vol. 228, pp. 389–404, 2016.

[127] N. Birbaumer *et al.*, “A spelling device for the paralysed,” *Nature*, vol. 398, no. 6725, pp. 297–298, 1999.

[128] J. Williamson, R. Murray-Smith, B. Blankertz, M. Krauledat, and K.-R. Müller, “Designing for uncertain, asymmetric control: Interaction design for brain–computer interfaces,” *Int. J. Hum. Comput. Stud.*, vol. 67, no. 10, pp. 827–841, 2009.

[129] M. Nakanishi, Y. Wang, Y.-T. Wang, Y. Mitsukura, and T.-P. Jung, “A HIGH-SPEED BRAIN SPELLER USING STEADY-STATE VISUAL EVOKED POTENTIALS,” *Int. J. Neural Syst.*, vol. 24, no. 6, p. 1450019, Sep. 2014.

[130] B. O. Mainsah *et al.*, “Increasing BCI communication rates with dynamic stopping towards more practical use: an ALS study,” *J. Neural Eng.*, vol. 12, no. 1, p. 16013, Feb. 2015.

[131] M. Bensch *et al.*, “Nessi: an EEG-controlled web browser for severely paralyzed patients.,” *Comput. Intell. Neurosci.*, vol. 2007, p. 71863, 2007.

[132] E. M. Mugler, C. A. Ruf, S. Halder, M. Bensch, and A. Kubler, “Design and Implementation of a P300-Based Brain-Computer Interface for Controlling an Internet Browser,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 6, pp. 599–609, Dec. 2010.

[133] F. Cincotti *et al.*, “Non-invasive brain–computer interface system: Towards its application as assistive technology,” *Brain Res. Bull.*, vol. 75, no. 6, pp. 796–803, Apr. 2008.

[134] C. H. Ho *et al.*, “Functional electrical stimulation and spinal cord injury,” *Phys. Med. Rehabil. Clin. N. Am.*, vol. 25, no. 3, p. 631–54, ix, Aug. 2014.

[135] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp, “‘Thought’ - Control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia,” *Neurosci. Lett.*, vol. 351, no. 1, pp. 33–36, 2003.

[136] M. Tavella, R. Leeb, R. Rupp, and J. D. R. Millán, “Towards natural non-invasive hand neuroprostheses for daily living,” *2010 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC’10*, pp. 126–129, 2010.

[137] S. Bermúdez i Badia, A. García Morgade, H. Samaha, and P. F. M. J.

Verschure, “Using a Hybrid Brain Computer Interface and Virtual Reality System to Monitor and Promote Cortical Reorganization through Motor Activity and Motor Imagery Training,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 174–181, Mar. 2013.

[138] A. Venkatakrishnan, G. E. Francisco, and J. L. Contreras-Vidal, “Applications of Brain-Machine Interface Systems in Stroke Recovery and Rehabilitation.,” *Curr. Phys. Med. Rehabil. reports*, vol. 2, no. 2, pp. 93–105, 2014.

[139] L. M. Alonso-Valerdi, R. A. Salido-Ruiz, and R. A. Ramirez-Mendoza, “Motor imagery based brain-computer interfaces: An emerging technology to rehabilitate motor deficits,” *Neuropsychologia*, vol. 79, pp. 354–363, 2015.

[140] G. R. Müller-Putz and G. Pfurtscheller, “Control of an electrical prosthesis with an SSVEP-based BCI,” *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 361–364, 2008.

[141] J. del R. Millán, F. Galán, D. Vanhooydonck, E. Lew, J. Philips, and M. Nuttin, “Asynchronous non-invasive brain-actuated control of an intelligent wheelchair,” in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 3361–3364.

[142] C. J. Bell, P. Shenoy, R. Chalodhorn, and R. P. N. Rao, “Control of a humanoid robot by a noninvasive brain–computer interface in humans,” *J. Neural Eng.*, vol. 5, no. 2, pp. 214–220, Jun. 2008.

[143] K. LaFleur, K. Cassady, A. Doud, K. Shades, E. Rogin, and B. He, “Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface,” *J. Neural Eng.*, vol. 10, no. 4, p. 46003, Aug. 2013.

[144] M. Middendorf, G. McMillan, G. Calhoun, and K. S. Jones, “Brain-computer interfaces based on the steady-state visual-evoked response.,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 211–4, Jun. 2000.

[145] M. Tangermann *et al.*, “Playing pinball with non-invasive BCI,” in *22nd Annual Conference on Neural Information Processing Systems*, 2008.

[146] A. Nijholt, “BCI for Games: A `State of the Art’ Survey,” in *Entertainment Computing - ICEC 2008: 7th International Conference, Pittsburgh, PA, USA, September 25-27, 2008. Proceedings*, S. M. Stevens and S. J.

Saldamarco, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 225–228.

[147] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain–Computer Communication: Motivation, Aim, and Impact of Exploring a Virtual Apartment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 473–482, Dec. 2007.

[148] A. Lecuyer, F. Lotte, R. B. Reilly, R. Leeb, M. Hirose, and M. Slater, "Brain-Computer Interfaces, Virtual Reality, and Videogames," *Computer (Long. Beach. Calif).*, vol. 41, no. 10, pp. 66–72, Oct. 2008.

[149] H. Touyama, "Photo Data Retrieval via P300 Evoked Potentials," *IEICE Trans. Inf. Syst.*, vol. E91–D, no. 8, pp. 2212–2213, Aug. 2008.

[150] E. R. Miranda, "Brain-computer music interface for composition and performance," *Int. J. Disabil. Hum. Dev.*, vol. 5, no. 2, pp. 119–126, Jan. 2006.

[151] A. Kübler, A. Furdea, S. Halder, and A. Hösle, "Brain painting - BCI meets art," in *Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course*, 2008.

[152] B. Zhang, Jianjun Wang, and T. Fuhlbrigge, "A review of the commercial brain-computer interface technology from perspective of industrial robotics," in *2010 IEEE International Conference on Automation and Logistics*, 2010, pp. 379–384.

[153] G. Udovicic, A. Topic, and M. Russo, "Wearable technologies for smart environments: A review with emphasis on BCI," in *2016 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 2016, pp. 1–9.

[154] "EMOTIV - Brainwear® Wireless EEG Technology." [Online]. Available: https://www.emotiv.com/. [Accessed: 06-Feb-2017].

[155] "EEG - ECG - Biosensors." [Online]. Available: http://neurosky.com/. [Accessed: 06-Feb-2017].

[156] "Products / ENOBIO - Neuroelectrics." [Online]. Available: http://www.neuroelectrics.com/products/enobio/. [Accessed: 06-Feb-2017].

[157] "Cognionics, Inc." [Online]. Available: http://www.cognionics.com/. [Accessed: 06-Feb-2017].

[158] "OpenBCI - Open Source Biosensing Tools (EEG, EMG, EKG, and

more).” [Online]. Available: http://openbci.com/. [Accessed: 06-Feb-2017].

[159] “BrainLink by Macrotellect | Healthy Brainwaves for Everyone.” [Online]. Available: http://o.macrotellect.com/#v2. [Accessed: 06-Feb-2017].

[160] S. Hanslmayr, P. Sauseng, M. Doppelmayr, M. Schabus, and W. Klimesch, “Increasing Individual Upper Alpha Power by Neurofeedback Improves Cognitive Performance in Human Subjects,” *Appl. Psychophysiol. Biofeedback*, vol. 30, no. 1, pp. 1–10, 2005.

[161] R. Sitaram, S. Lee, S. Ruiz, M. Rana, R. Veit, and N. Birbaumer, “Real-time support vector classification and feedback of multiple emotional brain states,” *Neuroimage*, vol. 56, no. 2, pp. 753–765, May 2011.

[162] V. Tulceanu, “Comprehensive brainwave authentication using emotional stimuli,” in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, 2012.

[163] M. A. Lopez-Gordo, R. Ron-Angevin, and F. Pelayo, “Authentication of Brain-Computer Interface Users in Network Applications,” in *Advances in Computational Intelligence: 13th International Work-Conference on Artificial Neural Networks, IWANN 2015, Palma de Mallorca, Spain, June 10-12, 2015. Proceedings, Part I*, I. Rojas, G. Joya, and A. Catala, Eds. Cham: Springer International Publishing, 2015, pp. 124–132.

[164] A. Bahramisharif, M. van Gerven, T. Heskes, and O. Jensen, “Covert attention allows for continuous control of brain-computer interfaces,” *Eur. J. Neurosci.*, vol. 31, no. 8, pp. 1501–1508, Apr. 2010.

[165] C. T. Lin, C. J. Chang, B. S. Lin, S. H. Hung, C. F. Chao, and I. J. Wang, “A real-time wireless brain-computer interface system for drowsiness detection,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 4, no. 4, pp. 214–222, 2010.

[166] P. Brunner, L. Bianchi, C. Guger, F. Cincotti, and G. Schalk, “Current trends in hardware and software for brain-computer interfaces (BCIs),” *J. Neural Eng.*, vol. 8, no. 2, p. 25001, 2011.

[167] J. Walker and G. P. Kozlowski, “Neurofeedback treatment of epilepsy,” *Child Adolesc. Psychiatr. Clin. N. Am.*, vol. 14, no. 1, pp. 163–176, Jan. 2005.

[168] U. Strehl, U. Leins, G. Goth, C. Klinger, T. Hinterberger, and N. Birbaumer, “Self-regulation of Slow Cortical Potentials: A New Treatment

for Children With Attention-Deficit/Hyperactivity Disorder," *Pediatrics*, vol. 118, no. 5, pp. e1530–e1540, Nov. 2006.

[169] F. Schneider *et al.*, "Self-regulation of slow cortical potentials in psychiatric patients: Schizophrenia," *Biofeedback Self. Regul.*, vol. 17, no. 4, pp. 277–292, 1992.

[170] F. Schneider, H. Heimann, R. Mattes, W. Lutzenberger, and N. Birbaumer, "Self-regulation of slow cortical potentials in psychiatric patients: Depression," *Biofeedback Self. Regul.*, vol. 17, no. 3, pp. 203–214, 1992.

[171] T. Ambler, A. Ioannides, and S. Rose, "Brands on the Brain: Neuro-Images of Advertising," *Bus. Strateg. Rev.*, vol. 11, no. 3, pp. 17–30, Sep. 2000.

[172] D. Ariely and G. S. Berns, "Neuromarketing: the hope and hype of neuroimaging in business.," *Nat. Rev. Neurosci.*, vol. 11, no. 4, pp. 284–92, Apr. 2010.

[173] "Consumer Neuroscience | Nielsen." [Online]. Available: http://www.nielsen.com/us/en/solutions/capabilities/consumer-neuroscience.html. [Accessed: 06-Feb-2017].

[174] "Neuro-Insight." [Online]. Available: http://www.neuro-insight.com/. [Accessed: 06-Feb-2017].

[175] "Forebrain :: Reaching thoughts." [Online]. Available: http://www.forebrain.com.br/en/index.php. [Accessed: 06-Feb-2017].

[176] B.-G. Xu and A.-G. Song, "Pattern Recognition of Motor Imagery EEG using Wavelet Transform," *J. Biomed. Sci. Eng.*, vol. 1, pp. 64–67, 2008.

[177] A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains.," *ISRN Neurosci.*, vol. 2014, 2014.

[178] Y. Zhang, Y. Zhang, J. Wang, and X. Zheng, "Comparison of classification methods on EEG signals based on wavelet packet decomposition," *Neural Comput. Appl.*, vol. 26, no. 5, pp. 1217–1225, 2015.

[179] Ö. Aydemir and T. Kayikçioglu, "Investigation of the most appropriate mother wavelet for characterizing imaginary EEG signals used in BCI systems," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 24, no. 1, pp. 38–49, 2016.

[180] K. P. Thomas, C. Guan, L. C. Tong, and A. P. Vinod, "A Study on the impact of spectral variability in brain-computer interface," in *Proceedings*

of 2010 IEEE International Symposium on Circuits and Systems, 2010, pp. 1189–1192.

[181] K. K. Ang *et al.*, "A Large Clinical Study on the Ability of Stroke Patients to Use an EEG-Based Motor Imagery Brain-Computer Interface," *Clin. EEG Neurosci.*, vol. 42, no. 4, pp. 253–258, 2011.

[182] W. Yi *et al.*, "EEG oscillatory patterns and classification of sequential compound limb motor imagery," *J. Neuroeng. Rehabil.*, vol. 13, no. 1, p. 11, 2016.

[183] L. Kauhanen *et al.*, "EEG and MEG Brain–Computer Interface for Tetraplegic Patients," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 190–193, Jun. 2006.

[184] A. Vuckovic and F. Sepulveda, "Delta band contribution in cue based single trial classification of real and imaginary wrist movements," *Med. {&} Biol. Eng. {&} Comput.*, vol. 46, no. 6, pp. 529–539, 2008.

[185] D. R. Achanccaray and M. A. Meggiolaro, "Brain Computer Interface based on Electroencephalographic Signal Processing," in *XVI IEEE International Congress of Electrical, Electronic and Systems Engineering - INTERCON 2009*, 2009.

[186] A. O. G. Barbosa, D. R. A. Diaz, M. M. B. R. Vellasco, M. A. Meggiolaro, and R. Tanscheit, "Mental Tasks Classification for a Noninvasive BCI Application," in *Artificial Neural Networks -- ICANN 2009: 19th International Conference, Limassol, Cyprus, September 14-17, 2009, Proceedings, Part II*, C. Alippi, M. Polycarpou, C. Panayiotou, and G. Ellinas, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 495–504.

[187] S.-H. Lee and J. S. Lim, "Minimum feature selection for epileptic seizure classification using wavelet-based feature extraction and a fuzzy neural network," *Appl. Math. Inf. Sci.*, vol. 8, no. 3, pp. 1295–1300, 2014.

[188] L. Vega-Escobar, A. E. Castro-Ospina, and L. Duque-Muñoz, "Feature extraction schemes for BCI systems," in *2015 20th Symposium on Signal Processing, Images and Computer Vision, STSIVA 2015 - Conference Proceedings*, 2015, no. 76.

[189] W.-Y. Hsu, "Wavelet-Coherence Feaures for Motor Imagery EEG Analysis Posterior to EOG Noise Elimination," *Int. J. Innov. Comput. Inf. Control*,

vol. 9, no. 2, pp. 465–475, 2013.

[190] J. del R. Millán, M. Franzé, J. Mouriño, F. Cincotti, and F. Babiloni, "Relevant EEG features for the classification of spontaneous motor-related tasks," *Biol. Cybern.*, vol. 86, no. 2, pp. 89–95, 2002.

[191] R. Aler, I. M. Galván, and J. M. Valls, "Applying evolution strategies to preprocessing EEG signals for brain–computer interfaces," *Inf. Sci. (Ny).*, vol. 215, pp. 53–66, 2012.

[192] I. Rejer, "Genetic Algorithms in EEG Feature Selection for the Classification of Movements of the Left and Right Hand," in *Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013*, R. Burduk, K. Jackowski, M. Kurzynski, M. Wozniak, and A. Zolnierek, Eds. Heidelberg: Springer International Publishing, 2013, pp. 579–589.

[193] A. Gupta, R. K. Agrawal, and B. Kaur, "Performance enhancement of mental task classification using EEG signal: a study of multivariate feature selection methods," *Soft Comput.*, vol. 19, no. 10, pp. 2799–2812, 2015.

[194] A. Celecia, R. González, and M. Vellasco, "Feature Selection Methods Applied to Motor Imagery Task Classification," in *LA-CCI 2016 Latin American Conference on Computational Intelligence*, 2016.

[195] C. S. Sanoj and S. Chitra, "An Ensembles Framework for Brain Computer Interface," *J. Theor. Appl. Inf. Technol.*, vol. 69, no. 1, pp. 127–135, 2014.

[196] B. Xu *et al.*, "Enhanced performance by time-frequency-phase feature for EEG-based BCI systems," *Sci. World J.*, 2014.

[197] L. Duan, H. Ge, W. Ma, and J. Miao, "EEG feature selection method based on decision tree," *Biomed. Mater. Eng.*, vol. 26, Aug. 2015.

[198] A. Schlögl, "Dataset IIIb: Non-stationary 2-class BCI data." BCI Competition III, 2005.

[199] M. K. Hazrati and A. Erfanian, "An on-line BCI for control of hand grasp sequence and holding using adaptive probabilistic neural network," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 1009–1012.

[200] J. Atkinson and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," *Expert Syst. Appl.*, vol. 47, pp. 35–41, 2016.

[201] J.-S. Woo, K.-R. Muller, and S.-W. Lee, “Classifying directions in continuous arm movement from EEG signals,” in *The 3rd International Winter Conference on Brain-Computer Interface*, 2015, pp. 1–2.

[202] C. Liu, H. Wang, and Z. Lu, “EEG classification for multiclass motor imagery BCI,” in *2013 25th Chinese Control and Decision Conference (CCDC)*, 2013, pp. 4450–4453.

[203] J. Machado, A. Balbinot, and A. Schuck, “A study of the Naive Bayes classifier for analyzing imaginary movement EEG signals using the Periodogram as spectral estimator,” in *2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*, 2013, pp. 1–4.

[204] I. Sturm, S. Lapuschkin, W. Samek, and K.-R. Müller, “Interpretable deep neural networks for single-trial EEG classification,” 2016.

[205] A. Soria-Frisch, “A Critical Review on the Usage of Ensembles for BCI,” in *Towards Practical Brain-Computer Interfaces: Bridging the Gap from Research to Real-World Applications*, B. Z. Allison, S. Dunne, R. Leeb, J. Del R. Millán, and A. Nijholt, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 41–65.

[206] S. Bhattacharyya, A. Konar, and D. N. Tibarewala, “Motor imagery, P300 and error-related EEG-based robot arm movement control for rehabilitation purpose,” *Med. Biol. Eng. Comput.*, vol. 52, no. 12, pp. 1007–1017, 2014.

[207] J. Ortega, J. Asensio-Cubero, J. Q. Gan, and A. Ortiz, “Classification of motor imagery tasks for BCI with multiresolution analysis and multiobjective feature selection,” *Biomed. Eng. Online*, vol. 15, no. 1, p. 73, 2016.

[208] O. AlZoubi, I. Koprinska, and R. A. Calvo, “Classification of brain-computer interface data,” in *Proceedings of the 7th Australasian Data Mining Conference*, 2008.

[209] S. R. Liyanage, C. Guan, H. Zhang, K. K. Ang, J. Xu, and T. H. Lee, “Dynamically weighted ensemble classification for non-stationary EEG processing,” *J. Neural Eng.*, vol. 10, no. 3, Jun. 2013.

[210] L. Gao, W. Cheng, J. Zhang, and J. Wang, “EEG classification for motor imagery and resting state in BCI applications using multi-class Adaboost extreme learning machine,” *Rev. Sci. Instrum.*, vol. 87, no. 8, Aug. 2016.

[211] A. Schlögl, “Data set: BCI-experiment.” BCI competition 2003, 2002.

[212] A. Schlögl, “Dataset IIIa : 4-class EEG data,” *BCI competition III*. 2005.

[213] C. Brunner, R. Leeb, G. R. Müller-Putz, A. Schlögl, and J. Pfurtscheller, “BCI Competition 2008 – Graz data set A,” *BCI competition IV*. 2008.

[214] C. Vidaurre, A. Schlögl, R. Cabeza, R. Scherer, and G. Pfurtscheller, “A Fully On-Line Adaptive BCI,” *IEEE Trans Biomed Eng.*, vol. 53, no. 6, pp. 1214–1219, 2006.

[215] O. Burmeister, M. Reischl, and R. Mikut, “BCI-Competition 2003, Data sets IIIb,” *BCI competition III*. 2003.

[216] X. Pei and G. Bin, “BCI competiton 2005: Data set IIIb,” *BCI competition III*. 2005.

[217] S. Parini, L. Piccini, L. Maggi, G. Panfili, and G. Andreoni, “BCI Competition III Challenge,” *BCI competition III*. 2005.

[218] D. Coyle, G. Prasad, and T. M. McGinnity, “BCI-competition III – the Graz data - dataset IIIb – Algorithm Description,” *BCI competition III*. 2005.

[219] F. Lotte, “The Use of Fuzzy Inference Systems for Classification in EEG-based Brain-Computer Interfaces,” *3rd Int. Brain-Computer Interfaces Work. Train. Course*, 2006.

[220] Y. Wu and Y. Ge, “A novel method for motor imagery EEG adaptive classification based biomimetic pattern recognition,” *Neurocomputing*, vol. 116, pp. 280–290, 2013.

[221] T. Omar, Z. Wassim, and B. M. Mohamed, “Brain-computer interface: Frequency domain approach using the linear and the quadratic discriminant analysis,” in *2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2014, pp. 346–349.

[222] R. Djemili, H. Bourouba, and M. C. Amara Korb, “Comparative Analysis of Spectral Estimation Methods for Brain-Computer Interfaces,” *2nd Int. Conf. Adv. Eng. Sci. Appl. Math. May 4-5, 2014 Istanbul*, pp. 2–6, 2014.

[223] M. Chen, Y. Fang, and X. Zheng, “Phase space reconstruction for improving the classification of single trial EEG,” *Biomed. Signal Process. Control*, vol. 11, no. 1, pp. 10–16, 2014.

[224] S. Lin, S. Guo, and Z. Huang, “Determining AR order for BCI based on motor imagery,” in *2015 8th International Conference on Biomedical*

Engineering and Informatics (BMEI), 2015, pp. 174–178.

[225] P. Tan, W. Sa, and L. Yu, "Applying Extreme Learning Machine to classification of EEG BCI," in *2016 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2016, pp. 228–232.

[226] P. Tan, G. Tan, Z. Cai, W. Sa, and Y. Zou, "Using ELM-based weighted probabilistic model in the classification of synchronous EEG BCI," *Med. Biol. Eng. Comput.*, vol. 55, no. 1, pp. 33–43, 2017.

[227] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures: Second Edition*. Boca Raton, Florida: Chapman & Hall, 2000.

[228] J. Demšar, "Statistical Comparisons of Classifiers over Multiple Data Sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.

[229] S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure," *Scand. J. Stat.*, vol. 6, no. 2, pp. 65–70, 1979.

[230] L. Martin, R. Leblanc, and N. K. Toan, "Tables for the Friedman rank test," *Can. J. Stat.*, vol. 21, no. 1, pp. 39–43, 1993.

[231] C. Garbin, "Friedman's Two-way Analysis of Variance by Ranks - Analysis of k-Within-Group Data with a Quantitative Response Variable," *Bivariate Statistics Hand-Computation Cache for Psych350 Course*. University of Nebraska-Lincoln.

[232] S. García, A. Fernández, A. D. Benítez, and F. Herrera, "Statistical Comparisons by Means of Non-Parametric Tests: A Case Study on Genetic Based Machine Learning," *Proc. II Congr. Español Informática (CEDI 2007). V Taller Nac. Minería Datos y Aprendiz. (TAMIDA 2007)*, pp. 95--104, 2007.

[233] S. García, A. Fernández, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," *Inf. Sci. (Ny).*, vol. 180, no. 10, pp. 2044–2064, 2010.

[234] N. Settouti, M. E. A. Bechar, and M. A. Chikh, "Statistical Comparisons of the Top 10 Algorithms in Data Mining for Classication Task," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 4, no. 1, p. 46, 2016.

[235] "Talon." [Online]. Available: http://www.beckerorthopedic.com/Affliated Companies/Oregon/ProductPages/Talon.html. [Accessed: 23-Feb-2017].

[236] “Bunnell® Splints.” [Online]. Available: http://www.bunnellsplints.com/28whfoextension.html. [Accessed: 23-Feb-2017].

[237] “SaeboFlex | Saebo.” [Online]. Available: https://www.saebo.com/saeboflex/. [Accessed: 27-Apr-2017].

[238] “SaeboGlove | Saebo.” [Online]. Available: https://www.saebo.com/saeboglove/. [Accessed: 23-Feb-2017].

[239] “JAECO Power Driven Flexor Hinge with 5 Position Wrist using ‘PowerGrip’; - Jaeco Orthopedic.” [Online]. Available: http://jaecoorthopedic.com/products/products/JAECO-Power-Driven-Flexor-Hinge-with-5-Position-Wrist-using-%22PowerGrip%22.html. [Accessed: 23-Feb-2017].

[240] “PROFESSIONAL 2 | Gloreha.” [Online]. Available: http://www.gloreha.com/professional/. [Accessed: 23-Feb-2017].

[241] J. Gómez-Gil, I. San-José-González, L. F. Nicolas-Alonso, and S. Alonso-García, “Steering a tractor by means of an EMG-based human-machine interface,” *Sensors*, vol. 11, no. 7, pp. 7110–7126, 2011.

[242] A. Tahmasebzadeh, M. Bahrani, and S. K. Setarehdan, “Development of a robust method for an online P300 Speller Brain Computer Interface,” in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2013, pp. 1070–1075.

