Deep learning methods for motor imagery detection from raw EEG : applications to brain-computer interfaces

Oleksii Avilov

▶ **To cite this version:**

Oleksii Avilov. Deep learning methods for motor imagery detection from raw EEG : applications to brain-computer interfaces. Computer Science [cs]. Université de Lorraine; Institut polytechnique de Kiev (Ukraine), 2021. English. NNT : 2021LORR0032 . tel-03229010

HAL Id: tel-03229010

https://hal.univ-lorraine.fr/tel-03229010v1

Submitted on 18 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Deep learning methods for motor imagery detection from raw EEG: applications to brain-computer interfaces

THÈSE

présentée et soutenue publiquement le Février 2021

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention informatique)

par

Oleksii Avilov

Composition du jury

Directeurs :	Patrick Hénaff, Professeur à l'Université de Lorraine Anton Popov, Associate professor à l'Institut Polytechnique de Kiev Igor Sikorsky
Co-encadrant :	Laurent Bougrain, Maître de conférences à l'Université de Lorraine
Rapporteurs :	Tetiana Aksenova, Chargé derecherche, CEA Leti MINATEC Sylvain Chevallier, Maître de conférences-HDR, Université de Versailles St Quentin
Examinateurs :	Ye-Qiong Song, Professeur à l'Université de Lorraine Vladimir Timofeyev, Professeur à l'Institut Polytechnique de Kiev Igor Sikorsky Kateryna Ivanko, Associate professor à l'Institut Polytechnique de Kiev Igor Sikorsky

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Mis en page avec la classe thesul.

I would like to dedicate this thesis to my grandfather Yakiv and all the people, who, unfortunately, died because of the COVID-19 global pandemic. My grandfather passed away during the last weeks of dissertation preparation. He was a man of strong will and spirit, who survived the Holodomor of 1932-33, the Second World War, and recovered from three strokes. He lived a wonderful 91 years and continued to smile even in the most difficult times. Your smile will always cheer me up and stay in my heart forever.

Remerciements

Carrying out the required work and then writing this thesis was, undoubtedly, the toughest task I have undertaken. However, one of the pleasures of having completed the thesis is looking back at everyone who has helped me over the past four, eleven, and twenty-eight years.

I would like to begin by thanking my four supervisors: Anton Popov, Laurent Bougrain, Patrick Hénaff, Vladimir Timofeyev, whose expertise was invaluable. Your insightful feedback pushed me to sharpen my thinking and brought me and my work to a higher level.

I would particularly like to single out my supervisor, comrade, and life mentor Anton Popov. Anton, I want to thank you for your patient support, for the inspiration, for the desire and interest in learning and research that you instilled in me, and for all of the opportunities I was given during all these years of study.

I would like to thank Sebastien Rimbert and Pierre Riff for the help during the experiments. I also wish to give thanks to all the participants.

I am also infinitely thankful to my Mum Liudmyla for her love, generosity, and support.

I would also like to thank my family: my brother Oleksandr, father Oleksandr, grandfather Yakiv, and grandmother Nina for their belief in me and constant support. It was extremely important for me to feel your support.

Also, I would like to thank my teachers from the high school Tamara Lenova, Valentyna Struk, and Lesia Zapadlovska who instilled in me curiosity and love to learn.

I want to thank the Neurorhythms members for speaking French all the time during lunch and encouraging and pushing me to learn French more.

I would like to thank Francesco Giovannini for your attempts to dissuade me from doing a Ph.D. In the last year, I finally understood what you mean. But have you see, I did it!

I would like to thank Dominique Martinez for all our talks about startups and business plans.

I would like to thank Rahaf Al-Chwa and Nathalie Azevedo-Carvalho for your help and for all our interesting conversations in the office.

I would like to thank my French friend Valentin Calisti for your help, support, and time we spent together in Ukraine and France. Thanks to you, I fell in love with Marseille and road bikes. And thank you for explaining Riemannian Geometry and teaching me how to flirt in French. Now I'm pro in both!

I would like to thank Annik and François Grande and the people from ARIV Association, who became my French family, always were open and kind to me and helped me a lot during my stay in France. Thanks to Annik and François I discovered French culture and a completely different way of life. It made me rethink many of my life's values. And I really appreciate the time and moments we spent together.