[243] Z. Vamvakousis and R. Ramirez, “Towards a Low Cost Mu-Rhythm Based BCI,” in *Proceedings of the Fifth International Brain-Computer Interface Meeting 2013*, 2013.

[244] J. O’Connor, “Real-time Control of a Robot Arm Using an Inexpensive System for Electroencephalography Aided by Artificial Intelligence,” *Comput. Sci. Honor. Pap.*, 2013.

[245] J. Lin and Z. Jiang, “Implementing Remote Presence Using Quadcopter Control by a Non-Invasive BCI Device,” *Comput. Sci. Inf. Technol.*, vol. 3, no. 4, pp. 122–126, 2015.

[246] J. A. Martinez-Leon, J. M. Cano-Izquierdo, and J. Ibarrola, “Are low cost Brain Computer Interface headsets ready for motor imagery applications?,” *Expert Syst. Appl.*, vol. 49, pp. 136–144, 2016.

Appendix 1. Confusion Matrix for each classifier without feature selection

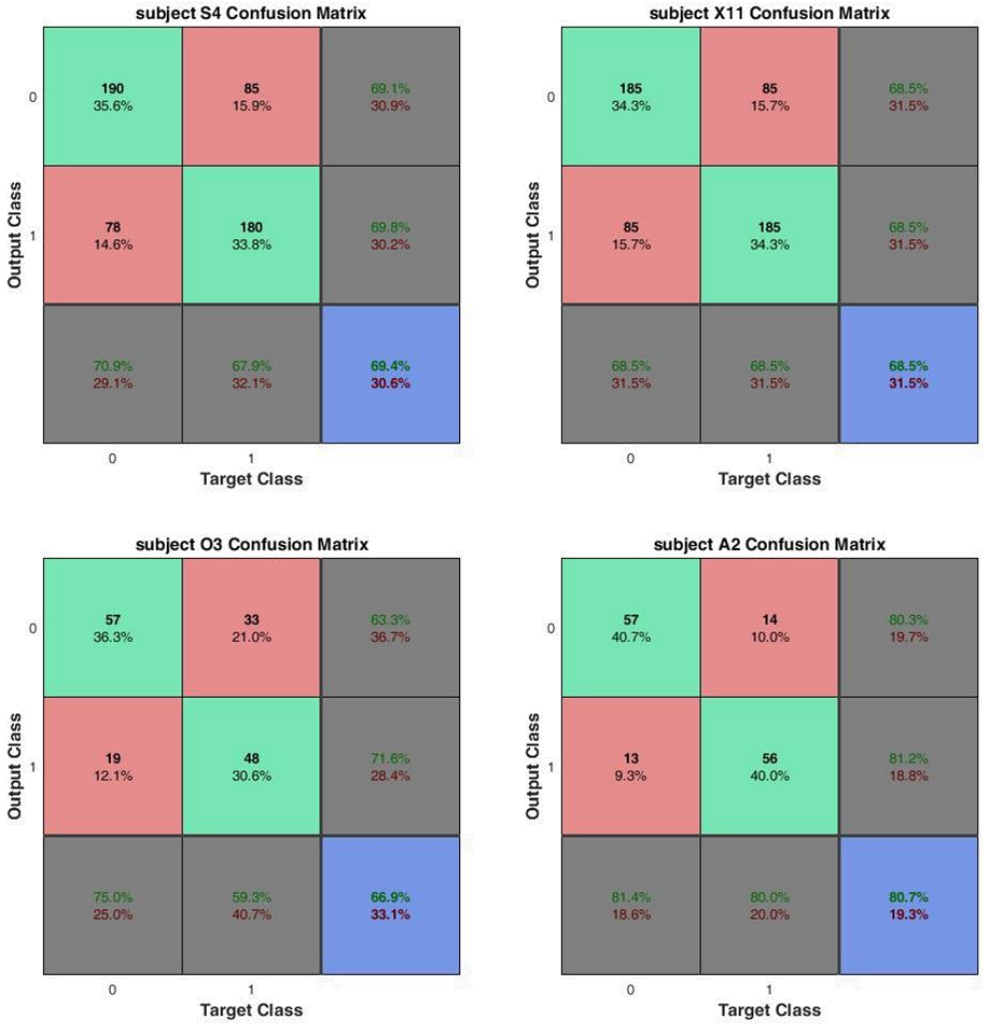

Figure 35. PNN

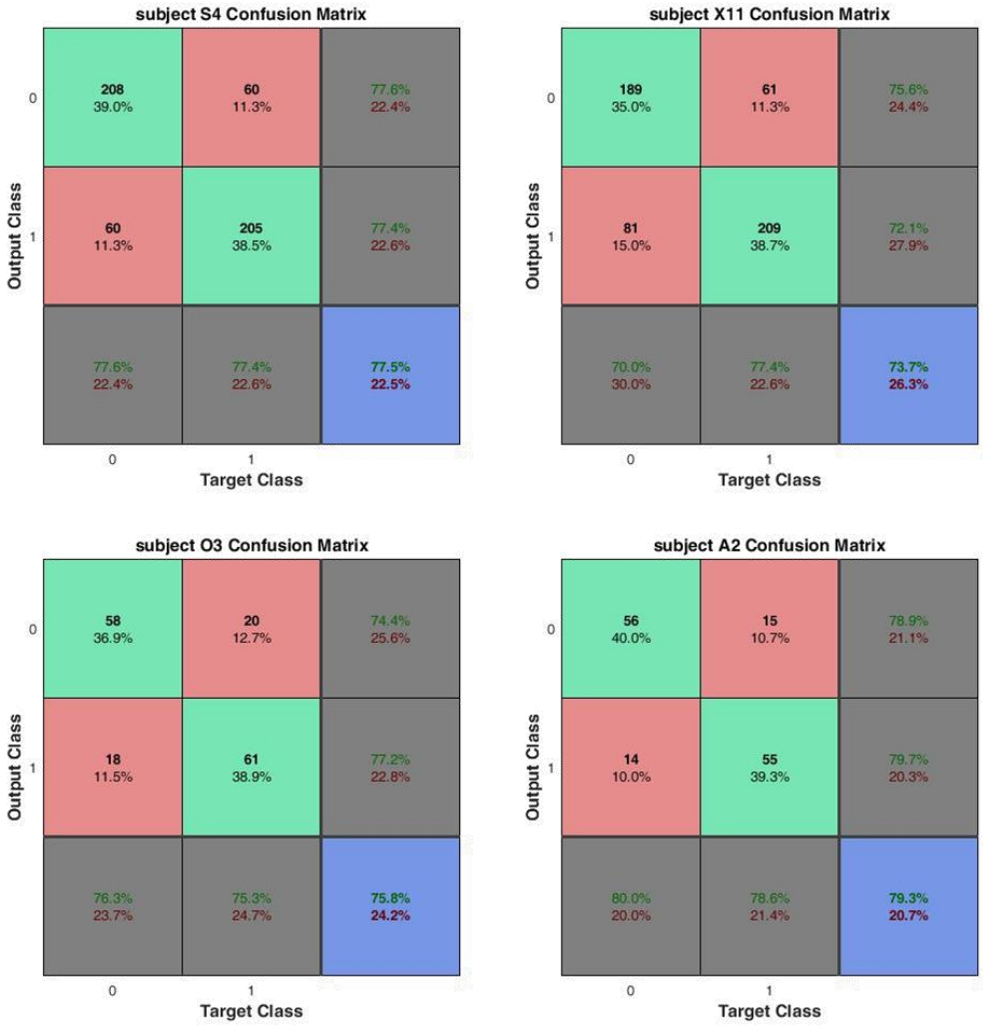

Figure 36. SVM with linear kernel

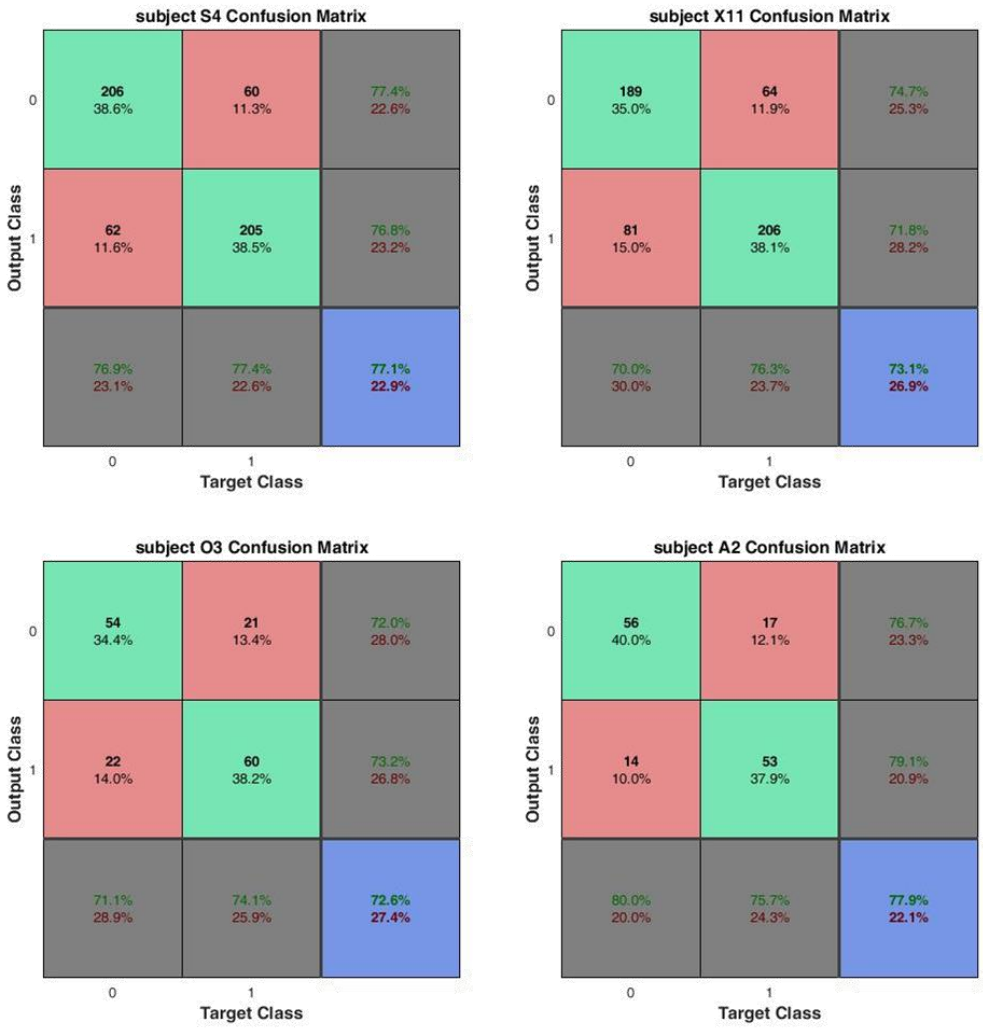

Figure 37. SVM with quadratic kernel

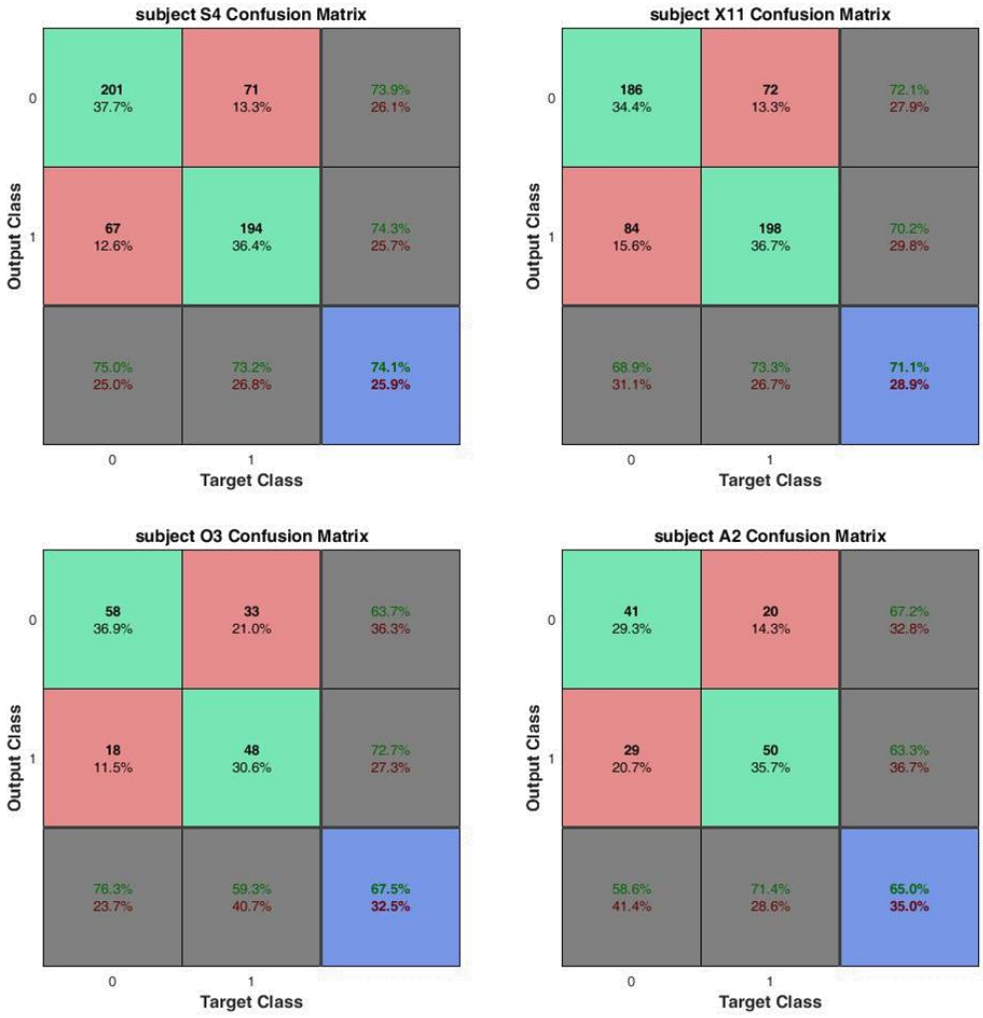

Figure 38. LDA

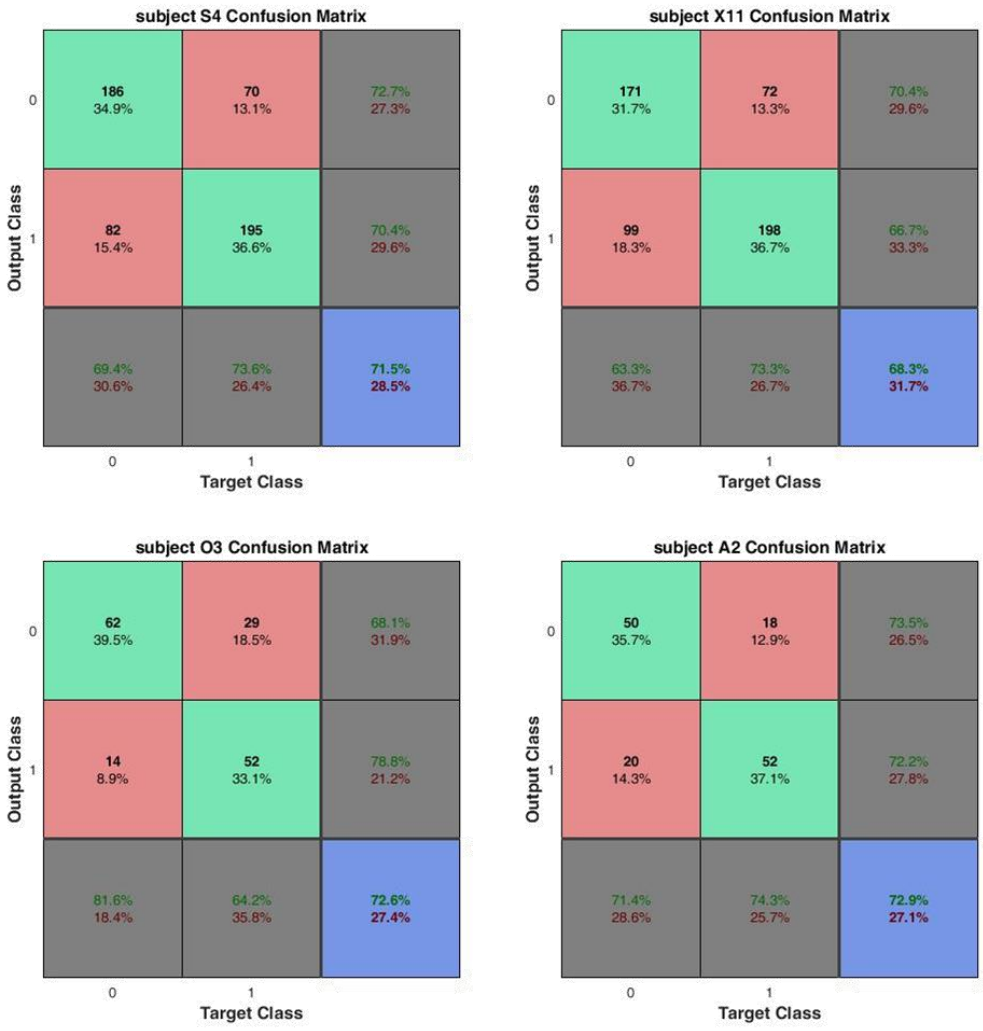

Figure 39. RBF

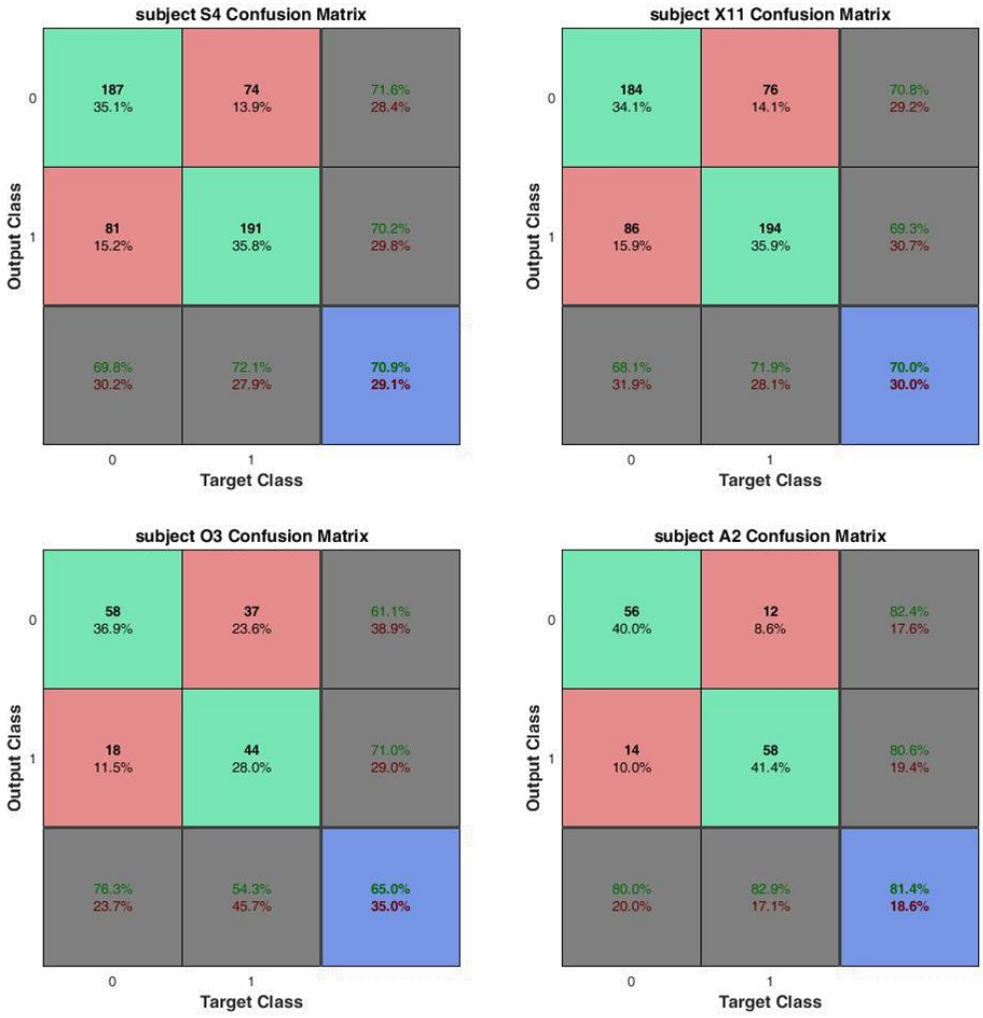

Figure 40. K-NN with Euclidean distance

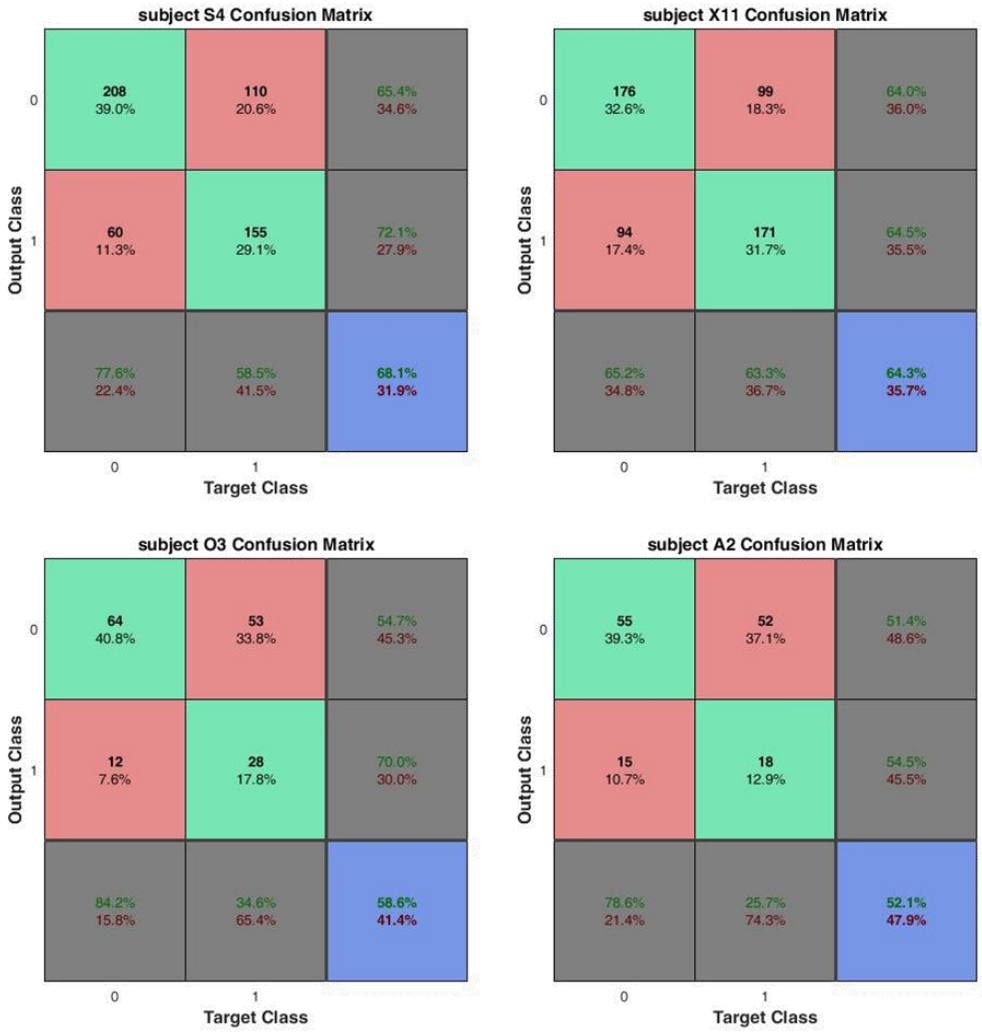

Figure 41. K-NN with Mahalanobis distance

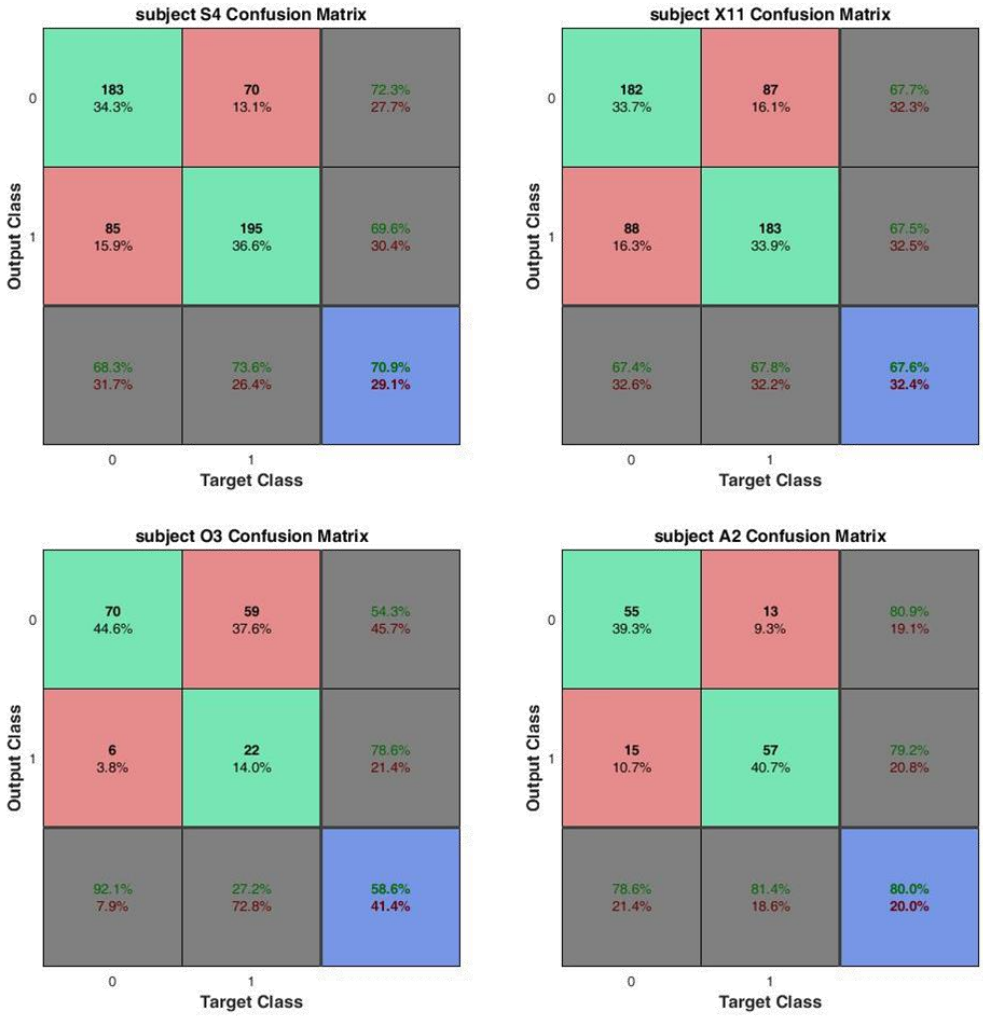

Figure 42. K-NN with cosine

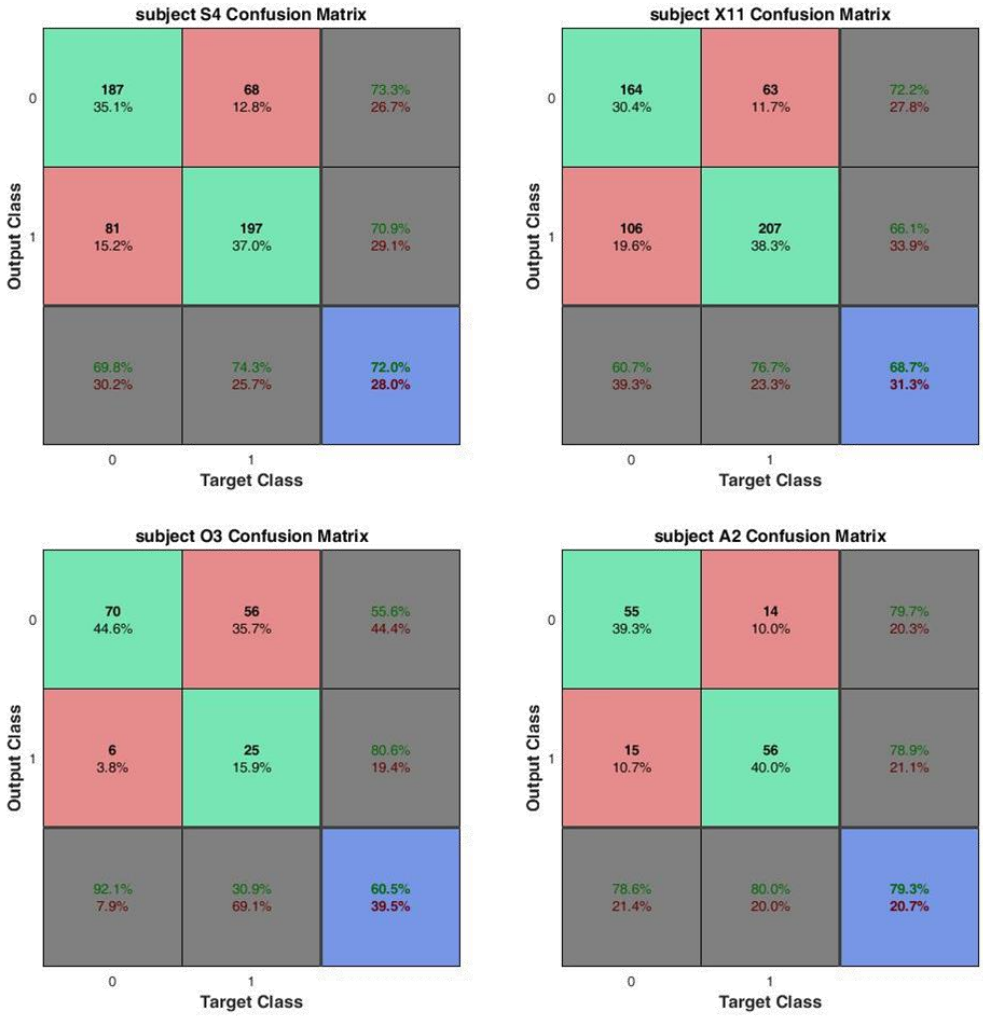

Figure 43. K-NN with correlation

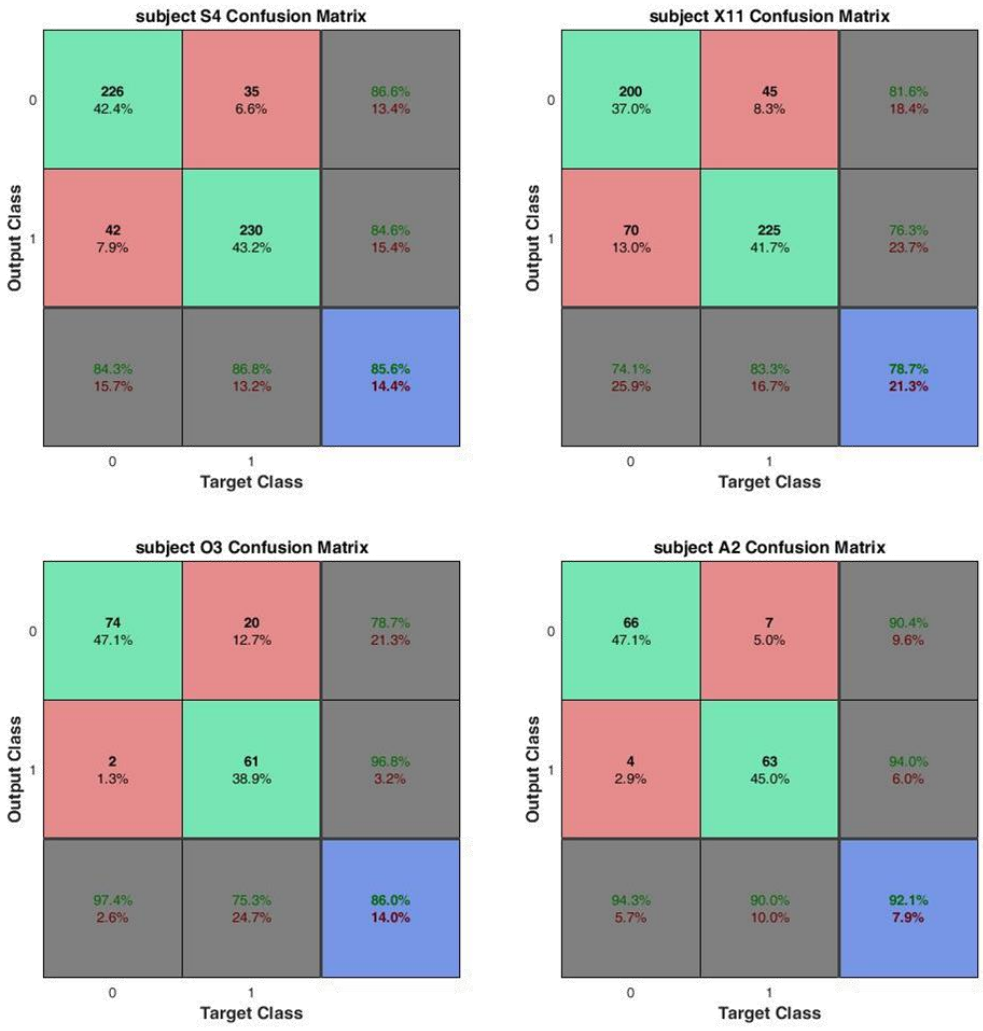

Figure 44. MLP

Appendix 2. Confusion Matrix for each classifier with feature selection

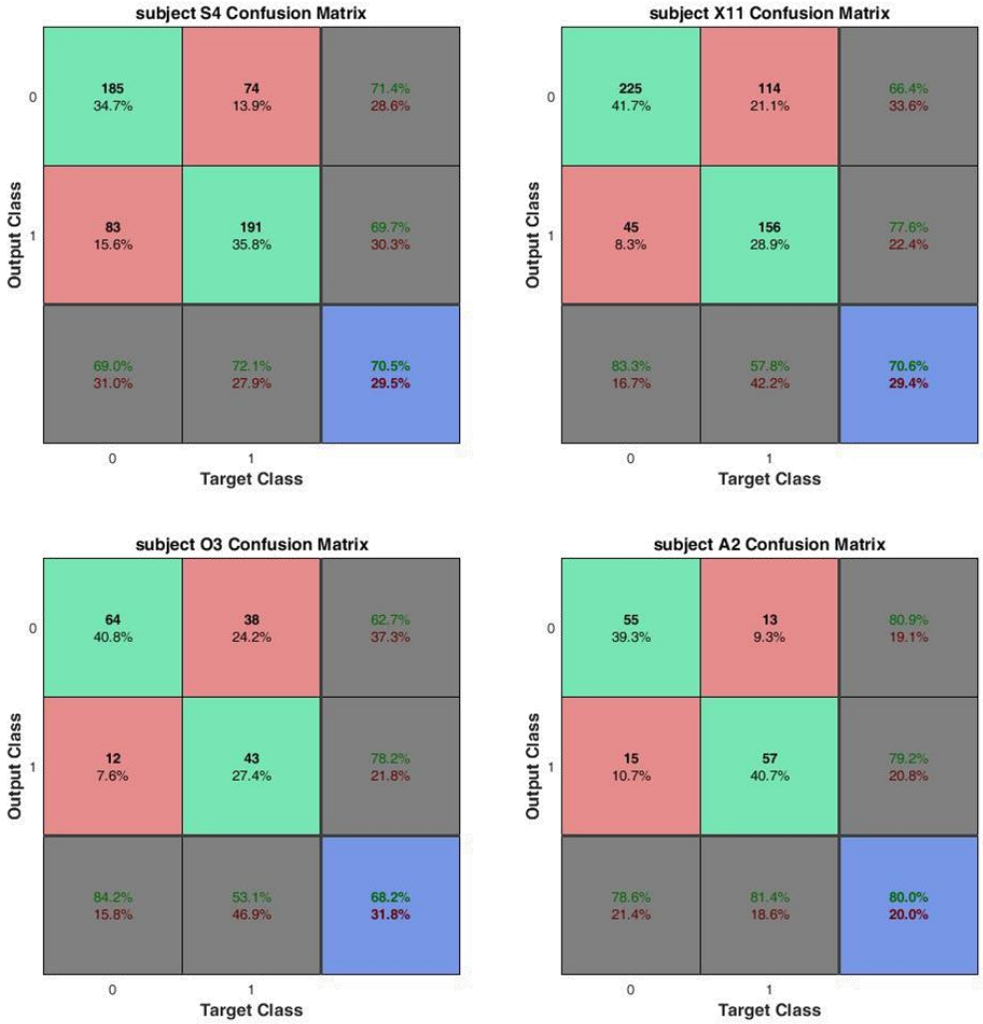

Figure 45. PNN

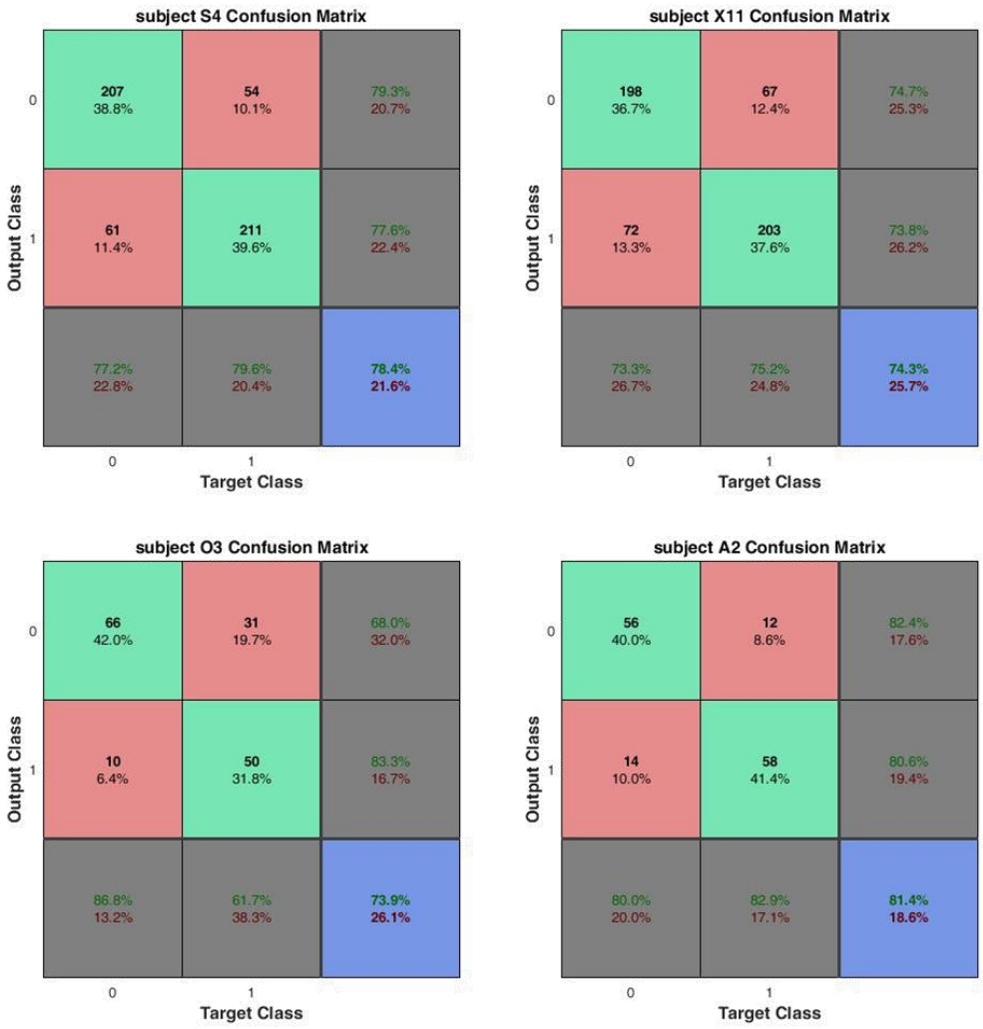

Figure 46. SVM with linear kernel

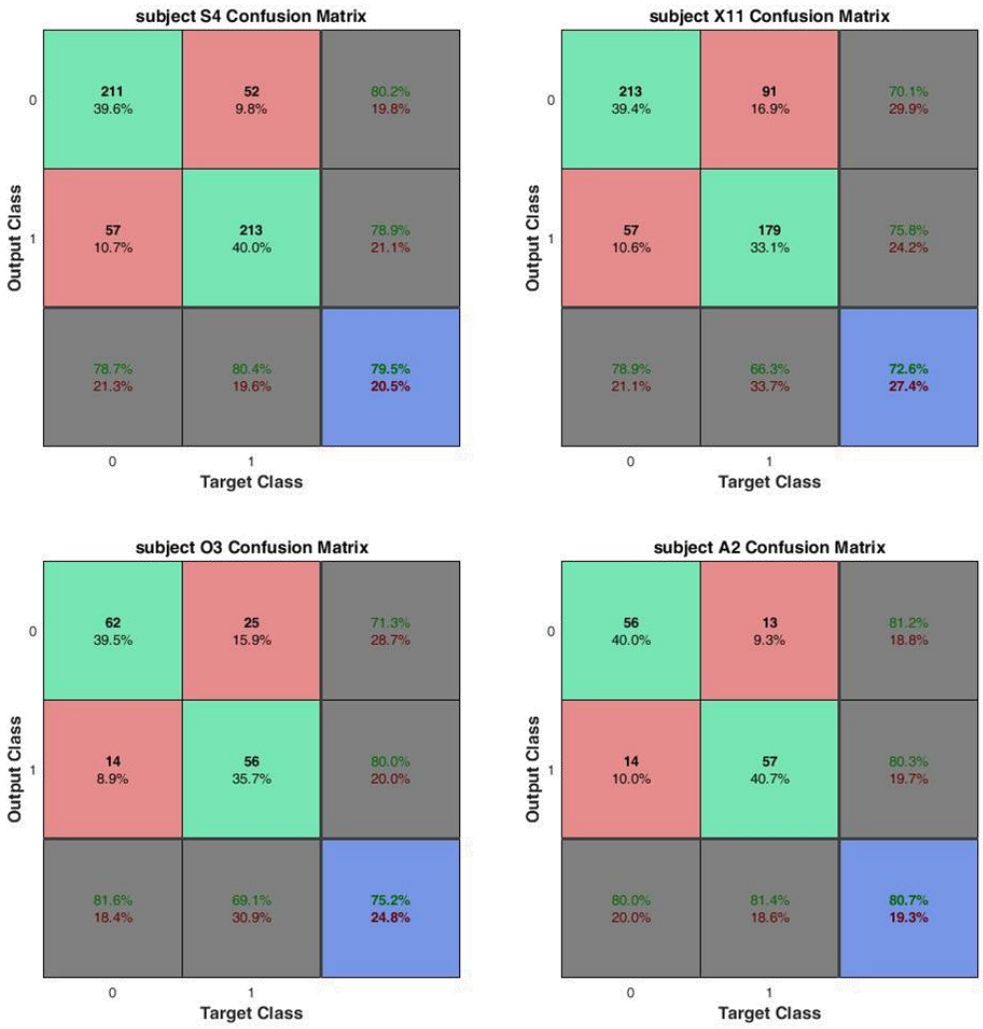

Figure 47. SVM with quadratic kernel

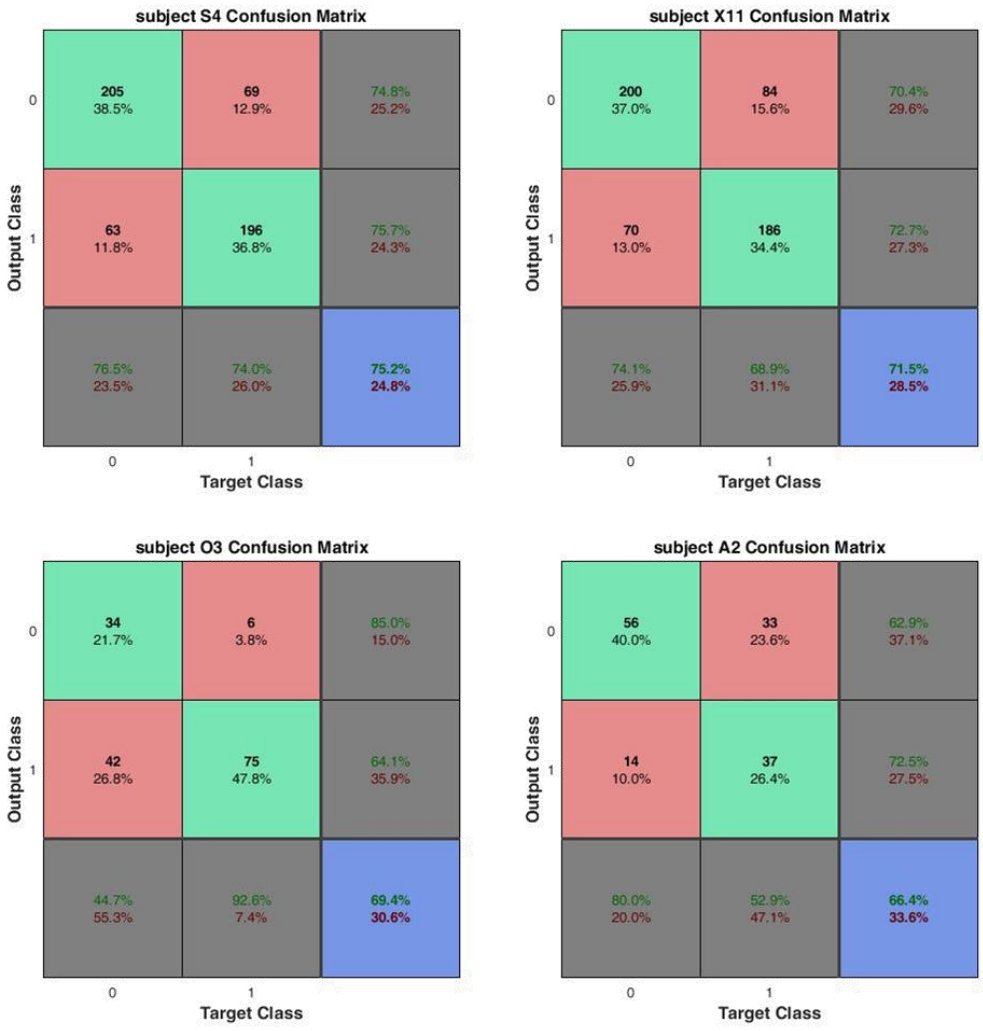

Figure 48. LDA

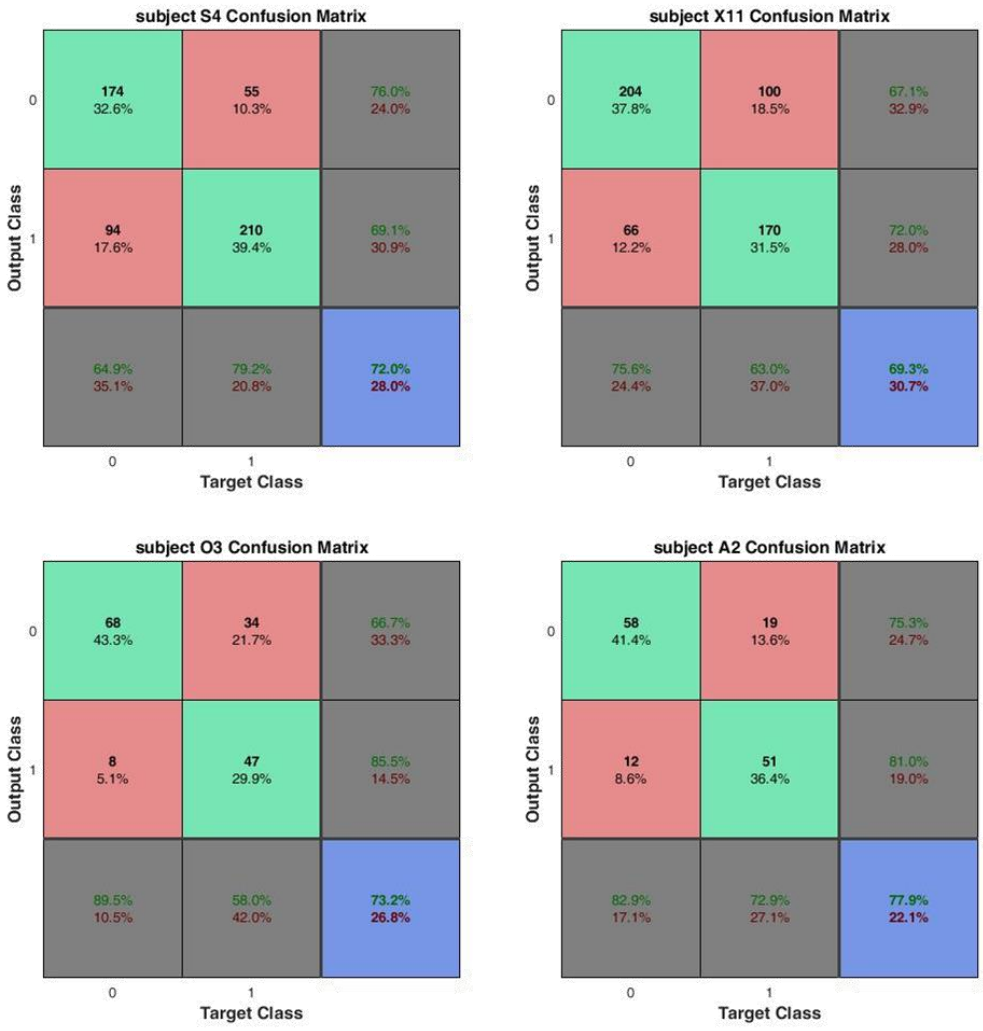

Figure 49. RBF

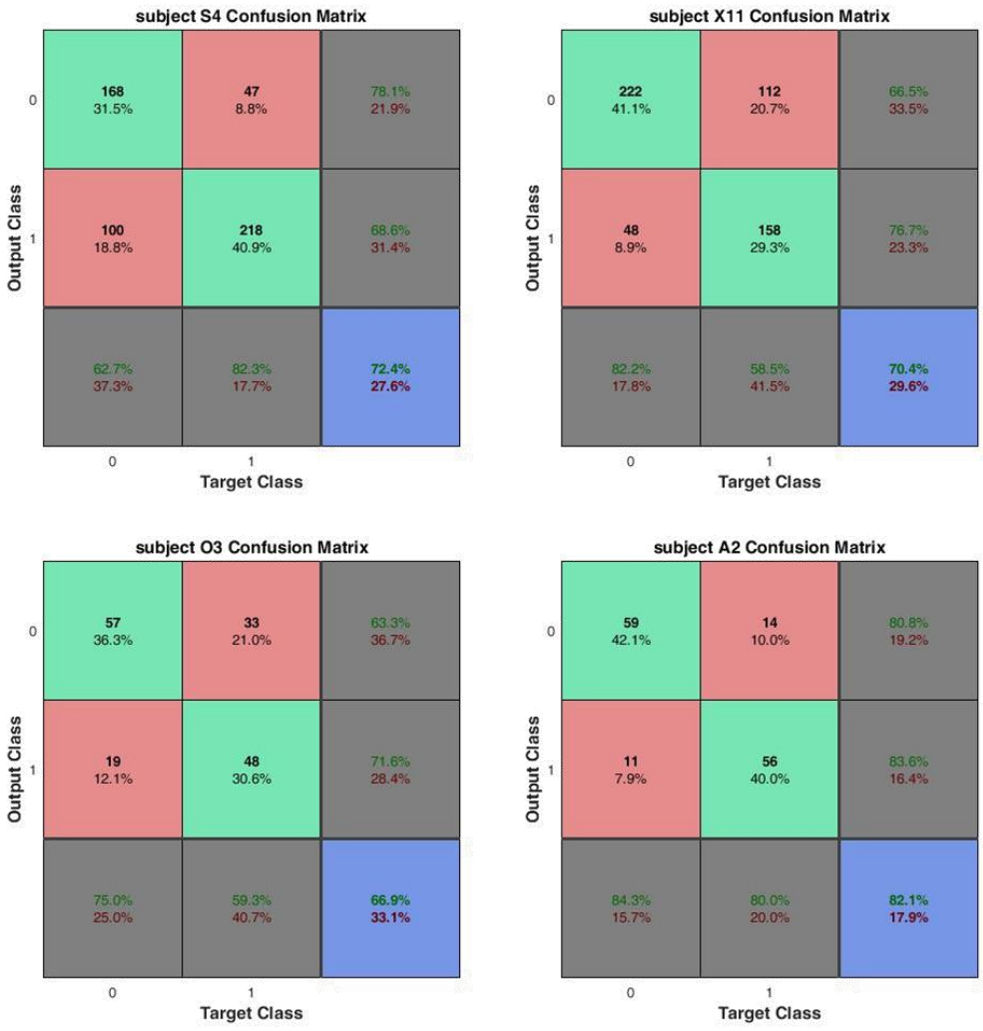

Figure 50. K-NN with Euclidean distance

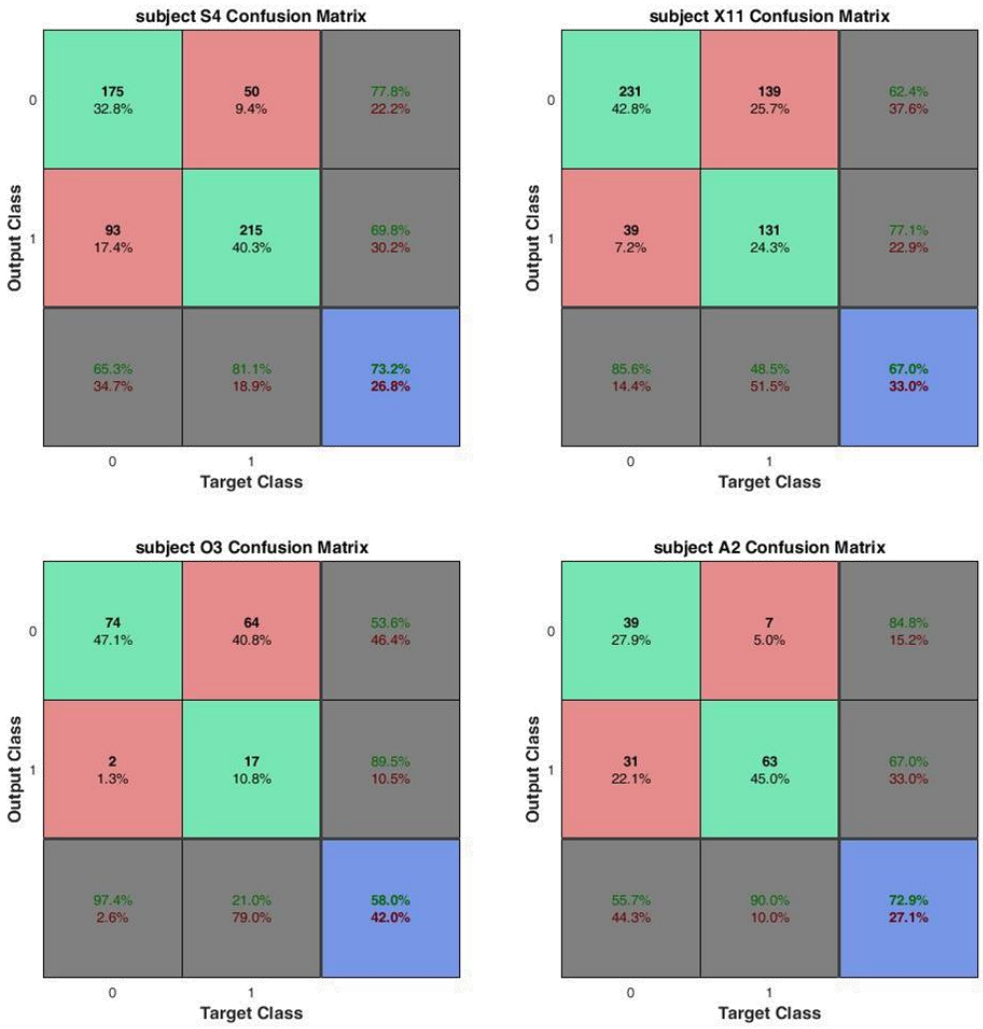

Figure 51. K-NN with Mahalanobis distance

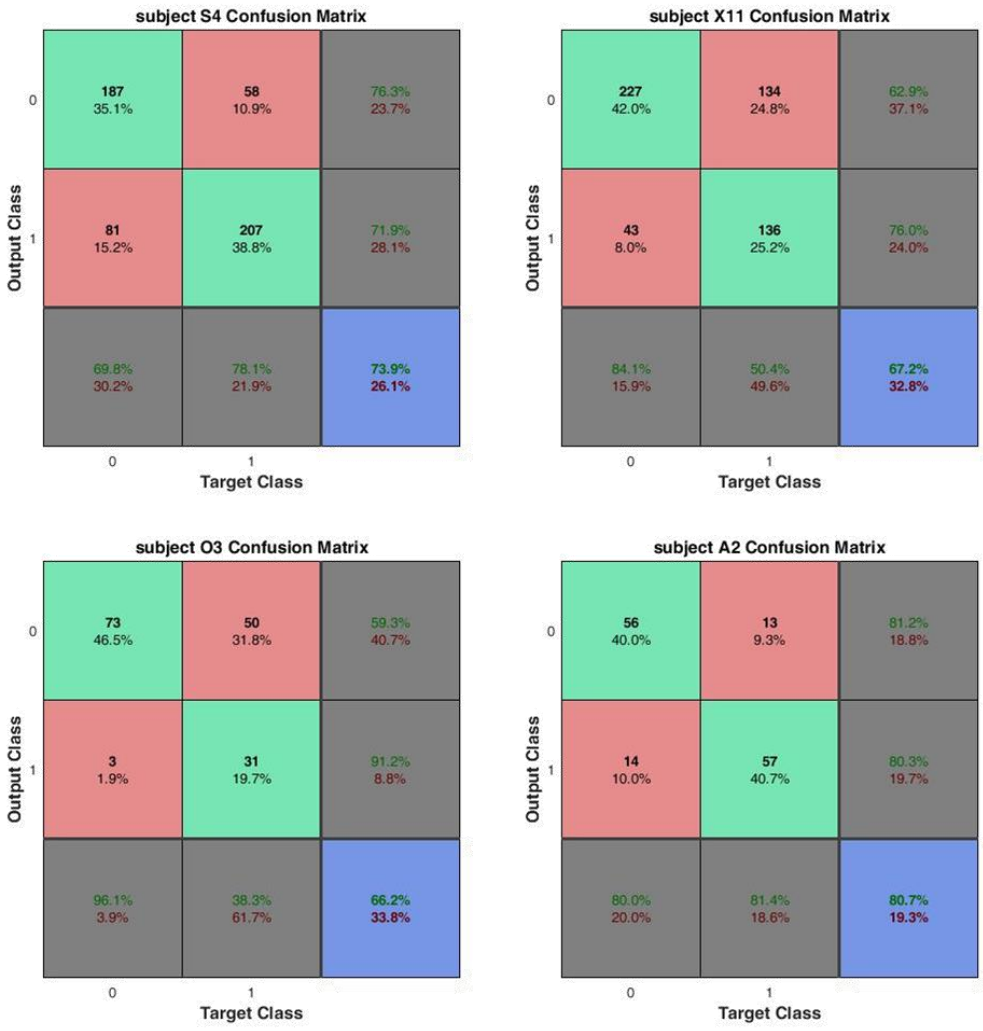

Figure 52. K-NN with cosine

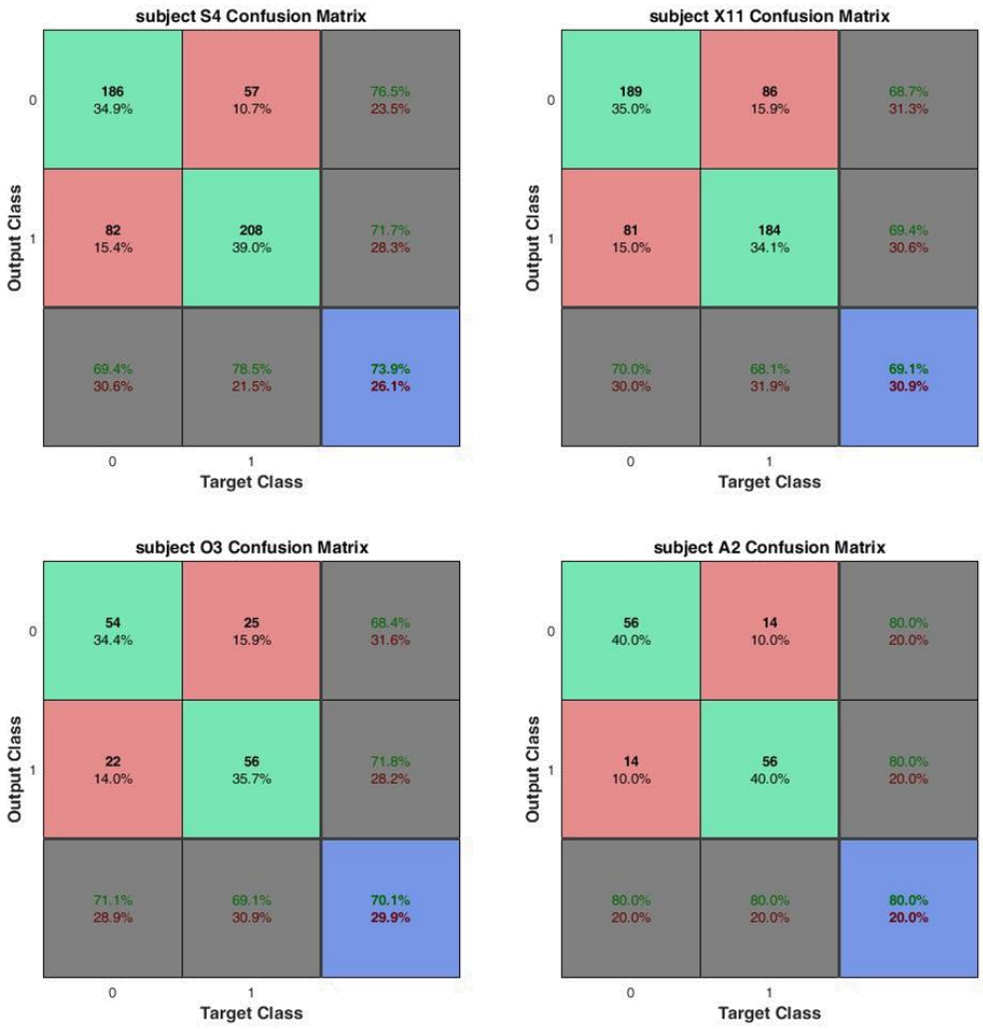

Figure 53. K-NN with correlation

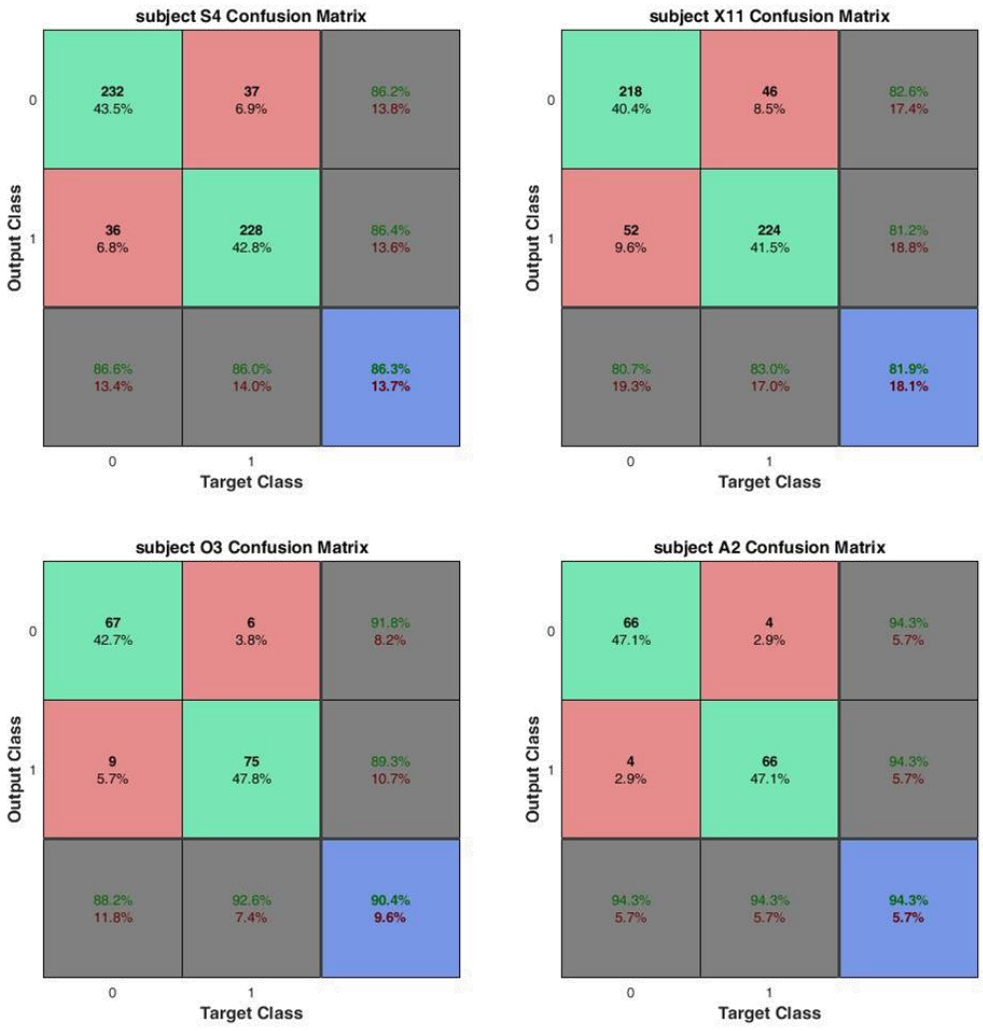

Figure 54. MLP

Appendix 3. Tables of selected features for subjects S4, X11, O3 and A2

Table 23. Subject S4

	PNN	SVM	SVM Q	LDA	RBF	KNN	KNN M	KNN COS	KNN COR	MLP	Total
Avg abs C3 sb0-4	1	1	1	1	1	1	1	1	1	1	**10**
Avg abs C3 sb4-8	0	0	1	0	0	0	0	0	1	0	**2**
Avg abs C3 sb8-12	1	1	1	1	1	0	1	0	1	1	**8**
Avg abs C3 sb12-16	1	0	0	1	0	0	0	0	1	1	**4**
Avg abs C3 sb16-20	0	0	1	0	0	0	0	0	1	0	**2**
Avg abs C3 sb20-24	1	0	0	0	1	1	0	1	1	0	**5**
Avg abs C3 sb24-28	1	1	0	1	0	1	0	0	0	1	**5**
Avg abs C3 sb28-32	1	1	0	0	1	1	1	1	1	1	**8**
Avg abs C4 sb0-4	0	0	1	1	1	1	1	0	1	0	**6**
Avg abs C4 sb4-8	1	1	1	1	1	0	0	0	0	1	**6**
Avg abs C4 sb8-12	0	0	0	0	0	0	0	0	0	1	**1**
Avg abs C4 sb12-16	1	1	0	1	0	0	1	1	0	0	**5**
Avg abs C4 sb16-20	0	0	1	0	0	1	0	0	1	0	**3**
Avg abs C4 sb20-24	1	1	1	1	0	1	0	1	1	0	**7**
Avg abs C4 sb24-28	0	0	0	0	0	0	0	0	0	1	**1**
Avg abs C4 sb28-32	0	0	0	0	0	0	1	0	0	1	**2**
Avg ampl C3 sb0-4	1	1	1	1	0	1	1	1	1	1	**9**
Avg ampl C3 sb4-8	1	0	1	1	1	0	0	0	0	1	**5**
Avg ampl C3 sb8-12	0	1	1	0	1	1	0	0	1	0	**5**
Avg ampl C3 sb12-16	0	0	1	1	1	1	0	1	0	0	**5**
Avg ampl C3 sb16-20	1	1	0	0	1	0	0	0	1	1	**5**
Avg ampl C3 sb20-24	1	1	1	0	1	0	0	1	0	1	**6**
Avg ampl C3 sb24-28	1	0	0	1	0	1	0	0	0	1	**4**
Avg ampl C3 sb28-32	1	1	1	0	0	1	1	1	0	0	**6**
Avg ampl C4 sb0-4	1	1	1	1	0	1	1	0	0	1	**7**
Avg ampl C4 sb4-8	1	0	0	1	0	0	0	0	1	0	**3**
Avg ampl C4 sb8-12	1	1	0	0	0	0	0	0	1	0	**3**
Avg ampl C4 sb12-16	0	1	1	1	1	1	1	0	0	0	**6**
Avg ampl C4 sb16-20	0	1	0	0	1	1	1	1	0	1	**6**
Avg ampl C4 sb20-24	1	0	1	0	1	0	1	0	1	1	**6**
Avg ampl C4 sb24-28	1	0	0	0	1	0	0	0	0	0	**2**
Avg ampl C4 sb28-32	0	0	0	1	1	0	1	0	0	1	**4**