But not only French families. During my Ph.D. and four stays in Nancy, I met and discovered many very interesting people, some of them became my close friends.

I would like to thank Valeria Razinina, Daryna Malashenko, and Ivan Kotiuchyi, we actually lived together in the first year of my Ph.D. I would also like to thank Max Fedorchuck, Yuliia Kunyk, Olha Kurhanska, Juliia Miliar, Nikita Vasilenko, Ksenia Yankina, Anastasya Kharchenko, Anna Liednikova, Luigi Penco, Erika Bartolomei, Houda Ennaim, Joanna Rogińska, and Taisiia Sikolenko, all of you supported me all the way to the goal. How many unforgettable moments we have experienced together. These 4 years would have been much harder without you.

I would like to thank my friends Dmytro Domashenko, Kirill Teplyakov, and Anna Turchaninova for your support. You were so brave to travel to Nancy for visiting me.

I want to thank my friend and colleague Dr. Daryna Panicheva, who provided stimulating discussions till 8 a.m. as well as happy entertainments and sports activities to rest my mind outside of the research.

Also, I want to say thanks to everyone who was around, supported and helped me during this long 4 years. My Ukrainian and French friends, thank you for your mental and motivational support! I am very grateful to many of you! I want to say special thanks to my friend Denys Ladchenko, for supporting my family when they needed help.

Also, I would like to thank Taylor Swift for two wonderful albums in 2020. Most of this thesis was written while I was listening to these albums.

The work reported in this thesis would not have been possible without the financial support of the LORIA laboratory, Erasmus, and the French Embassy to Ukraine. Experiments presented in this paper were carried out using

the Grid'5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER, and several Universities as well as other organizations.

Résumé

Cette thèse présente trois contributions pour améliorer la reconnaissance d'imaginations motrices utilisées par de nombreuses interfaces cerveau-ordinateur (BCI) comme moyen d'interaction. Tout d'abord, nous proposons d'estimer la qualité des images motrices en détectant des valeurs aberrantes et de les supprimer avant apprentissage. Ensuite, nous étudions la sélection des caractéristiques pour sept imaginations de mouvements. Enfin, nous présentons une architecture d'apprentissage profond reprenant les principes du réseau EEGnet applicable directement sur des signaux électroencéphalographiques simplement filtrés et adaptés au nombre d'électrodes. Nous montrons en particulier ses bénéfices pour l'amélioration de la détection des réveils peropératoires et d'autres applications.

Nous introduisons d'abord les concepts de base des interfaces cerveau-ordinateur basées sur l'imagerie motrice. Nous présentons les bases neurophysiologiques des rythmes cérébraux et de leurs localisations.

Nous abordons les méthodes de prétraitement, d'extraction de caractéristiques, de sélection de caractéristiques, de détection des valeurs aberrantes et de classification ainsi que des méthodes d'apprentissage automatique fréquemment utilisées dans les BCI basées sur l'IM pour la tâche de classification, telles que l'analyse discriminante linéaire, les machines à vaste marge et les algorithmes de régression logistique.

Nous introduison le concept de réseaux de neurones artificiels, leurs architectures et algorithmes d'optimisation.

Enfin, nous présentons les architectures de réseaux de neurones de convolutions (CNN) et leurs adaptations pour la classification des signaux EEG bruts tels que Deep Convolutional Network, Shallow Convolutional Network et EEGNet.

Pour implémenter et valider les méthodes proposées, nous avons présenté trois jeux de données différents. Le dernier jeu de données se compose de conditions qui nous permettent de comparer deux tâches de discrimination : l'imagination motrice et l'état de repos d'une part et la réponse à une stimulation du nerf médian et la réponse à une stimulation du nerf médian pendant une imagination motrice d'autre part.

Les résultats sur l'estimation de la qualité de l'imagerie motrice à l'aide de la détection des valeurs aberrantes a montré qu'elle permet d'augmenter considérablement la précision pour un groupe de sujets.