Std dev C3 sb0-4	0	0	1	1	0	0	0	1	0	1	**4**

Std dev C3 sb4-8	1	0	1	0	0	0	0	1	0	0	**3**
Std dev C3 sb8-12	0	1	1	0	0	0	1	0	0	1	**4**
Std dev C3 sb12-16	0	1	1	1	1	1	1	1	0	0	**7**
Std dev C3 sb16-20	0	0	1	0	1	1	0	1	0	1	**5**
Std dev C3 sb20-24	0	1	1	0	1	0	0	1	1	1	**6**
Std dev C3 sb24-28	1	1	1	1	1	1	0	0	0	0	**6**
Std dev C3 sb28-32	0	0	0	1	1	1	0	1	1	0	**5**
Std dev C4 sb0-4	1	1	1	1	1	1	0	1	1	1	**9**
Std dev C4 sb4-8	1	0	1	0	1	0	0	1	1	0	**5**
Std dev C4 sb8-12	0	0	0	1	0	0	0	0	0	1	**2**
Std dev C4 sb12-16	1	0	0	0	0	0	0	1	0	0	**2**
Std dev C4 sb16-20	0	0	1	1	0	1	0	1	1	0	**5**
Std dev C4 sb20-24	0	0	0	0	1	1	0	1	1	1	**5**
Std dev C4 sb24-28	0	0	0	0	1	0	0	0	0	0	**1**
Std dev C4 sb28-32	0	0	0	0	0	0	0	0	0	0	**0**
Energy C3 sb0-4	0	0	0	0	1	0	0	0	0	1	**2**
Energy C3 sb4-8	0	0	0	1	0	0	1	0	0	0	**2**
Energy C3 sb8-12	1	0	0	0	0	0	0	0	0	1	**2**
Energy C3 sb12-16	0	0	1	1	0	0	0	0	0	1	**3**
Energy C3 sb16-20	1	1	0	1	1	1	1	1	0	1	**8**
Energy C3 sb20-24	0	0	1	1	1	1	0	1	1	0	**6**
Energy C3 sb24-28	1	1	1	0	1	1	1	1	1	0	**8**
Energy C3 sb28-32	1	1	0	1	1	1	1	1	1	1	**9**
Energy C4 sb0-4	0	0	1	0	0	0	0	0	0	1	**2**
Energy C4 sb4-8	0	0	0	1	1	0	0	0	0	1	**3**
Energy C4 sb8-12	0	0	0	0	0	1	0	0	1	0	**2**
Energy C4 sb12-16	0	0	0	1	1	1	1	0	0	0	**4**
Energy C4 sb16-20	0	0	0	1	0	1	1	0	1	0	**4**
Energy C4 sb20-24	1	1	0	0	0	0	1	0	1	1	**5**
Energy C4 sb24-28	0	1	1	1	0	1	1	1	0	1	**7**
Energy C4 sb28-32	0	0	0	0	0	0	0	0	0	1	**1**
Rt En C3 sb0-4 and sb4-8	1	1	0	0	0	0	0	0	1	1	**4**
Rt En C3 sb4-8 and sb8-12	1	1	1	0	1	1	0	1	1	0	**7**
Rt En C3 sb8-12 and sb12-16	1	1	1	0	0	1	0	1	1	0	**6**
Rt En C3 sb12-16 and sb16-20	0	0	0	0	1	0	1	0	0	0	**2**
Rt En C3 sb16-20 and sb20-24	1	0	0	0	0	0	0	1	1	1	**4**
Rt En C3 sb20-24 and sb24-28	0	0	0	1	0	1	1	1	0	0	**4**
Rt En C3 sb24-28 and sb28-32	1	0	0	0	0	0	0	0	0	1	**2**
Rt En C4 sb0-4 and sb4-8	0	0	0	0	0	1	0	0	1	0	**2**
Rt En C4 sb4-8 and sb8-12	1	1	0	0	0	0	1	1	1	1	**6**
Rt En C4 sb8-12 and sb12-16	0	1	1	1	0	0	0	0	1	1	**5**

Rt En C4 sb12-16 and sb16-20	0	1	0	0	1	0	0	0	1	0	**3**
Rt En C4 sb16-20 and sb20-24	0	0	1	0	0	0	1	1	0	0	**3**
Rt En C4 sb20-24 and sb24-28	1	0	1	0	1	0	0	1	0	0	**4**
Rt En C4 sb24-28 and sb28-32	0	0	0	0	0	1	0	1	0	1	**3**
Entropy C3 sb0-4	1	1	1	1	1	0	1	0	0	1	**7**
Entropy C3 sb4-8	0	0	1	0	0	1	1	1	1	0	**5**
Entropy C3 sb8-12	0	1	0	1	1	1	0	0	0	0	**4**
Entropy C3 sb12-16	0	1	0	1	1	1	0	1	0	1	**6**
Entropy C3 sb16-20	1	0	0	0	1	0	0	1	1	0	**4**
Entropy C3 sb20-24	1	1	1	0	1	1	0	1	1	0	**7**
Entropy C3 sb24-28	1	1	1	1	0	0	0	0	0	0	**4**
Entropy C3 sb28-32	0	0	1	1	1	1	1	1	1	0	**7**
Entropy C4 sb0-4	0	1	0	1	0	0	1	1	1	0	**5**
Entropy C4 sb4-8	1	0	0	1	0	1	0	1	0	1	**5**
Entropy C4 sb8-12	0	1	0	0	1	1	0	0	0	1	**4**
Entropy C4 sb12-16	0	1	1	0	1	1	0	1	0	0	**5**
Entropy C4 sb16-20	1		1	0	0	1	0	0	1	0	**4**
Entropy C4 sb20-24	0	0	1	0	0	0	0	0	0	0	**1**
Entropy C4 sb24-28	1	1	0	0	0	0	0	1	1	1	**5**
Entropy C4 sb28-32	0	0	1	1	1	0	1	1	0	0	**5**
PLV C3 and C4 sb0-4	1	1	1	0	0	1	1	1	1	1	**8**
PLV C3 and C4 sb4-8	0	0	0	1	0	0	0	1	0	0	**2**
PLV C3 and C4 sb8-12	1	0	1	1	0	1	0	0	1	1	**6**
PLV C3 and C4 sb12-16	0	0	1	0	0	0	1	1	1	0	**4**
PLV C3 and C4 sb16-20	0	0	0	0	0	0	0	0	0	1	**1**
PLV C3 and C4 sb20-24	1	0	0	0	1	1	1	0	0	1	**5**
PLV C3 and C4 sb24-28	0	1	1	0	0	1	1	1	1	0	**6**
PLV C3 and C4 sb28-32	1	1	1	0	1	0	1	0	1	0	**6**
Root mean sq C3 sb0-4	0	0	1	1	0	1	1	1	1	0	**6**
Root mean sq C3 sb4-8	1	0	1	1	1	1	1	0	1	0	**7**
Root mean sq C3 sb8-12	1	1	0	0	1	0	1	1	1	1	**7**
Root mean sq C3 sb12-16	1	1	0	1	1	1	1	0	1	1	**8**
Root mean sq C3 sb16-20	0	0	0	0	0	0	0	0	1	0	**1**
Root mean sq C3 sb20-24	1	0	0	0	1	1	1	1	1	0	**6**
Root mean sq C3 sb24-28	0	0	0	1	1	0	1	1	0	1	**5**
Root mean sq C3 sb28-32	0	1	0	0	0	0	1	1	1	0	**4**
Root mean sq C4 sb0-4	0	1	1	0	0	1	0	1	1	1	**6**
Root mean sq C4 sb4-8	1	1	1	1	1	1	0	1	0	0	**7**
Root mean sq C4 sb8-12	0	1	0	1	1	0	0	1	1	0	**5**
Root mean sq C4 sb12-16	0	1	0	0	1	0	0	1	0	1	**4**
Root mean sq C4 sb16-20	1	1	0	0	1	1	0	0	0	1	**5**

Root mean sq C4 sb20-24	0	1	0	0	0	1	0	0	0	0	**2**
Root mean sq C4 sb24-28	0	0	0	1	1	0	0	1	0	0	**3**
Root mean sq C4 sb28-32	0	0	0	0	0	0	0	0	0	0	**0**
Variance C3 sb0-4	1	0	0	0	1	0	1	0	0	0	**3**
Variance C3 sb4-8	0	0	0	1	0	0	1	1	1	1	**5**
Variance C3 sb8-12	0	0	0	0	0	0	0	0	0	0	**0**
Variance C3 sb12-16	0	1	0	1	0	0	1	0	1	0	**4**
Variance C3 sb16-20	1	1	1	0	0	1	0	0	0	1	**5**
Variance C3 sb20-24	1	1	0	1	0	0	0	1	1	0	**5**
Variance C3 sb24-28	0	1	0	0	1	0	0	0	1	1	**4**
Variance C3 sb28-32	1	1	1	1	1	1	1	1	1	1	**10**
Variance C4 sb0-4	0	0	1	1	1	0	0	0	0	0	**3**
Variance C4 sb4-8	1	1	1	1	0	1	0	1	0	0	**6**
Variance C4 sb8-12	0	0	1	0	1	0	0	0	0	1	**3**
Variance C4 sb12-16	0	1	0	0	0	1	1	1	1	0	**5**
Variance C4 sb16-20	1	0	0	1	1	0	0	1	0	1	**5**
Variance C4 sb20-24	0	0	0	0	0	0	0	1	1	0	**2**
Variance C4 sb24-28	1	1	0	1	0	1	0	0	0	0	**4**
Variance C4 sb28-32	0	0	0	0	1	0	0	0	0	0	**1**

Table 24. Subject X11

	PNN	**SVM**	**SVM Q**	**LDA**	**RBF**	**KNN**	**KNN M**	**KNN COS**	**KNN COR**	**MLP**	**Total**
Avg abs C3 sb0-4	1	0	0	1	1	0	0	0	0	0	**3**
Avg abs C3 sb4-8	0	1	1	0	1	0	1	1	0	0	**5**
Avg abs C3 sb8-12	0	1	1	0	1	1	1	0	1	1	**7**
Avg abs C3 sb12-16	0	0	1	0	1	1	0	0	1	0	**4**
Avg abs C3 sb16-20	1	0	0	1	1	0	1	1	0	1	**6**
Avg abs C3 sb20-24	1	0	0	0	1	0	0	0	0	1	**3**
Avg abs C3 sb24-28	0	1	0	0	1	0	0	0	1	0	**3**
Avg abs C3 sb28-32	0	1	0	0	0	1	1	0	0	0	**3**
Avg abs C4 sb0-4	1	0	0	1	1	0	0	1	0	0	**4**
Avg abs C4 sb4-8	1	0	0	0	0	1	0	1	0	0	**3**
Avg abs C4 sb8-12	0	0	0	1	0	0	0	0	0	0	**1**
Avg abs C4 sb12-16	0	1	0	1	0	0	1	0	1	0	**4**
Avg abs C4 sb16-20	1	1	1	0	1	1	0	1	1	0	**7**
Avg abs C4 sb20-24	0	1	0	1	0	0	1	0	0	1	**4**
Avg abs C4 sb24-28	0	0	0	1	0	0	1	0	1	1	**4**
Avg abs C4 sb28-32	1	1	1	1	1	1	0	0	1	0	**7**
Avg ampl C3 sb0-4	0	1	1	1	0	1	0	1	0	1	**6**
Avg ampl C3 sb4-8	0	0	0	0	0	0	0	1	0	0	**1**
Avg ampl C3 sb8-12	1	0	0	0	1	0	0	0	1	1	**4**
Avg ampl C3 sb12-16	0	1	1	0	1	1	0	0	1	1	**6**
Avg ampl C3 sb16-20	1	1	1	0	0	0	1	1	0	1	**6**
Avg ampl C3 sb20-24	0	0	0	1	0	0	0	0	1	0	**2**

Avg ampl C3 sb24-28	1	1	1	1	1	0	0	0	1	0	**6**
Avg ampl C3 sb28-32	0	0	0	1	1	1	0	1	0	1	**5**
Avg ampl C4 sb0-4	1	0	0	0	1	1	1	0	1	0	**5**
Avg ampl C4 sb4-8	0	0	1	0	0	0	0	0	1	0	**2**
Avg ampl C4 sb8-12	1	1	0	0	0	0	0	1	0	1	**4**
Avg ampl C4 sb12-16	0	1	1	1	0	0	0	0	1	1	**5**
Avg ampl C4 sb16-20	1	1	0	0	1	1	1	1	1	0	**7**
Avg ampl C4 sb20-24	0	1	1	0	0	1	0	0	1	1	**5**
Avg ampl C4 sb24-28	1	1	0	0	1	1	0	0	1	1	**6**
Avg ampl C4 sb28-32	0	0	1	1	1	1	0	1	1	0	**6**
Std dev C3 sb0-4	0	0	1	0	0	0	0	1	0	0	**2**
Std dev C3 sb4-8	1	0	0	0	1	0	1	0	1	1	**5**
Std dev C3 sb8-12	0	1	0	0	0	1	1	0	0	1	**4**
Std dev C3 sb12-16	0	0	1	1	1	0	0	0	0	0	**3**
Std dev C3 sb16-20	0	1	0	0	0	1	1	1	0	0	**4**
Std dev C3 sb20-24	1	0	1	0	0	1	0	0	1	1	**5**
Std dev C3 sb24-28	0	1	0	1	0	0	1	0	1	0	**4**
Std dev C3 sb28-32	0	0	0	0	0	0	1	0	0	1	**2**
Std dev C4 sb0-4	0	0	1	0	0	0	1	1	0	1	**4**
Std dev C4 sb4-8	1	1	1	0	1	0	1	1	1	1	**8**
Std dev C4 sb8-12	0	0	0	1	0	0	0	0	0	1	**2**
Std dev C4 sb12-16	0	0	0	0	0	0	0	1	1	1	**3**
Std dev C4 sb16-20	1	0	1	0	1	1	0	1	1	1	**7**
Std dev C4 sb20-24	0	1	1	1	0	1	0	0	1	1	**6**
Std dev C4 sb24-28	0	1	0	1	0	1	1	0	0	1	**5**
Std dev C4 sb28-32	1	0	1	0	0	1	1	1	1	1	**7**
Energy C3 sb0-4	1	1	1	0	1	0	0	1	1	0	**6**
Energy C3 sb4-8	1	0	0	1	1	0	1	1	0	1	**6**
Energy C3 sb8-12	1	0	0	1	1	0	1	1	1	1	**7**
Energy C3 sb12-16	1	1	0	0	0	1	0	1	1	1	**6**
Energy C3 sb16-20	1	0	1	0	1	0	0	1	0	1	**5**
Energy C3 sb20-24	1	0	1	0	0	1	1	1	1	0	**6**
Energy C3 sb24-28	0	0	0	1	0	1	0	0	0	0	**2**
Energy C3 sb28-32	0	0	1	0	0	0	0	0	0	1	**2**
Energy C4 sb0-4	0	0	1	0	1	0	0	0	0	1	**3**
Energy C4 sb4-8	1	1	0	1	1	0	1	0	0	0	**5**
Energy C4 sb8-12	0	1	0	1	0	0	0	0	0	0	**2**
Energy C4 sb12-16	0	0	0	0	1	1	1	0	1	0	**4**
Energy C4 sb16-20	0	0	1	0	0	0	1	1	0	1	**4**
Energy C4 sb20-24	1	0	1	1	0	0	1	0	1	0	**5**
Energy C4 sb24-28	1	1	0	0	1	1	0	0	0	0	**4**
Energy C4 sb28-32	1	0	1	1	0	1	1	1	1	1	**8**
Rt En C3 sb0-4 and sb4-8	0	0	0	0	0	0	1	1	0	1	**3**
Rt En C3 sb4-8 and sb8-12	1	0	0	1	1	0	0	0	1	0	**4**
Rt En C3 sb8-12 and sb12-16	0	0	1	0	0	1	0	1	0	0	**3**
Rt En C3 sb12-16 and sb16-20	1	1	0	1	1	1	1	0	0	1	**7**
Rt En C3 sb16-20 and sb20-24	0	0	1	0	1	0	1	1	1	0	**5**
Rt En C3 sb20-24 and sb24-28	1	1	1	1	1	1	1	1	1	0	**9**
Rt En C3 sb24-28 and sb28-32	1	0	0	1	1	0	0	0	0	0	**3**

Rt En C4 sb0-4 and sb4-8	0	1	0	1	1	0	1	1	0	1	**6**
Rt En C4 sb4-8 and sb8-12	1	1	1	0	0	0	0	0	0	0	**3**
Rt En C4 sb8-12 and sb12-16	0	1	0	1	0	0	1	1	0	0	**4**
Rt En C4 sb12-16 and sb16-20	0	0	1	0	0	0	0	0	1	1	**3**
Rt En C4 sb16-20 and sb20-24	0	1	0	1	1	0	0	0	1	1	**5**
Rt En C4 sb20-24 and sb24-28	0	1	0	1	0	0	1	0	0	1	**4**
Rt En C4 sb24-28 and sb28-32	1	0	0	0	1	1	1	1	1	0	**6**
Entropy C3 sb0-4	1	0	1	0	1	0	0	0	1	1	**5**
Entropy C3 sb4-8	1	1	0	1	1	0	1	1	1	0	**7**
Entropy C3 sb8-12	0	0	1	0	0	1	0	0	0	0	**2**
Entropy C3 sb12-16	0	0	0	0	0	1	0	0	0	0	**1**
Entropy C3 sb16-20	1	1	0	0	1	0	0	1	0	1	**5**
Entropy C3 sb20-24	1	1	1	1	1	0	1	1	1	1	**9**
Entropy C3 sb24-28	1	0	1	0	1	1	0	0	0	1	**5**
Entropy C3 sb28-32	1	0	1	0	1	1	1	0	1	1	**7**
Entropy C4 sb0-4	0	1	0	0	0	1	1	0	1	1	**5**
Entropy C4 sb4-8	1	0	1	0	0	1	0	1	0	0	**4**
Entropy C4 sb8-12	0	0	1	1	0	0	0	0	0	0	**2**
Entropy C4 sb12-16	1	0	0	1	1	1	1	1	0	0	**6**
Entropy C4 sb16-20	0	0	1	1	0	1	1	0	0	0	**4**
Entropy C4 sb20-24	1	1	1	1	1	0	1	1	1	0	**8**
Entropy C4 sb24-28	1	0	1	0	1	0	0	1	1	1	**6**
Entropy C4 sb28-32	1	0	1	0	0	0	0	1	1	0	**4**
PLV C3 and C4 sb0-4	0	1	0	0	0	0	0	0	1	0	**2**
PLV C3 and C4 sb4-8	0	1	1	0	1	0	0	1	0	1	**5**
PLV C3 and C4 sb8-12	0	0	0	0	1	1	0	0	1	0	**3**
PLV C3 and C4 sb12-16	0	0	0	0	1	0	0	1	1	0	**3**
PLV C3 and C4 sb16-20	0	0	1	0	0	0	0	0	1	1	**3**
PLV C3 and C4 sb20-24	1	1	1	1	1	1	1	0	0	1	**8**
PLV C3 and C4 sb24-28	0	0	0	0	0	0	0	0	1	1	**2**
PLV C3 and C4 sb28-32	1	1	1	0	0	1	1	1	0	1	**7**
Root mean sq C3 sb0-4	0	1	1	0	0	0	0	1	0	0	**3**
Root mean sq C3 sb4-8	0	0	1	1	1	0	0	0	0	0	**3**
Root mean sq C3 sb8-12	0	0	0	1	0	1	0	1	1	1	**5**
Root mean sq C3 sb12-16	1	0	0	1	1	0	1	0	1	0	**5**
Root mean sq C3 sb16-20	0	0	1	0	1	0	1	1	1	0	**5**
Root mean sq C3 sb20-24	0	0	0	1	0	0	0	0	0	0	**1**
Root mean sq C3 sb24-28	0	0	0	1	0	0	0	1	1	0	**3**
Root mean sq C3 sb28-32	1	1	1	0	1	1	1	0	0	1	**7**
Root mean sq C4 sb0-4	1	1	0	0	0	0	0	1	0	0	**3**
Root mean sq C4 sb4-8	0	0	0	0	1	1	0	1	1	0	**4**
Root mean sq C4 sb8-12	1	0	0	1	0	0	0	0	0	0	**2**
Root mean sq C4 sb12-16	0	1	0	1	0	0	0	1	1	0	**4**
Root mean sq C4 sb16-20	1	0	0	1	1	1	1	1	1	1	**8**
Root mean sq C4 sb20-24	0	1	0	1	1	0	0	1	0	1	**5**
Root mean sq C4 sb24-28	0	0	1	0	0	0	1	0	1	1	**4**
Root mean sq C4 sb28-32	1	1	1	0	0	1	1	1	1	1	**8**
Variance C3 sb0-4	1	0	0	0	0	1	0	1	0	0	**3**
Variance C3 sb4-8	0	0	1	1	1	0	0	0	0	0	**3**

Variance C3 sb8-12	1	1	0	0	0	1	1	1	0	1	**6**
Variance C3 sb12-16	0	1	1	0	0	1	1	0	1	1	**6**
Variance C3 sb16-20	1	1	0	0	1	1	1	1	1	0	**7**
Variance C3 sb20-24	0	1	0	1	1	0	0	0	0	0	**3**
Variance C3 sb24-28	0	0	1	0	0	1	1	0	0	0	**3**
Variance C3 sb28-32	0	1	0	0	0	1	1	0	0	0	**3**
Variance C4 sb0-4	1	0	1	0	0	0	1	1	0	1	**5**
Variance C4 sb4-8	0	1	0	1	1	0	1	0	1	0	**5**
Variance C4 sb8-12	1	1	0	1	0	1	0	0	1	1	**6**
Variance C4 sb12-16	1	0	0	1	1	0	0	0	0	0	**3**
Variance C4 sb16-20	1	1	0	1	0	1	1	1	1	0	**7**
Variance C4 sb20-24	1	0	0	0	1	0	1	1	1	1	**6**
Variance C4 sb24-28	0	0	0	1	0	0	0	1	0	1	**3**
Variance C4 sb28-32	0	0	0	1	1	0	1	1	1	0	**5**

Table 25. Subject O3

	PNN	**SVM**	**SVM Q**	**LDA**	**RBF**	**KNN**	**KNN M**	**KNN COS**	**KNN COR**	**MLP**	**Total**
Avg abs C3 sb0-4	0	1	1	1	0	0	1	1	0	1	**6**
Avg abs C3 sb4-8	0	0	0	1	0	0	0	0	0	1	**2**
Avg abs C3 sb8-12	1	0	0	1	1	0	0	0	1	0	**4**
Avg abs C3 sb12-16	1	1	1	0	1	1	0	1	1	0	**7**
Avg abs C3 sb16-20	0	0	1	1	0	0	1	0	0	1	**4**
Avg abs C3 sb20-24	1	1	1	0	0	1	1	1	1	0	**7**
Avg abs C3 sb24-28	1	0	0	0	0	0	1	1	0	1	**4**
Avg abs C3 sb28-32	0	0	0	0	0	1	0	1	1	1	**4**
Avg abs C4 sb0-4	1	1	0	1	0	1	0	0	0	0	**4**
Avg abs C4 sb4-8	0	1	0	0	1	0	0	0	0	0	**2**
Avg abs C4 sb8-12	1	1	0	0	1	0	0	0	1	1	**5**
Avg abs C4 sb12-16	1	1	0	1	1	0	0	1	1	1	**7**
Avg abs C4 sb16-20	0	0	1	0	0	0	1	0	0	0	**2**
Avg abs C4 sb20-24	0	0	0	1	1	1	0	1	1	0	**5**
Avg abs C4 sb24-28	1	1	0	1	0	0	0	0	1	1	**5**
Avg abs C4 sb28-32	0	0	1	0	1	1	1	1	1	1	**7**
Avg ampl C3 sb0-4	0	1	0	0	0	1	0	0	0	0	**2**
Avg ampl C3 sb4-8	0	1	0	1	1	0	1	0	0	0	**4**
Avg ampl C3 sb8-12	0	1	1	0	0	0	0	0	0	1	**3**
Avg ampl C3 sb12-16	0	0	1	1	0	1	1	0	1	1	**6**
Avg ampl C3 sb16-20	1	0	0	1	1	1	0	0	0	1	**5**
Avg ampl C3 sb20-24	1	1	0	0	0	1	1	1	0	0	**5**
Avg ampl C3 sb24-28	1	0	1	0	1	1	0	0	0	0	**4**
Avg ampl C3 sb28-32	1	1	1	1	0	1	0	1	1	0	**7**
Avg ampl C4 sb0-4	0	1	0	0	1	0	0	1	0	1	**4**

Avg ampl C4 sb4-8	1	0	0	0	0	1	0	0	0	1	**3**
Avg ampl C4 sb8-12	1	0	1	1	1	0	0	1	1	1	**7**
Avg ampl C4 sb12-16	1	1	1	0	1	1	1	1	1	1	**9**
Avg ampl C4 sb16-20	0	0	0	1	0	0	0	1	1	1	**4**
Avg ampl C4 sb20-24	0	1	1	0	0	0	1	0	1	0	**4**
Avg ampl C4 sb24-28	0	0	0	1	1	0	0	1	1	0	**4**
Avg ampl C4 sb28-32	1	1	1	1	0	1	0	0	1	1	**7**
Std dev C3 sb0-4	0	0	0	0	0	1	0	1	0	0	**2**
Std dev C3 sb4-8	0	0	1	1	1	0	1	1	1	0	**6**
Std dev C3 sb8-12	0	1	1	1	0	1	1	1	1	1	**8**
Std dev C3 sb12-16	0	1	1	0	1	1	0	0	0	0	**4**
Std dev C3 sb16-20	0	1	0	1	1	1	1	0	1	1	**7**
Std dev C3 sb20-24	0	0	1	1	1	0	1	0	1	0	**5**
Std dev C3 sb24-28	0	1	1	0	1	1	1	0	0	0	**5**
Std dev C3 sb28-32	0	1	1	1	1	1	1	1	0	0	**7**
Std dev C4 sb0-4	0	1	0	0	1	0	1	1	0	0	**4**
Std dev C4 sb4-8	1	1	1	1	0	1	0	0	0	1	**6**
Std dev C4 sb8-12	1	1	0	0	0	1	1	1	1	1	**7**
Std dev C4 sb12-16	1	1	0	0	0	0	1	0	0	0	**3**
Std dev C4 sb16-20	0	1	1	1	0	1	1	0	1	0	**6**
Std dev C4 sb20-24	1	1	0	1	1	0	0	0	0	1	**5**
Std dev C4 sb24-28	0	0	0	1	0	0	0	1	0	1	**3**
Std dev C4 sb28-32	0	1	1	0	0	0	0	0	1	1	**4**
Energy C3 sb0-4	1	0	0	0	1	0	1	1	0	0	**4**
Energy C3 sb4-8	0	0	0	1	1	1	1	1	1	0	**6**
Energy C3 sb8-12	1	1	1	0	0	0	1	1	0	0	**5**
Energy C3 sb12-16	0	1	0	0	1	0	0	0	0	0	**2**
Energy C3 sb16-20	0	1	0	1	0	0	0	1	1	1	**5**
Energy C3 sb20-24	0	1	1	1	1	0	0	1	0	1	**6**
Energy C3 sb24-28	0	0	0	0	0	0	0	0	0	1	**1**
Energy C3 sb28-32	1	1	0	0	0	0	0	1	0	0	**3**
Energy C4 sb0-4	0	1	1	0	0	1	1	0	0	1	**5**
Energy C4 sb4-8	1	0	0	0	0	0	0	0	1	0	**2**
Energy C4 sb8-12	1	1	1	0	1	1	1	0	1	1	**8**
Energy C4 sb12-16	0	0	0	0	1	0	0	1	1	1	**4**
Energy C4 sb16-20	1	0	0	0	1	0	1	1	1	0	**5**
Energy C4 sb20-24	1	0	1	0	0	1	1	1	0	1	**6**
Energy C4 sb24-28	1	0	0	0	1	1	1	0	0	0	**4**
Energy C4 sb28-32	0	0	1	0	1	1	0	0	0	0	**3**
Rt En C3 sb0-4 and sb4-8	1	0	1	1	0	1	0	1	1	1	**7**
Rt En C3 sb4-8 and sb8-12	0	1	1	0	0	0	1	1	1	1	**6**