Nous avons également présenté une approche pour la classification de l'imagerie motrice de mouvement combinée qui comprend la sélection et la classification des caractéristiques à l'aide d'une architecture avec trois classificateurs binaires en parallèle. Nous avons montré qu'en utilisant cette approche, il est possible d'atteindre le même niveau de précision qu'en utilisant l'approche multi-classes standard, mais avec un temps réduit pour la calibration de la BCI.

Enfin, nous avons proposé d'adapter la taille temporelle du noyau de la première couche convolutive et la profondeur du multiplicateur (qui détermine le nombre de filtres spatiaux) de la couche de convolution pour mieux gérer les composantes de moyenne et haute fréquences à partir de données EEG brutes

Nous avons montré que l'architecture de réseau neuronal convolutif EEGNet-4.8 proposée pour la tâche de discrimination MNS vs MI + MNS ainsi qu'un ensemble de paramètres spécifiques (telles que la bande de fréquences, la disposition des électrodes, la valeur de dropout) améliorent de manière significative les prédictions en particulier avec un petit nombre d'électrodes : 84% pour 6 électrodes, 88% avec 13 électrodes; 94,5% avec 128 électrodes).

Enfin nous avons analysé les résultats, examiné les limites des solutions proposées et discuté des moyens de les dépasser dans les travaux futurs. Ces perspectives sont très prometteuses en particulier pour l'amélioration de la détection des réveils peropératoires.

Mots-clés: Apprentisssage profond, Electroencéphalographie, Interfaces cerveau-ordinateur, extraction de caractéristiques, imagination motrices.

Abstract

Brain-Computer Interfaces (BCI) are communication and control systems that enable users to exchange information between the brain and an external device. Electroencephalography (EEG) is a popular technique for signal acquisition in the BCI domain because this method is non-invasive, mobile, cheap, and easy to use. Motor imagery (MI) is the most common discriminating task which is widely used in the BCI applications for controlling devices (computers, robots, wheelchairs, prostheses), stroke patients rehabilitation, monitoring consciousness during general anesthesia, and in entertainment. But despite all the advantages of this method, classification accuracy for imaginary movements from EEG signals remains low mainly due to the low signal-to-noise ratio.

The present work introduces three contributions to improve motor imagery-based BCI performance. First, we propose to estimate the quality of motor imageries using outlier detection before training the classifier. Second, we study feature selection in a combined movement motor imagery classification. Finally, a new deep learning architecture for EEG motor imagery-based BCI systems is developed, which affords a high level of accuracy to be potentially useful for preventing intraoperative awareness and other applications.

We start from introducing the basic concepts of motor imagery based BCI. We present neurophysiological concepts of brain rhythms and their localizations.

We discuss the methods of EEG preprocessing, feature extraction, feature selection, outlier detection, as well as machine learning methods frequently used in MI-based BCIs for classification tasks, such as linear discriminant analysis, support vector machine, and logistic regression algorithms.

We introduce the concept of artificial neural networks, their architectures, and optimization algorithms.

Finally, we present the architectures of convolutional neural networks (CNN) and their adaptations such as Deep Convolutional Network, Shallow Convolutional Network, and EEGNet for the classification of raw EEG signals.

To implement and validate the proposed methods, we have presented three different datasets. The latest dataset consists of conditions that allow us to compare two discriminative tasks: i) motor imagery vs. resting state; ii) the response to median nerve stimulation vs. the response to median nerve stimulation during motor imagery.

The results of estimating the quality of motor imagery using outlier detection showed that it can significantly increase the accuracy of a group of subjects.

Moreover, we have presented an approach for combined movement motor imagery classification which includes feature selection and classification using an architecture with three binary parallel-connected classifiers. We have shown that using this approach it is possible to achieve the comparable accuracy level as using the standard multi-class approach but with reduced time for BCI calibration.

We proposed to adapt lengths of temporal kernel size for the first convolutional layer and multiplier depth, which determines the number of spatial filters of the EEGNet depthwise convolution layer to better handle mid- and high-frequency components from raw EEG data. We have shown that proposed convolutional neural network architecture for the MNS vs. MI+MNS discriminating task as well as a set of specific parameters (such as frequency band, electrode layout, dropout date) significantly improve the prediction especially for layouts with a small number of electrodes (84% for 6 electrodes; 88% for 13 electrodes; 94.5% for 128 electrodes).