Rt En C3 sb8-12 and sb12-16	1	0	1	1	1	1	0	0	0	0	**5**
Rt En C3 sb12-16 and sb16-20	1	0	1	0	1	0	0	1	0	0	**4**
Rt En C3 sb16-20 and sb20-24	0	1	0	0	0	1	0	0	1	1	**4**
Rt En C3 sb20-24 and sb24-28	1	1	1	0	0	0	0	0	0	1	**4**
Rt En C3 sb24-28 and sb28-32	1	0	1	0	1	1	0	1	1	1	**7**
Rt En C4 sb0-4 and sb4-8	0	0	0	0	0	1	0	1	0	0	**2**
Rt En C4 sb4-8 and sb8-12	0	1	0	0	0	0	0	1	1	1	**4**
Rt En C4 sb8-12 and sb12-16	1	0	1	0	0	0	0	1	1	1	**5**
Rt En C4 sb12-16 and sb16-20	1	0	1	0	1	1	0	0	1	0	**5**
Rt En C4 sb16-20 and sb20-24	0	0	0	0	0	0	0	1	0	0	**1**
Rt En C4 sb20-24 and sb24-28	0	0	0	1	0	0	0	0	1	0	**2**
Rt En C4 sb24-28 and sb28-32	0	0	0	1	0	0	0	0	0	0	**1**
Entropy C3 sb0-4	1	0	0	0	1	0	0	1	0	0	**3**
Entropy C3 sb4-8	0	1	0	0	0	1	0	0	1	1	**4**
Entropy C3 sb8-12	0	1	1	1	1	0	0	1	1	0	**6**
Entropy C3 sb12-16	1	0	1	1	0	0	0	1	0	0	**4**
Entropy C3 sb16-20	0	1	1	0	0	0	0	0	0	0	**2**
Entropy C3 sb20-24	1	0	1	0	1	1	1	0	1	0	**6**
Entropy C3 sb24-28	1	0	1	0	1	0	1	1	1	1	**7**
Entropy C3 sb28-32	1	0	1	1	0	1	1	1	1	0	**7**
Entropy C4 sb0-4	1	1	0	1	0	1	0	0	0	0	**4**
Entropy C4 sb4-8	0	1	1	1	1	1	1	0	0	0	**6**
Entropy C4 sb8-12	1	0	0	0	0	0	1	1	0	0	**3**
Entropy C4 sb12-16	0	1	1	0	0	0	1	0	0	0	**3**
Entropy C4 sb16-20	0	0	1	0	0	0	0	0	1	1	**3**
Entropy C4 sb20-24	1	1	0	1	1	1	0	0	0	0	**5**
Entropy C4 sb24-28	0	1	1	1	0	1	1	0	0	0	**5**
Entropy C4 sb28-32	1	0	1	1	1	0	1	0	1	1	**7**
PLV C3 and C4 sb0-4	1	0	0	0	0	0	0	0	0	0	**1**
PLV C3 and C4 sb4-8	1	0	1	0	1	0	0	1	0	0	**4**
PLV C3 and C4 sb8-12	0	1	1	0	0	0	0	1	1	1	**5**
PLV C3 and C4 sb12-16	0	1	0	1	0	0	0	1	0	1	**4**
PLV C3 and C4 sb16-20	0	1	0	0	0	0	0	0	0	0	**1**
PLV C3 and C4 sb20-24	0	0	0	0	1	0	1	0	1	1	**4**
PLV C3 and C4 sb24-28	1	0	0	0	0	1	1	0	0	0	**3**
PLV C3 and C4 sb28-32	0	0	0	0	1	0	1	1	0	0	**3**
Root mean sq C3 sb0-4	0	0	0	0	1	0	1	1	1	0	**4**
Root mean sq C3 sb4-8	0	0	1	1	0	0	1	0	1	1	**5**
Root mean sq C3 sb8-12	0	1	0	1	0	1	1	0	1	0	**5**
Root mean sq C3 sb12-16	1	0	0	1	1	1	0	0	1	1	**6**
Root mean sq C3 sb16-20	0	1	1	1	1	1	0	0	1	0	**6**

Root mean sq C3 sb20-24	1	0	1	0	1	1	0	1	1	0	**6**
Root mean sq C3 sb24-28	1	0	0	0	1	1	1	1	1	1	**7**
Root mean sq C3 sb28-32	1	0	0	0	1	1	0	1	1	1	**6**
Root mean sq C4 sb0-4	1	1	0	0	1	0	1	0	0	1	**5**
Root mean sq C4 sb4-8	1	1	0	1	0	0	1	0	0	0	**4**
Root mean sq C4 sb8-12	0	0	1	1	1	1	0	1	1	0	**6**
Root mean sq C4 sb12-16	0	0	0	1	1	1	1	1	0	1	**6**
Root mean sq C4 sb16-20	0	0	1	0	1	1	0	0	0	1	**4**
Root mean sq C4 sb20-24	1	0	1	0	1	0	1	1	0	1	**6**
Root mean sq C4 sb24-28	1	0	0	1	1	0	0	1	0	0	**4**
Root mean sq C4 sb28-32	1	1	0	0	0	1	0	0	0	1	**4**
Variance C3 sb0-4	1	0	0	1	1	1	0	0	0	1	**5**
Variance C3 sb4-8	0	0	0	1	0	0	0	0	1	1	**3**
Variance C3 sb8-12	0	1	1	0	0	0	1	0	1	1	**5**
Variance C3 sb12-16	1	1	1	0	0	0	1	1	0	0	**5**
Variance C3 sb16-20	0	0	1	0	1	0	0	0	1	0	**3**
Variance C3 sb20-24	1	1	1	1	1	0	1	0	0	0	**6**
Variance C3 sb24-28	1	0	1	1	0	0	1	1	1	1	**7**
Variance C3 sb28-32	0	0	1	1	1	1	1	1	1	0	**7**
Variance C4 sb0-4	0	0	0	0	0	0	1	0	0	0	**1**
Variance C4 sb4-8	1	0	0	1	0	0	1	1	0	1	**5**
Variance C4 sb8-12	1	0	0	1	0	1	1	1	1	1	**7**
Variance C4 sb12-16	1	1	1	1	1	1	0	0	0	0	**6**
Variance C4 sb16-20	0	1	1	1	0	1	1	1	1	1	**8**
Variance C4 sb20-24	1	0	0	1	1	1	0	0	0	0	**4**
Variance C4 sb24-28	0	0	0	1	0	0	0	0	0	1	**2**
Variance C4 sb28-32	0	0	1	0	0	0	0	0	0	1	**2**

Table 26. Subject A2

	PNN	**SVM**	**SVM Q**	**LDA**	**RBF**	**KNN**	**KNN M**	**KNN COS**	**KNN COR**	**MLP**	**Total**
Avg abs C3 sb0-4	0	0	1	0	1	0	1	1	1	0	**5**
Avg abs C3 sb4-8	1	1	1	1	0	1	0	1	1	1	**8**
Avg abs C3 sb8-12	0	1	0	0	0	0	0	0	0	0	**1**
Avg abs C3 sb12-16	0	1	1	1	0	0	0	0	1	1	**5**
Avg abs C3 sb16-20	1	1	1	1	1	0	1	1	1	0	**8**
Avg abs C3 sb20-24	0	1	0	0	0	1	0	1	0	0	**3**
Avg abs C3 sb24-28	0	1	1	0	0	0	1	0	0	0	**3**
Avg abs C3 sb28-32	0	1	0	0	1	1	0	1	0	1	**5**
Avg abs C4 sb0-4	1	0	0	0	0	0	1	0	0	0	**2**

Avg abs C4 sb4-8	0	0	0	0	1	1	0	1	1	1	**5**
Avg abs C4 sb8-12	1	0	0	0	1	0	0	1	1	1	**5**
Avg abs C4 sb12-16	1	1	1	0	1	0	0	0	0	1	**5**
Avg abs C4 sb16-20	1	0	1	1	1	0	0	0	1	1	**6**
Avg abs C4 sb20-24	0	1	0	0	1	0	0	0	0	1	**3**
Avg abs C4 sb24-28	0	0	0	0	0	0	1	0	0	1	**2**
Avg abs C4 sb28-32	1	0	0	1	0	0	0	0	1	1	**4**
Avg ampl C3 sb0-4	0	0	0	1	0	1	0	1	0	0	**3**
Avg ampl C3 sb4-8	1	1	1	0	0	1	0	1	0	0	**5**
Avg ampl C3 sb8-12	0	0	1	0	1	1	1	1	0	0	**5**
Avg ampl C3 sb12-16	1	1	1	0	0	1	0	1	1	1	**7**
Avg ampl C3 sb16-20	0	1	1	1	0	1	0	1	1	0	**6**
Avg ampl C3 sb20-24	0	1	1	0	0	1	0	1	1	1	**6**
Avg ampl C3 sb24-28	0	1	0	0	0	0	1	1	0	1	**4**
Avg ampl C3 sb28-32	0	0	1	0	1	1	0	1	1	1	**6**
Avg ampl C4 sb0-4	0	1	0	1	1	0	1	0	1	0	**5**
Avg ampl C4 sb4-8	0	1	0	0	1	1	0	0	0	1	**4**
Avg ampl C4 sb8-12	1	0	1	0	1	0	1	0	0	1	**5**
Avg ampl C4 sb12-16	0	0	1	1	1	0	0	0	1	1	**5**
Avg ampl C4 sb16-20	1	0	1	1	1	0	1	1	1	1	**8**
Avg ampl C4 sb20-24	1	1	0	0	1	1	0	1	0	1	**6**
Avg ampl C4 sb24-28	1	1	0	0	0	1	1	0	0	0	**4**
Avg ampl C4 sb28-32	1	0	1	1	1	0	1	1	0	0	**6**
Std dev C3 sb0-4	1	1	0	1	0	0	0	1	0	0	**4**
Std dev C3 sb4-8	0	0	0	1	1	1	1	1	0	1	**6**
Std dev C3 sb8-12	1	0	1	0	0	1	1	1	0	1	**6**
Std dev C3 sb12-16	1	0	0	1	0	1	0	1	1	0	**5**
Std dev C3 sb16-20	0	1	0	0	0	0	0	0	0	0	**1**
Std dev C3 sb20-24	1	0	1	0	1	1	1	1	0	1	**7**
Std dev C3 sb24-28	0	1	0	0	0	0	1	0	0	0	**2**
Std dev C3 sb28-32	0	1	0	1	0	1	0	1	0	0	**4**
Std dev C4 sb0-4	1	0	0	0	1	1	0	1	0	0	**4**
Std dev C4 sb4-8	0	0	0	1	0	0	0	0	1	1	**3**
Std dev C4 sb8-12	0	0	1	0	0	0	0	0	1	0	**2**
Std dev C4 sb12-16	1	1	0	0	1	0	1	0	1	0	**5**
Std dev C4 sb16-20	1	1	1	1	1	0	1	0	0	0	**6**
Std dev C4 sb20-24	0	0	0	0	0	1	0	0	1	0	**2**
Std dev C4 sb24-28	0	1	0	0	0	0	1	0	1	1	**4**
Std dev C4 sb28-32	0	0	0	0	1	0	1	0	1	1	**4**
Energy C3 sb0-4	0	1	0	0	0	1	1	0	1	1	**5**
Energy C3 sb4-8	1	0	1	1	0	1	0	0	0	1	**5**

Energy C3 sb8-12	0	1	1	0	1	1	0	1	0	1	**6**
Energy C3 sb12-16	1	0	0	1	1	0	1	1	1	1	**7**
Energy C3 sb16-20	1	0	0	0	0	0	1	0	0	1	**3**
Energy C3 sb20-24	1	1	0	0	1	1	0	0	1	0	**5**
Energy C3 sb24-28	1	1	0	1	1	1	1	1	0	1	**8**
Energy C3 sb28-32	1	1	0	0	1	1	0	1	1	1	**7**
Energy C4 sb0-4	1	0	0	1	0	0	0	1	1	1	**5**
Energy C4 sb4-8	0	0	0	1	0	1	1	0	0	0	**3**
Energy C4 sb8-12	0	0	0	0	0	1	1	0	0	1	**3**
Energy C4 sb12-16	0	1	1	0	0	1	0	1	1	1	**6**
Energy C4 sb16-20	0	0	0	0	1	1	1	0	0	0	**3**
Energy C4 sb20-24	0	0	0	0	1	0	0	0	1	0	**2**
Energy C4 sb24-28	1	1	0	1	0	0	0	0	1	1	**5**
Energy C4 sb28-32	0	1	1	1	1	1	1	1	0	1	**8**
Rt En C3 sb0-4 and sb4-8	1	0	0	1	0	1	0	1	0	0	**4**
Rt En C3 sb4-8 and sb8-12	0	1	1	1	1	1	0	1	1	0	**7**
Rt En C3 sb8-12 and sb12-16	1	0	1	0	0	0	1	1	1	1	**6**
Rt En C3 sb12-16 and sb16-20	1	0	1	1	0	0	0	0	1	0	**4**
Rt En C3 sb16-20 and sb20-24	0	1	0	0	0	0	0	0	0	0	**1**
Rt En C3 sb20-24 and sb24-28	1	0	1	0	1	1	0	0	1	0	**5**
Rt En C3 sb24-28 and sb28-32	1	1	1	1	0	0	1	1	1	1	**8**
Rt En C4 sb0-4 and sb4-8	0	0	0	1	0	1	0	0	0	1	**3**
Rt En C4 sb4-8 and sb8-12	0	1	0	1	0	0	0	1	0	1	**4**
Rt En C4 sb8-12 and sb12-16	1	1	0	1	0	1	0	0	0	0	**4**
Rt En C4 sb12-16 and sb16-20	1	1	1	0	0	0	0	1	0	1	**5**
Rt En C4 sb16-20 and sb20-24	0	1	0	0	0	1	1	1	0	1	**5**
Rt En C4 sb20-24 and sb24-28	0	0	0	1	1	1	0	0	1	0	**4**
Rt En C4 sb24-28 and sb28-32	0	1	0	1	1	1	1	0	1	0	**6**
Entropy C3 sb0-4	1	1	0	1	0	1	0	1	0	0	**5**
Entropy C3 sb4-8	0	1	1	1	0	1	1	0	1	1	**7**
Entropy C3 sb8-12	0	0	1	1	0	1	0	1	0	0	**4**
Entropy C3 sb12-16	1	0	1	0	0	1	0	1	1	0	**5**
Entropy C3 sb16-20	0	1	0	0	0	0	0	1	1	0	**3**
Entropy C3 sb20-24	1	0	1	0	1	1	1	0	0	0	**5**
Entropy C3 sb24-28	1	1	0	1	1	1	1	1	1	0	**8**
Entropy C3 sb28-32	1	1	1	0	0	1	1	0	0	0	**5**
Entropy C4 sb0-4	1	0	1	0	1	0	0	1	0	0	**4**
Entropy C4 sb4-8	0	0	0	0	0	1	0	0	0	1	**2**
Entropy C4 sb8-12	1	0	1	0	0	1	0	1	1	0	**5**
Entropy C4 sb12-16	0	1	0	1	1	0	0	1	0	1	**5**
Entropy C4 sb16-20	0	0	1	0	1	0	1	0	1	0	**4**

Entropy C4 sb20-24	1	0	1	0	1	1	1	0	1	0	**6**
Entropy C4 sb24-28	0	0	0	1	0	1	1	0	0	0	**3**
Entropy C4 sb28-32	1	0	0	0	1	0	1	1	0	0	**4**
PLV C3 and C4 sb0-4	1	1	1	1	0	0	0	1	0	1	**6**
PLV C3 and C4 sb4-8	1	1	0	0	0	0	0	0	0	1	**3**
PLV C3 and C4 sb8-12	0	1	0	1	1	1	1	1	0	0	**6**
PLV C3 and C4 sb12-16	0	0	0	1	0	0	1	0	1	0	**3**
PLV C3 and C4 sb16-20	0	1	0	1	1	1	1	0	0	0	**5**
PLV C3 and C4 sb20-24	0	0	1	1	0	0	0	1	0	1	**4**
PLV C3 and C4 sb24-28	1	0	0	1	0	0	1	0	0	0	**3**
PLV C3 and C4 sb28-32	0	0	0	0	1	0	0	1	0	1	**3**
Root mean sq C3 sb0-4	0	0	0	0	1	1	0	1	1	1	**5**
Root mean sq C3 sb4-8	0	0	1	0	1	1	0	1	0	0	**4**
Root mean sq C3 sb8-12	0	1	0	1	0	0	0	1	0	1	**4**
Root mean sq C3 sb12-16	1	1	0	0	0	0	1	1	1	0	**5**
Root mean sq C3 sb16-20	0	0	1	0	0	1	1	1	0	0	**4**
Root mean sq C3 sb20-24	1	0	1	0	1	0	1	0	0	0	**4**
Root mean sq C3 sb24-28	1	1	0	0	0	1	0	0	0	1	**4**
Root mean sq C3 sb28-32	1	1	0	0	1	1	0	1	1	1	**7**
Root mean sq C4 sb0-4	1	1	0	0	0	1	0	0	0	1	**4**
Root mean sq C4 sb4-8	1	1	0	0	1	0	1	1	1	0	**6**
Root mean sq C4 sb8-12	0	0	0	0	1	0	0	0	1	1	**3**
Root mean sq C4 sb12-16	0	0	0	0	0	0	0	1	0	0	**1**
Root mean sq C4 sb16-20	0	1	0	1	0	0	0	0	1	0	**3**
Root mean sq C4 sb20-24	1	1	1	0	0	1	0	0	1	1	**6**
Root mean sq C4 sb24-28	1	0	1	1	1	0	1	0	1	1	**7**
Root mean sq C4 sb28-32	1	1	1	0	0	1	0	0	1	1	**6**
Variance C3 sb0-4	1	1	0	0	1	0	1	0	0	0	**4**
Variance C3 sb4-8	0	1	1	1	1	0	0	1	1	1	**7**
Variance C3 sb8-12	0	0	1	0	0	1	0	1	0	0	**3**
Variance C3 sb12-16	1	1	0	1	0	0	0	0	1	0	**4**
Variance C3 sb16-20	0	1	0	1	1	0	0	1	0	1	**5**
Variance C3 sb20-24	1	1	1	0	0	0	1	1	1	1	**7**
Variance C3 sb24-28	1	0	0	0	0	1	1	1	1	1	**6**
Variance C3 sb28-32	0	0	1	1	0	0	1	1	0	1	**5**
Variance C4 sb0-4	0	0	0	0	0	1	0	0	1	0	**2**
Variance C4 sb4-8	1	0	0	0	1	0	1	0	1	1	**5**
Variance C4 sb8-12	1	0	0	0	1	0	1	1	0	1	**5**
Variance C4 sb12-16	1	0	0	1	1	1	0	0	0	0	**4**
Variance C4 sb16-20	0	1	1	0	1	0	0	1	0	0	**4**
Variance C4 sb20-24	1	0	0	0	1	0	1	0	1	0	**4**

Variance C4 sb24-28	0	0	0	1	0	0	0	0	0	1	**2**
Variance C4 sb28-32	0	1	1	1	0	0	1	1	0	0	**5**