Finally, we have analyzed results, considered the limitations of proposed solutions, and discussed ways of solving them in future work. Perspectives of proposed methods in the field of MI-based BCI were represented. These results are very promising, in particular for improving the detection of intraoperative awareness.

Keywords: Brain-computer interface (BCI), deep learning, electroencephalogram (EEG), intraoperative awareness during general anesthesia, machine learning, median nerve stimulation, motor imagery.

Contents

List of Figures

List of Tables

Glossary

1d-AX : OneDimension-Aggregate Approximation
AAGA : Accidental Awareness during General Anesthesia
AAR : AutomaticArtifact Removal
ADAM : Adaptive Moment Estimation
AE : Autoencoder
ANN : Artificial Neural Networks
AVG : Average
BCI : Brain-Computer Interface
BGRU : Bidirectional Gated Recurrent Unit
BSS : Blind Source Separation
CAR : Common Average Reference
CNN : Convolutional Neural Network
CSP : Common Spatial Pattern
CWD : Choi-Willimas Distribution
CWT : Continuous Wavelet Transform
DBN : Deep Belief Networks
DL : Deep Learning
DNN : Deep Neural Network
ECoG : Electrocorticography
EEG : Electroencephalography
ELM : Extreme Learning Machine
ELU : Exponential Linear Unit
EMD : Empirical Mode Decomposition
EOG : Electrooculogram
ERD : Event-Related Desynchronization
ERP : Event-Related Potential
ERS : Event-Related Synchronization
FBCSP : Filter Bank Common Spatial Pattern
FC : Fully Connected
FFT : Fast Fourier Transform
FFTEM : Fast Fourier Transform Energy Map
fMRI : Functional Magnetic Resonance Imaging
fNIRS : Functional Near-Infrared Spectroscopy
FPR : False-Positive Rate
GAN : Generative Adversarial Network
GD : Gradient Descent
GPU : Graphical Processor Unit
GRU : Gated Recurrent Unit
HGD : High-Gamma Dataset
ICA : IndependentComponent Analysis
KMI : Kinesthetic Motor Imagery
LDA : Linear Discriminant Analysis
LOF : Local Outlier Factor

LSTM : Long Short-Term Memory
MEG : Magnetoencephalography
MI : Motor Imagery
ML : Machine Learning
MNS : Median Nerve Stimulation
PCA : Principal Component Analysis
PSD : Power Spectral Density
PSO : Particle Swarm Optimization
PTSD : post-traumatic stress disorder
RBF : radial Basis Function
RBM : Restricted Boltzmann Machine
ReLU : Rectified Linear Unit
RM : Real Movement
RNN : Recurrent Neural Network
SAE : Stacked Autoencoder
SELU : Scaled Exponential Linear Unit
SGD : Stochastic Gradient Descent
SNR : Signal-to-Noise Ratio
SReLU : S-shaped Rectified Linear Unit
SSVEP : Steady-StateVisual Evoked Potentials
STFT : Short-Time Fourier Transform
SVM : Support Vector Machine
ULM : Upper Limb Movement
VAE : Variational Autoencoder
VEP : Visually Evoked Potential
VI : Visual Imagery
WPD : Wavelet Package Decomposition

Introduction

Brain-computer interfaces (BCI) are comparably new technologies that are becoming more and more used. The goal of BCI is to build a communication channel for interaction between the human brain and physical electronic devices. For example, such application might be to create applications for paralyzed patients that allow to control robots or wheelchairs. Recently, however, the areas of potential application of the technology have gone beyond communication and control and could be found in the fields of rehabilitation [1, 2, 3, 4], treatment [5, 6], mental state monitoring [7, 8, 9, 10, 11] and entertainment [12, 13, 14, 15, 16].

Motor imagery (MI) is the most frequent interaction paradigm in the Brain-Computer Interface (BCI) domain and useful therapy for motor recovery, but the accuracy remains low and should be improved to create a reliable device that can be used in hospitals.

The use of machine learning in the field of BCI is very promising because machine learning techniques can detect activity and simulate complex processes in the human brain, recognize subject profiles, identify, assess, classify events, all of these allow to speed up and automate many processes in BCI. Most often, EEG signals are used to obtain information about the electrical activity of the brain, as the safest technology due to its non-invasiveness.

Typically, features are extracted from filtered EEG signals in alpha and beta frequency bands using various algorithms directly before classification. Common spatial pattern (CSP) algorithm is still the most popular method for feature extraction [17, 18, 19] from EEG signal and showed both advantages (i.e., high accuracy, flexibility, simplicity) and limitations (i.e., need for frequency band selection). Recently, other algorithms based on Riemannian geometry (such as tangent space (TS) algorithm [20, 21] and minimum distance to Riemannian mean (MDRM) algorithm [21, 22, 23]) showed promising results for MI detection.

BCI classification accuracy and performance levels have been greatly improved in the past decade mainly due to machine learning algorithms. However, there are still several issues that need to be addressed before BCI can be applied in real-life conditions. These problems include unstable classification performance, long calibration time, and poor reliability. Research on new developments is expected to bring solutions in this regard.

Along with the development of BCI, deep learning has been successfully applied in a variety of areas such as computer vision and natural language processing, significantly outperforming traditional machine learning approaches. Due to the considerable amount of investigations in this area, there are such technologies as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs).

While deep neural networks show better performances compared to standard machine learning methods in almost all application domains (such as computer vision or natural language processing), analysis of EEG signals particularly in the BCI domain remains an area where the use of deep learning techniques is still limited [24, 25]. First of all, the reason is the small amount of data available for training compared to image data in computer vision domains. Moreover, EEG signals are high-dimensional, non-stationary, and have a low signal-to-noise ratio. The potential of deep learning architectures that can deal with raw EEG signals especially for motor imagery-based tasks is rather unexplored and is active research field [25]. Several deep learning architectures already showed competitive results for MI vs. rest task in BCI domain [24, 25, 26, 27]. At the same time, we already showed the potential of convolutional neural networks to classify MI + MNS vs. MNS task on the smaller dataset [28]. Thus, further research on applications of deep learning in the BCI domain and an assessment of how these applications can be used in practice to implement stable robust BCI would be beneficial.

Hundreds of millions of general anesthesia are performed each year on patients all over the world [29]. Among these patients, 0.2-1.3% suffer from Accidental Awareness during General Anesthesia (AAGA) [30], i.e. an unexpected awakening during a surgical procedure under general anesthesia. This happens when general anesthesia

is not deep enough to compensate for all surgical stimuli related to the surgical procedure [31]. While the precise number of incidents is still debated, there is no doubt about the negative consequences for a patient experiencing an AAGA. Indeed, it can lead to physical suffering or psychological damage called post-traumatic stress disorder (PTSD) [32, 33]. People with explicit memories of PTSD may have long-term post-operative effects including anxiety, irritability, fear of death, sleep disturbances, and depressive syndromes [34]. In cases of implicit memories, the consequences are not fully established, but all evidence suggests that the patient may suffer from it [35, 30]. AAGAs also generate a high level of anxiety for anesthesiologists and are one of the three main causes of legal attacks by patients against the hospital [36], with a significant economic impact.

Prevention of AAGA during surgery is a challenge because technologies used by anesthesiologists are not able to predict all patients' recovery. Currently, there are two ways to monitor the depth of general anesthesia: (i) by observing the clinical features (i.e. heart rate, blood pressure, movements, sweating), which may be objective or subjective in nature, (ii) or by recording a portion of the electroencephalographic (EEG) signal at the frontal cortex (i.e. Bispectral Index, Patient State Index, Entropy) [37]. Unfortunately, both clinical signs and new brain indices (based on frontal lobe activity) are not satisfactory enough to successfully assess the depth of general anesthesia and prevent AAGA [38, 39, 36]. In particular, these cerebral indices do not ensure that consciousness during general anesthesia can be detected [40, 41]. Indeed, some studies have failed to demonstrate the superiority of these monitors over physical monitoring [42, 30] and others have demonstrated the unreliability of these techniques [43].