Appendix 4. Confusion Matrix for each Ensemble fusion method

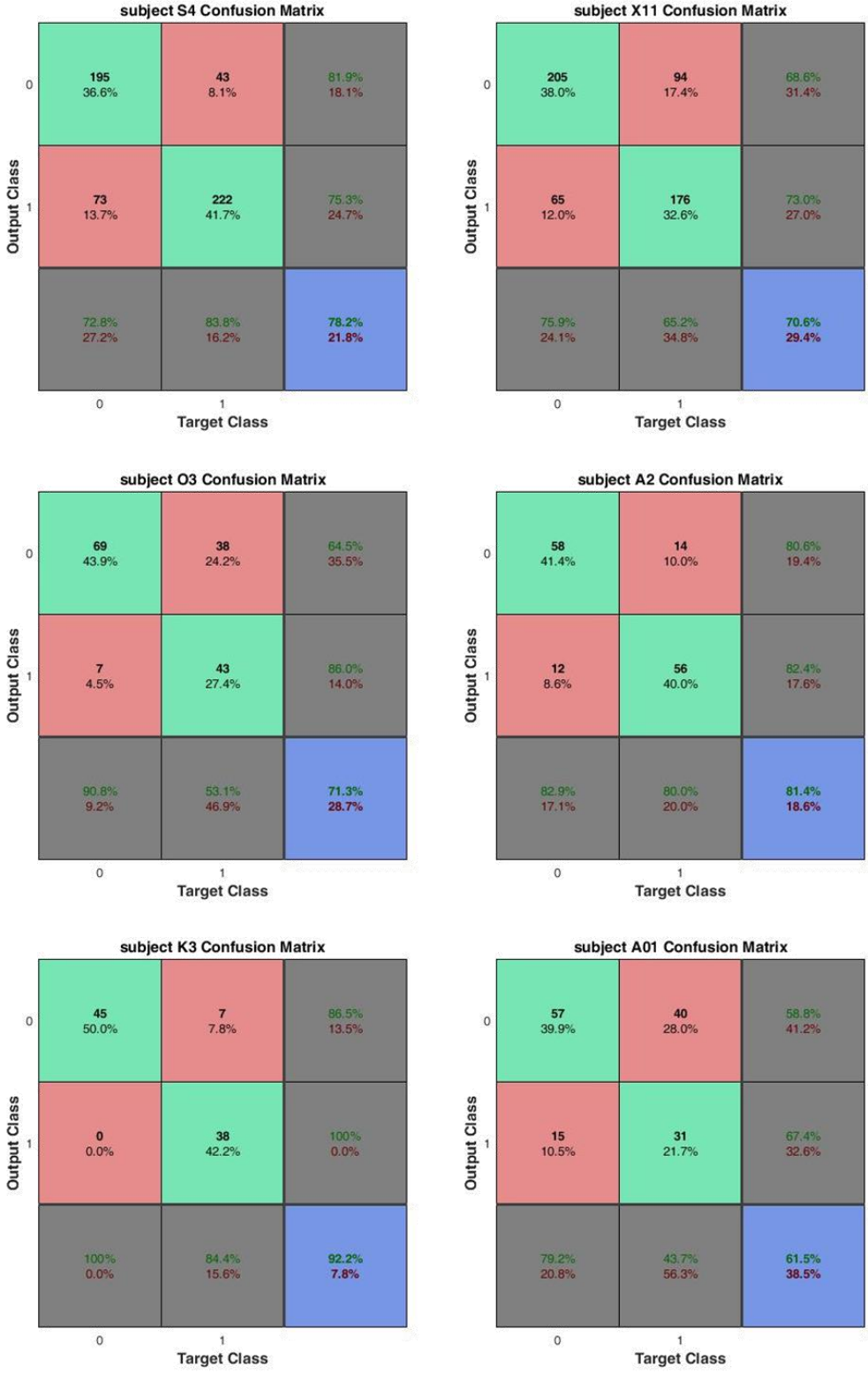

Figure 55. Naïve Bayes

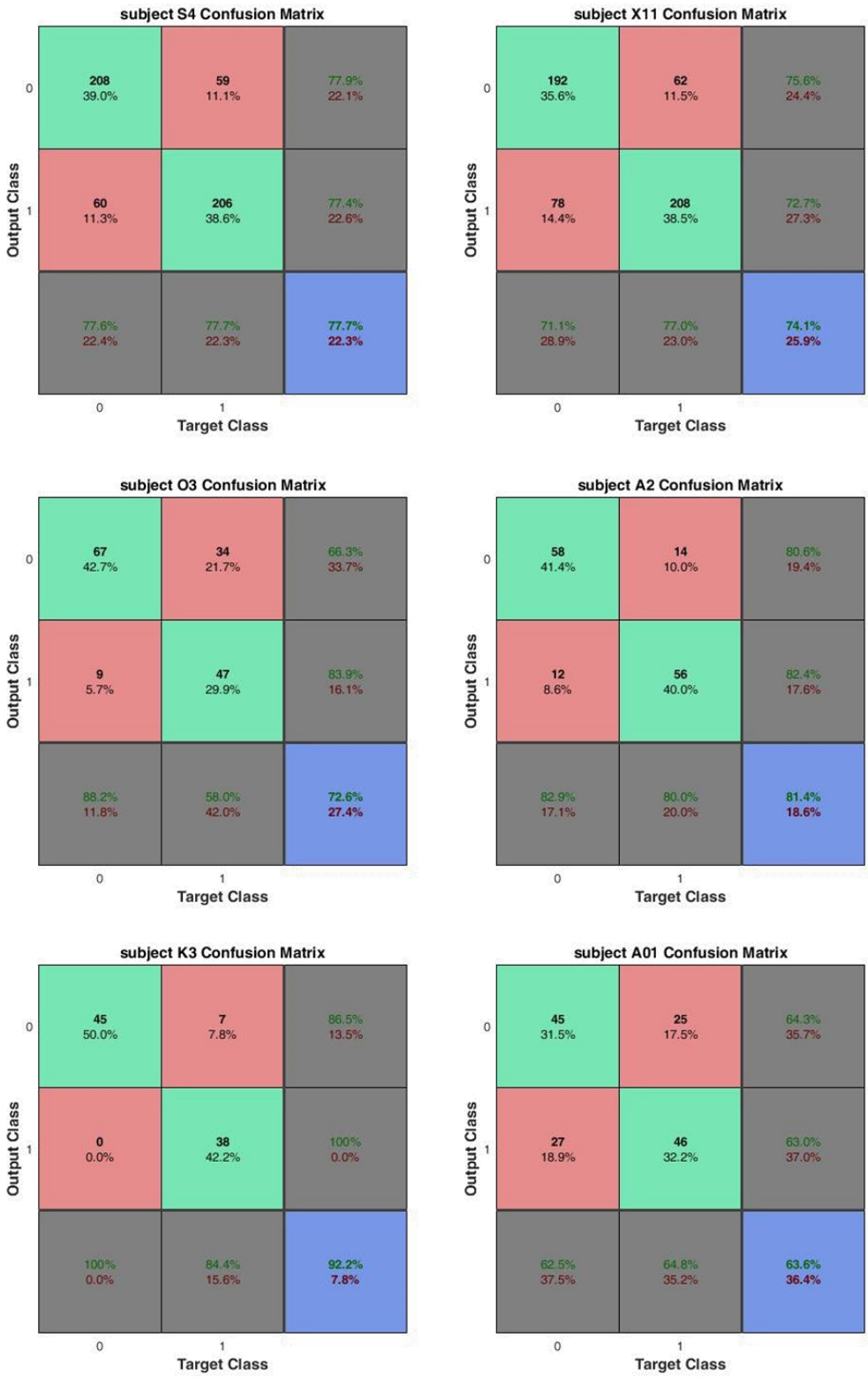

Figure 56. Majority Voting without Feature Selection

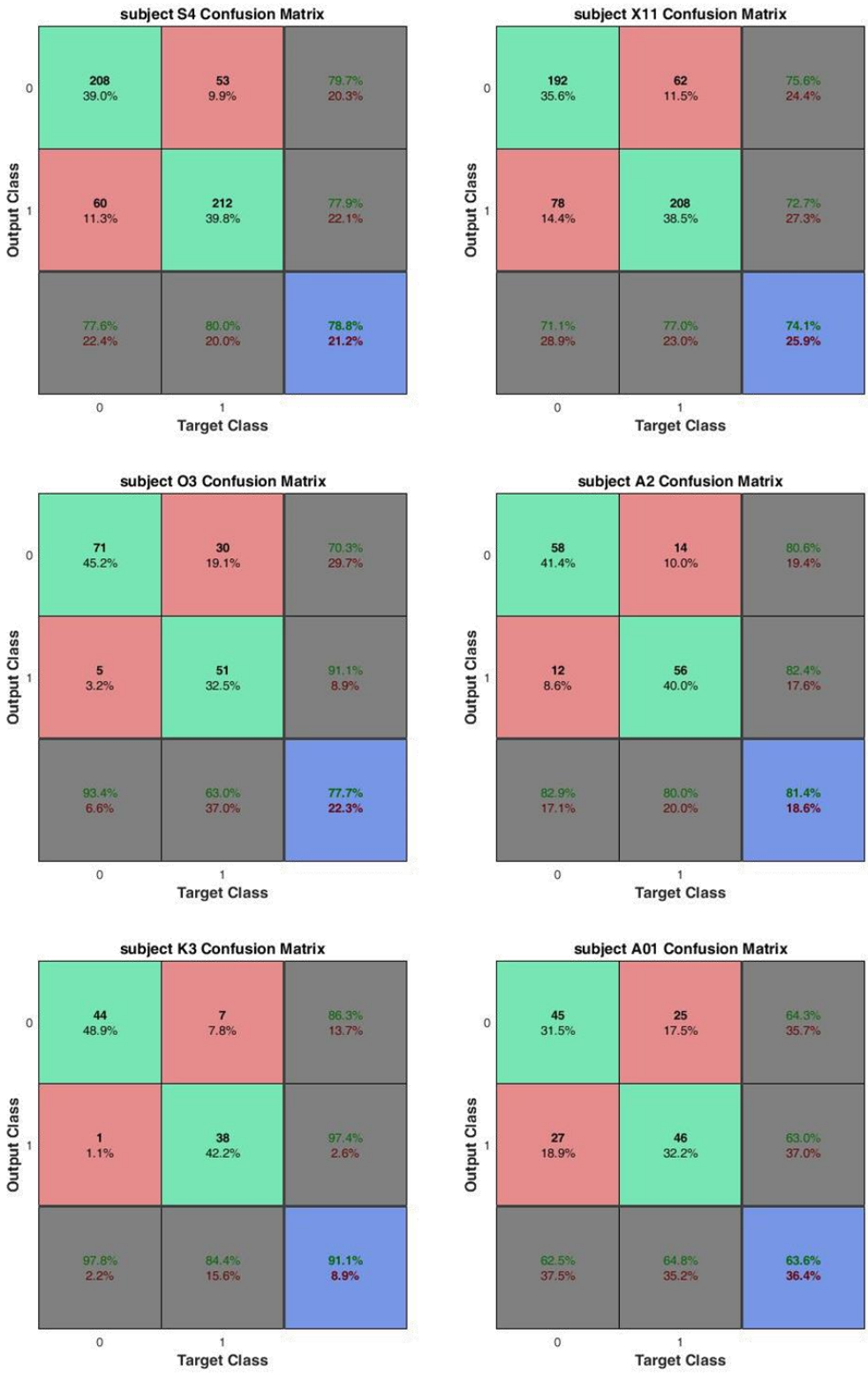

Figure 57. Weighted Voting without Feature Selection

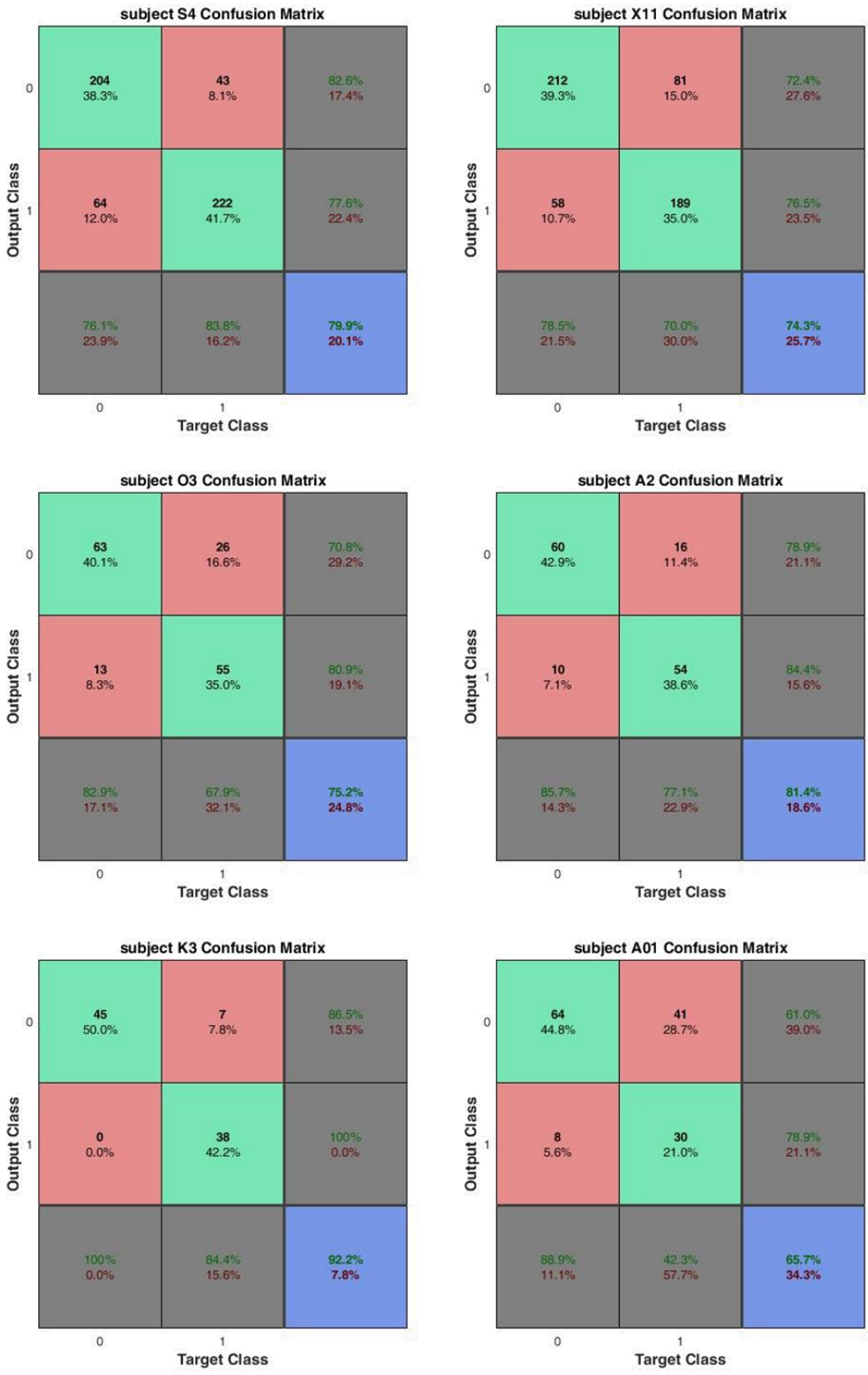

Figure 58. Majority Voting with Feature Selection for an ensemble composed of SVM with linear kernel, LDA, KNN, PNN, RBF and MLP

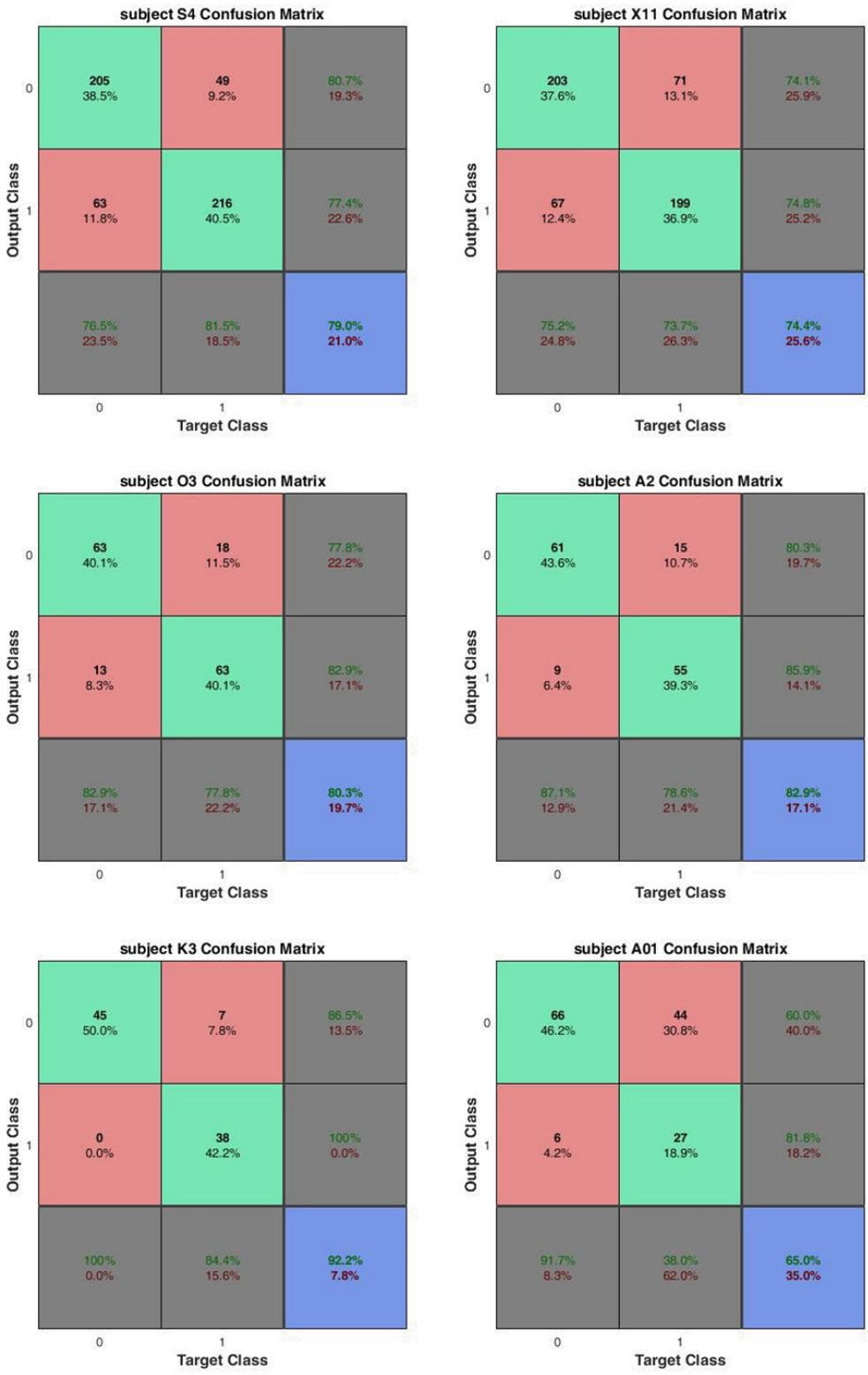

Figure 59. Majority Voting with Feature Selection for an ensemble composed of SVM, LDA, SVM with quadratic kernel, and MLP

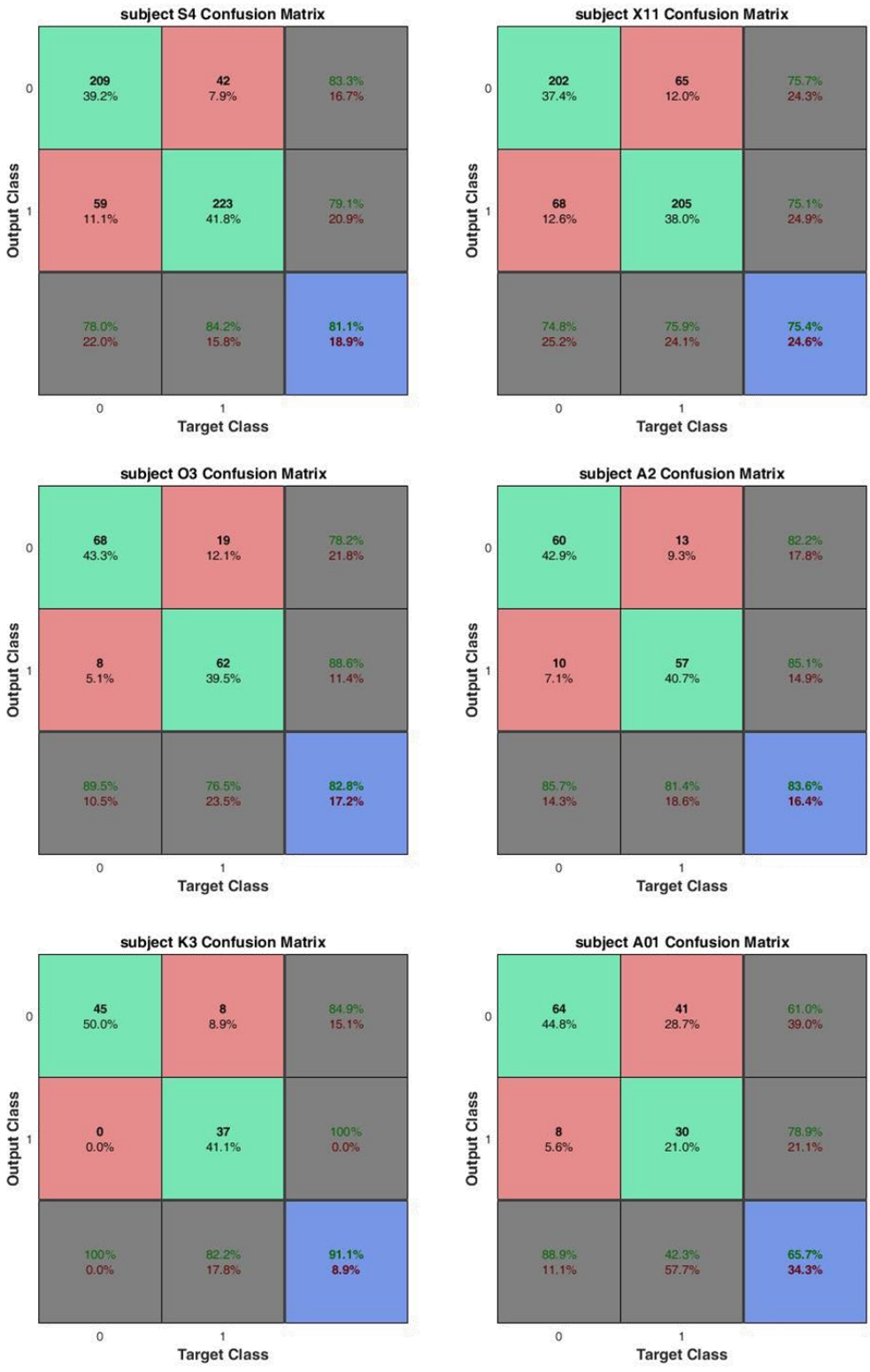

Figure 60. Majority Voting with Feature Selection for an ensemble composed of all the classifiers

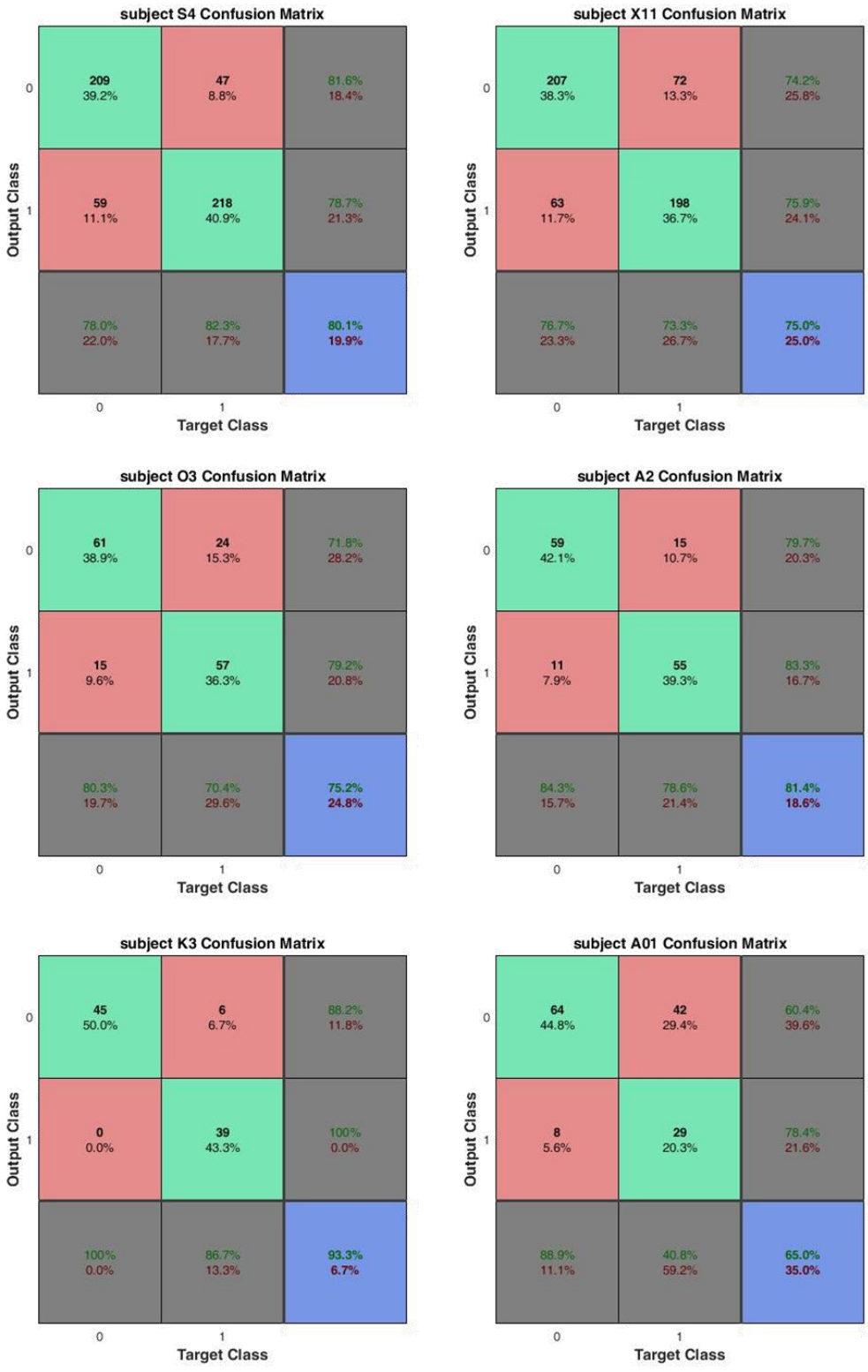

Figure 61. Weighted Voting with Feature Selection for an ensemble composed of SVM with linear kernel, LDA, KNN, PNN, RBF and MLP

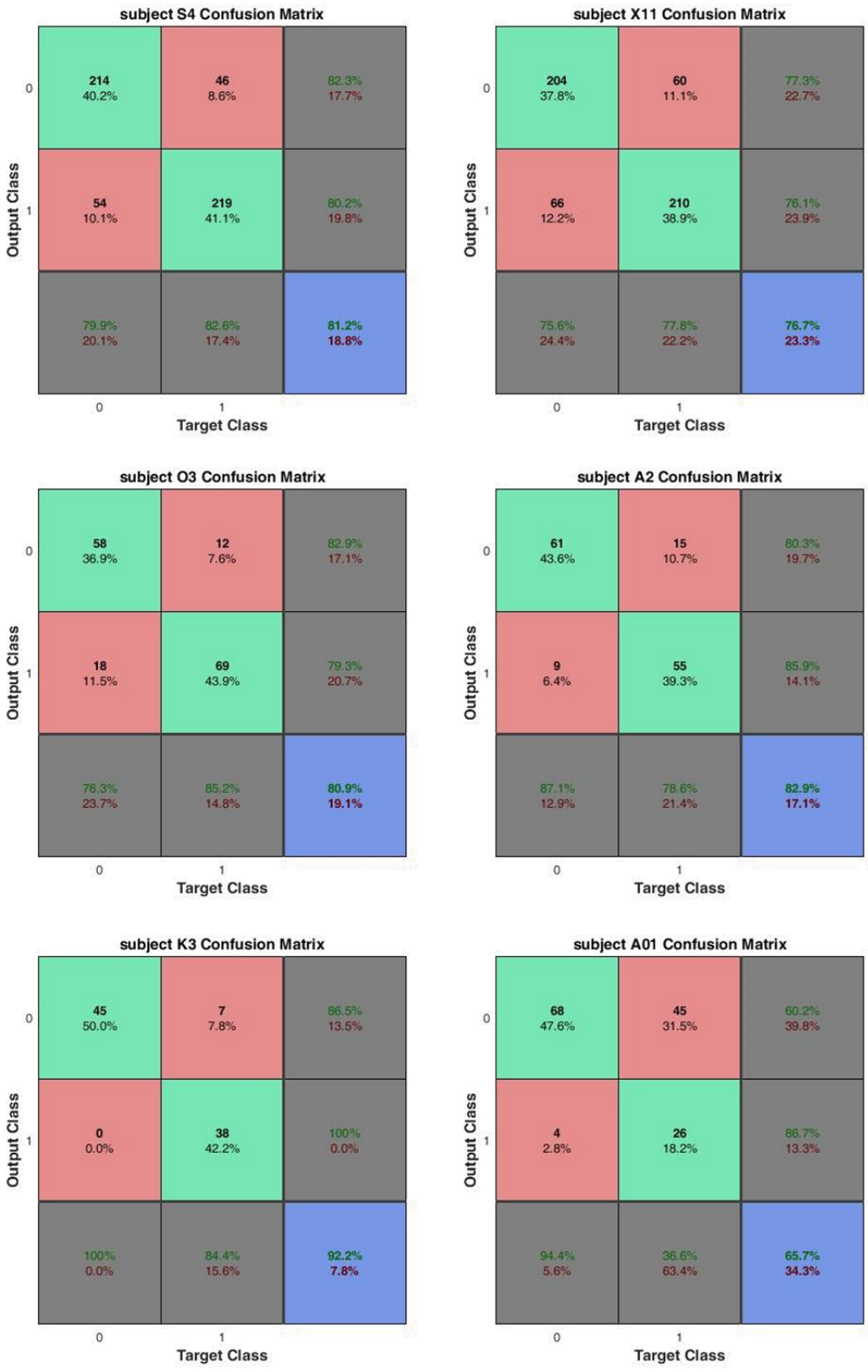

Figure 62. Weighted Voting with Feature Selection for an ensemble composed of SVM, LDA, SVM with quadratic kernel, and MLP

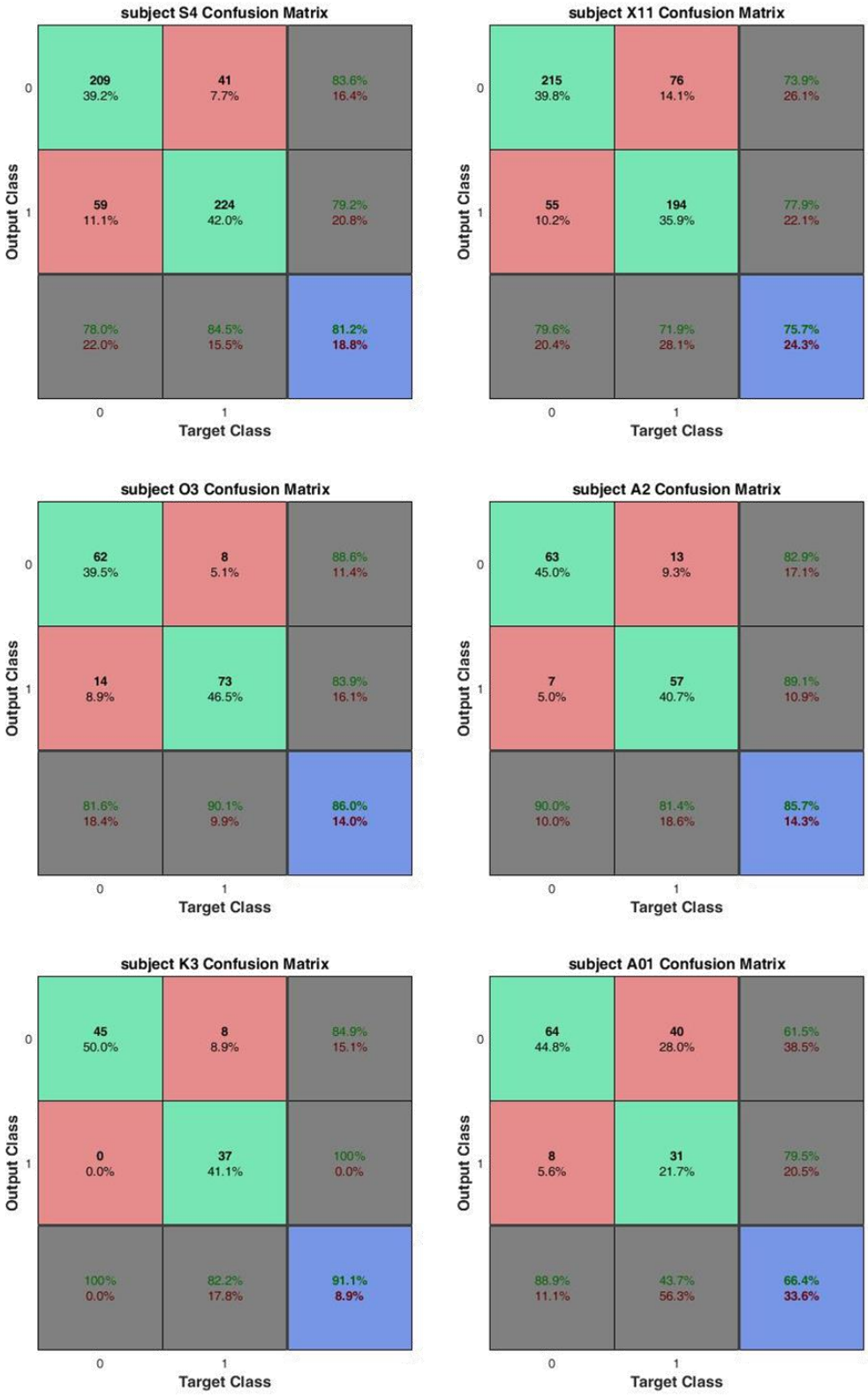

Figure 63. Weighted Voting with Feature Selection for an ensemble composed of all the classifiers

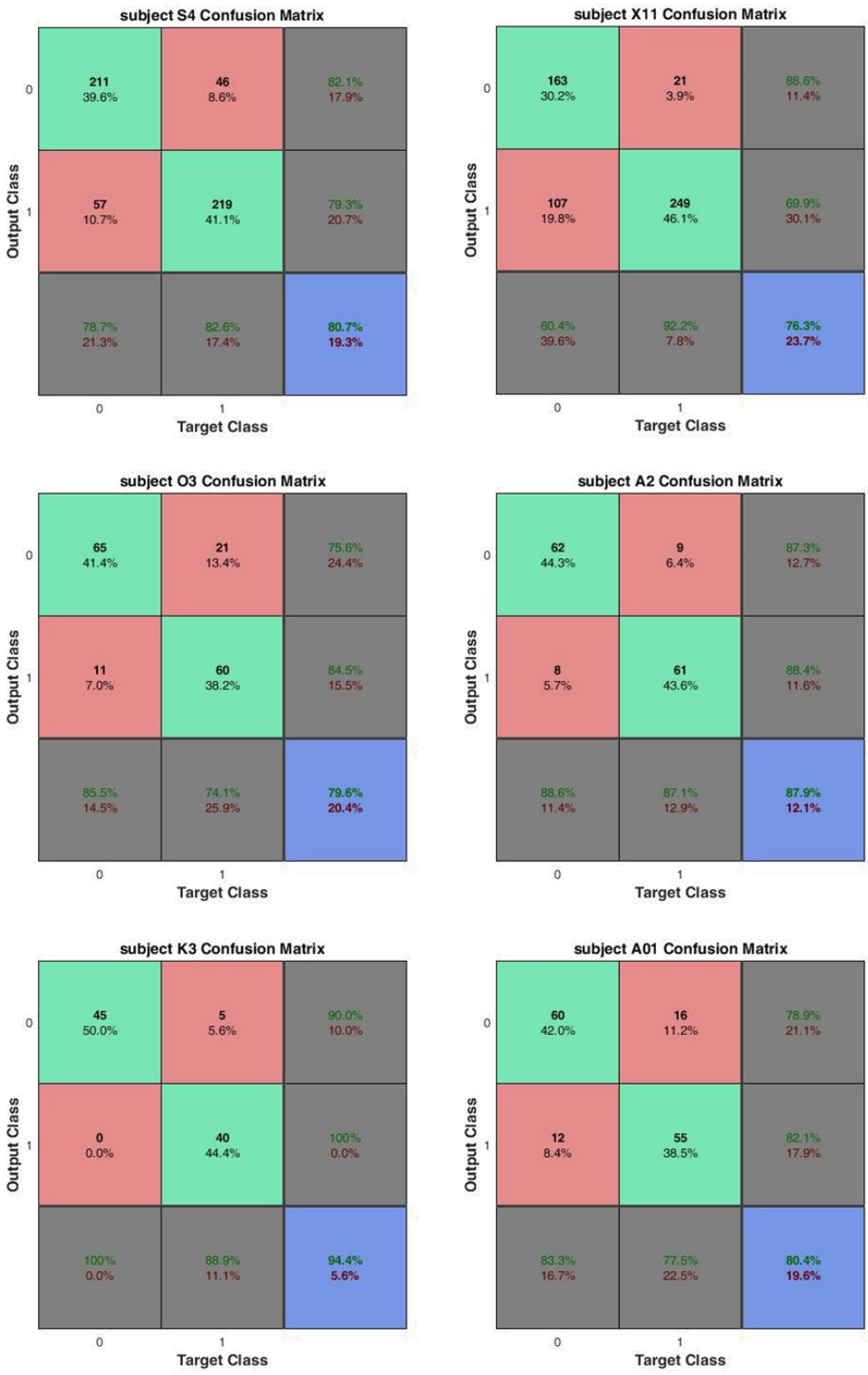

Figure 64. GA for selection of classifiers

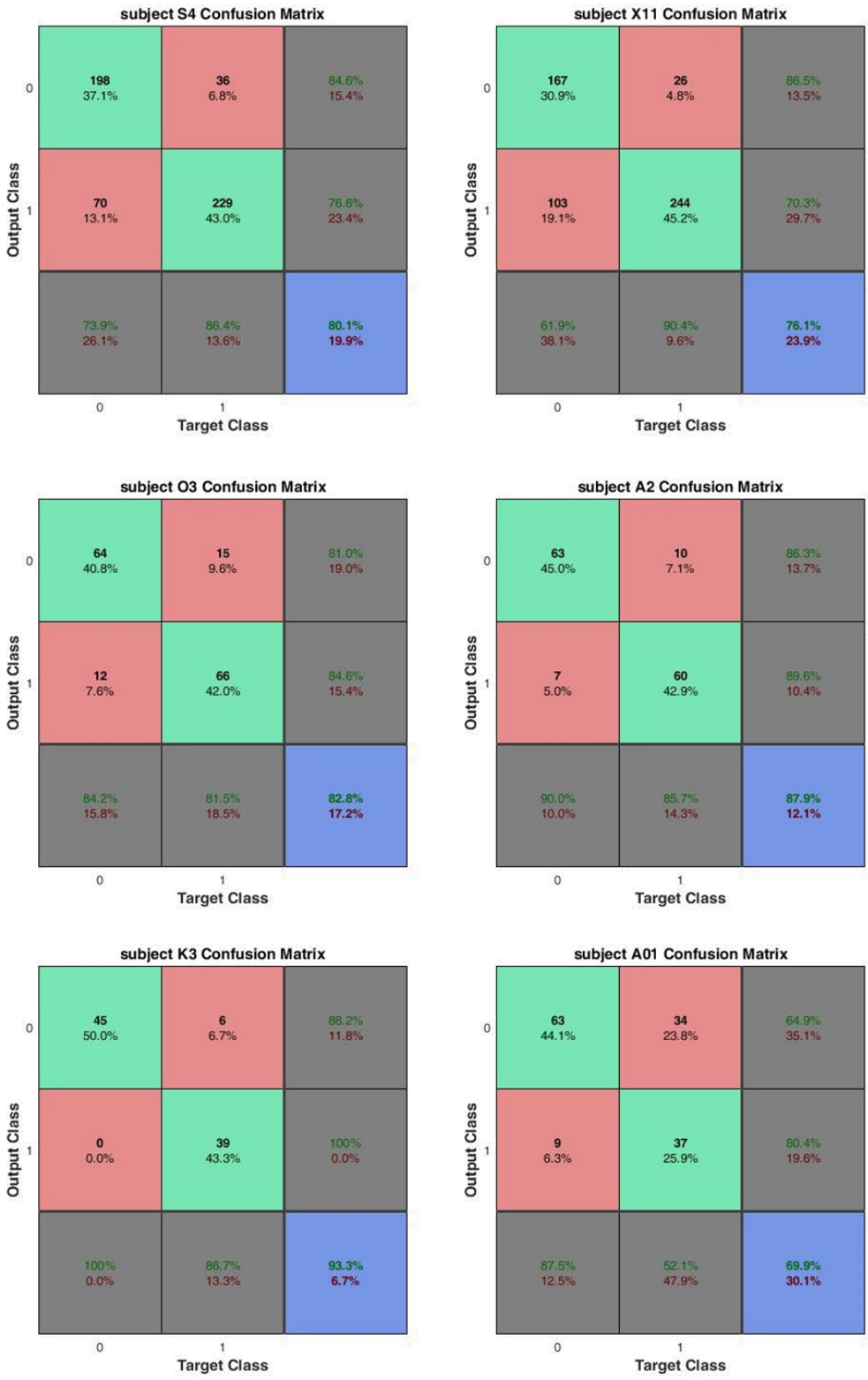

Figure 65. GA for weights definition between 0 and 1

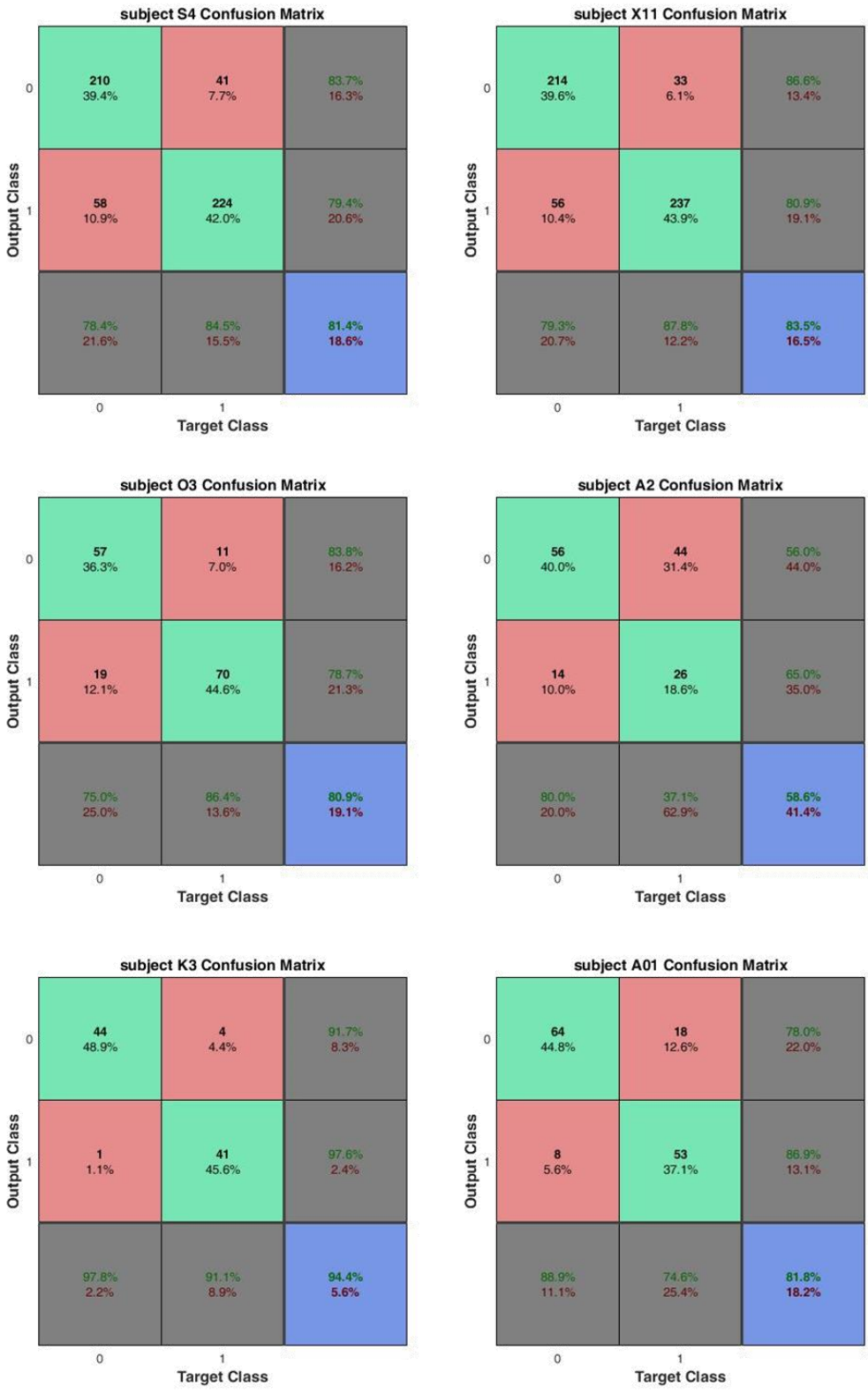

Figure 66. GA for weights definition between 0 and 1summing 1

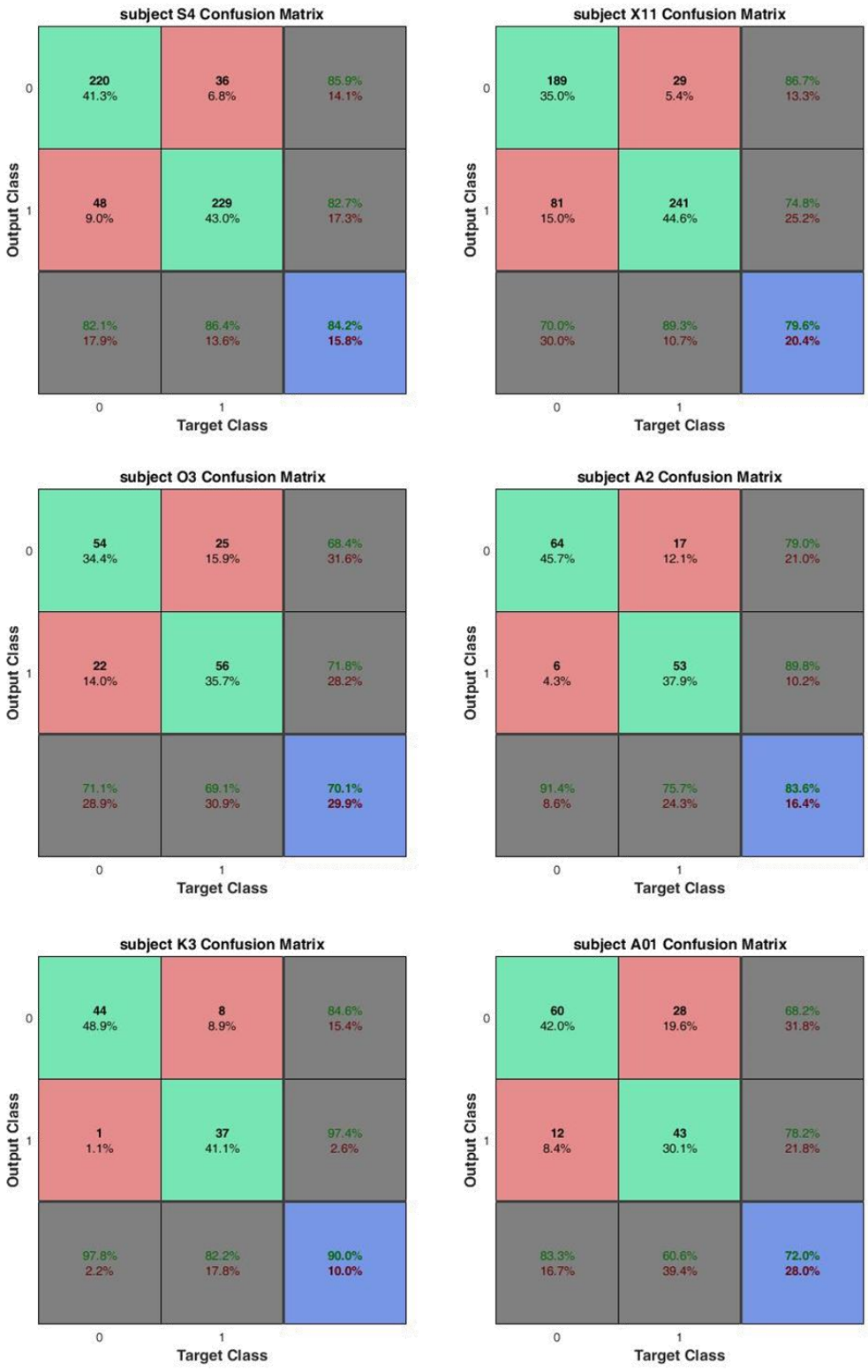

Figure 67. GA for weights definition between -1 and 1summing 1

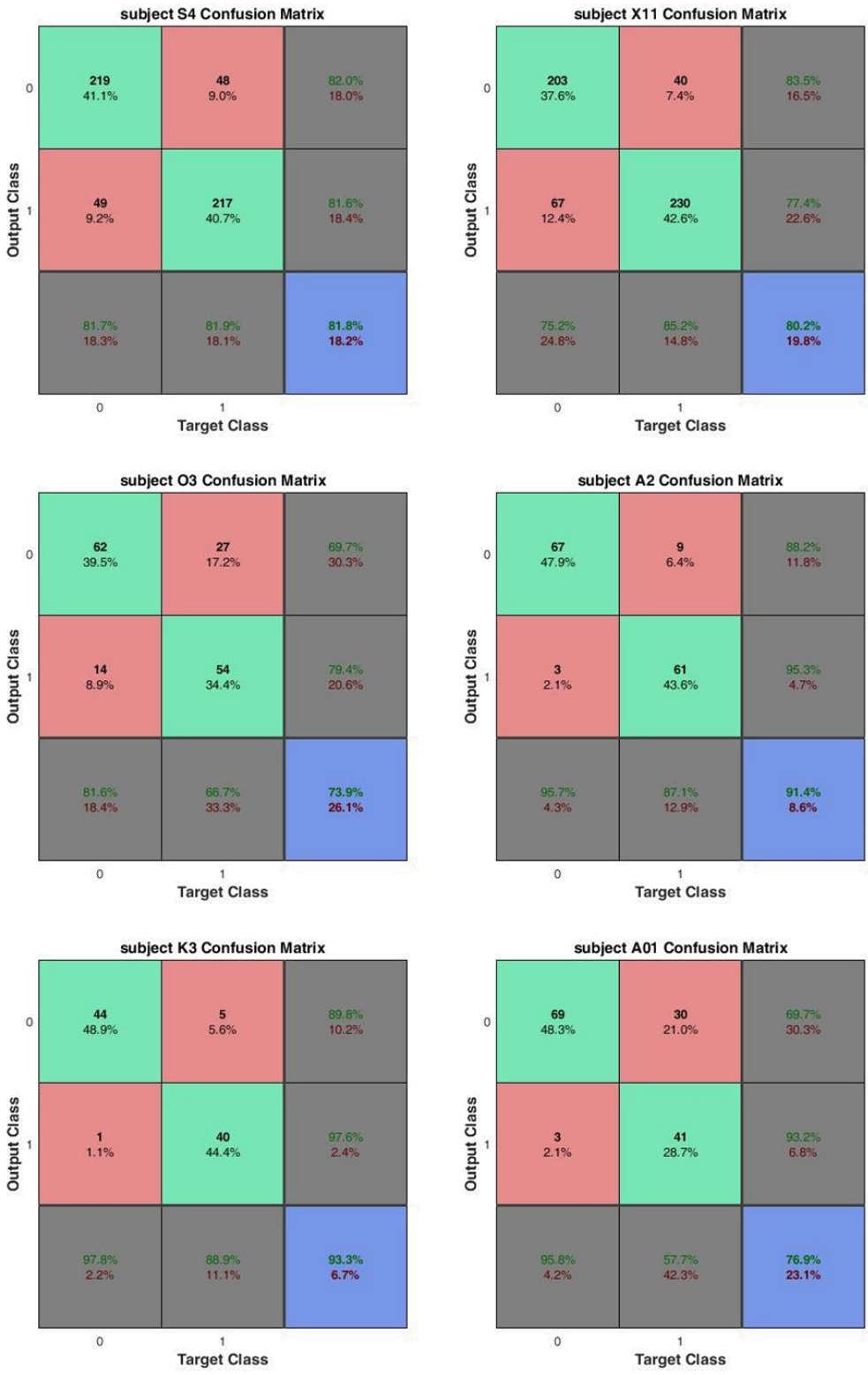

Figure 68. GAs combined

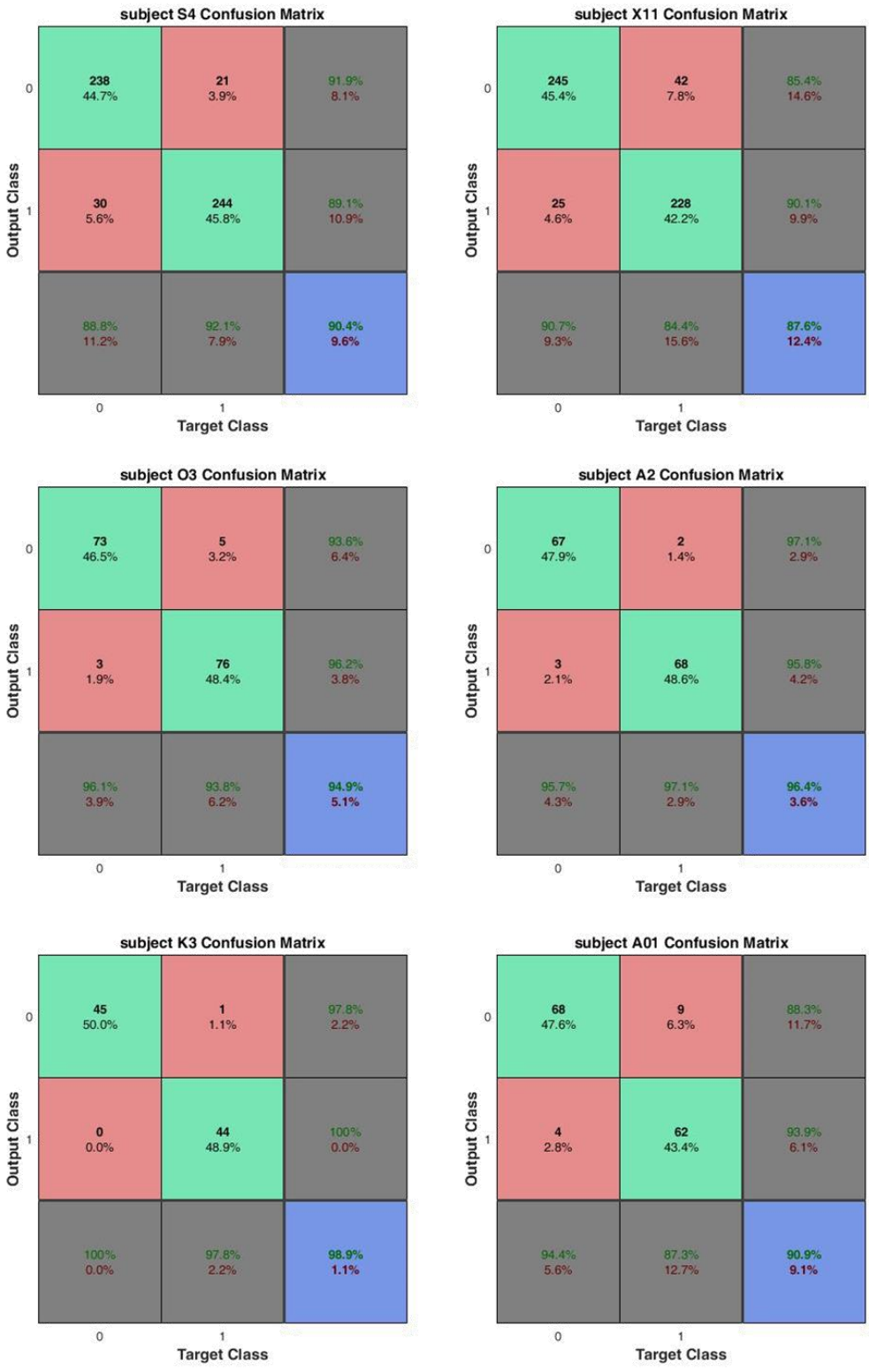

Figure 69. MLP as meta-classifier