The NeuroRhythms team in Loria proposed concept of an innovative passive brain-computer interface (BCI) based on detection of movement intentions to prevent AAGA earlier [44, 45, 20, 28]. Indeed, during AAGA patients typically try to move to alert the medical staff during an AAGA, but discover that they are unable to move [46, 33]. The proposed solution incorporates a Median Nerve Stimulation (MNS) causing specific modulations of activity in the sensorimotor cortex [47], which are confirmed to be altered by an intention of movement [20, 48]. Indeed, during a movement or motor imagery, a minimal power level is maintained in both alpha and beta bands [49], called event-related desynchronization (ERD). Then, 300–500 ms after the end of the task, there is an increase of power referred to as an event-related-synchronization (ERS) in the beta band, also known as a post-movement beta rebound, lasting 1 s [47, 50]. Concurrently, in the alpha band, the power returns to a baseline after several seconds. Typically, these patterns occur in the motor cortex, in the contralateral hemisphere [51, 52, 53] but can appear bilaterally [54, 55, 56]. More interestingly, a MNS induces an ERD during the stimulation while an ERS appears after the stimulation [57, 48, 47] but the ERS is strongly abolished if a MNS is performed during an intention of movement [20, 45, 48]. This interesting process could make the detection of AAGA with a passive BCI possible. Indeed, we can imagine a routine system where the patient would be stimulated at the median nerve, and the analysis of ERD and ERS modulations of the motor cortex would be used to find out if the patient has an intention to move [20].

This new MNS-based BCI provides a solution for two major challenges: (i) be able to detect the intention to move of a person who is a victim of AAGA without any time marker, which is equivalent to continuously analyzing the EEG signal with limited indications concerning the temporal phases to be investigated. A second challenge is, therefore (ii) to obtain appropriate accuracy with an asynchronous BCI, which guarantees the reliability of a device that can be used in a medical context for which the motor intention is self-paced. Previous results show that this may provide a foundation for an innovative BCI that would allow the detection of AAGA [28, 20]. In previous studies [28], it was already shown that MI + MNS vs. MNS classification results have better accuracy than a MI vs. rest classification, suggesting that MNS could be used as a trigger. Although the use of median nerve stimulation is promising, it still requires more detailed investigations, particularly concerning which frequency bands can be used to maximize discrimination between the two classes.

Firstly, the goal of this work is to increase classification performance for the task of MI detection. Secondly, it is to decrease the number of EEG electrodes. Since applying EEG electrodes takes time and increases the preparation time it is often not possible to install 64 or 128 electrodes before the surgery. The solution might be to select the electrode setup which will show acceptable classification accuracy with a minimum number of electrodes. As a threshold accepted in the present study, 80% of accuracy were selected.

The objective of this thesis is to better investigate the ability of deep learning techniques such as convolutional neural networks (CNN) [58] to extract features from raw EEG signals applied to the problem of detection of intraoperative awareness during general anesthesia in the case when MNS is applied. Our goal is not only to increase accuracy for MI + MNS vs. MNS task, but also to optimize the proposed method by (i) selecting the best frequency band, (ii) selecting the best feature extraction and classifier method and its parameters, and (iii) reducing

the number of EEG electrodes to use.

Present thesis work is structured as follows.

In chapter 1 we have introduced the concepts of motor imagery (MI) based BCI. Signals which are frequently used in the BCI domain were described and we selected EEG technique for the further investigations because it is non-invasive, mobile, cheap, and easy to use method. We have presented neurophysiological concepts of brain rhythms and their localizations. Motor imagery as a frequently used interaction method for BCIs has been introduced as well as changes in the electrical activity of the human brain that usually comes with motor intentions.

Finally, the concept of a novel method based on median nerve stimulation for detection of accidental awareness during general anesthesia was presented and will be used as target application. Moreover we have shown influence of the median nerve stimulation on the oscillations in the sensorimotor cortex.

In chapter 2 we have presented preprocessing, feature extraction, feature selection, outlier detection and classification methods used in motor imagery based BCI.

As feature extraction methods, we have introduced time-frequency analysis, band power features, spatial analysis, and Riemannian geometry techniques. Common spatial pattern and filter bank common spatial pattern algorithms are still the most popular methods for feature extraction from EEG signals. Algorithms based on Riemannian geometry (such as tangent space algorithm and minimum distance to Riemannian mean algorithm) are becoming the new standard methods and could achieved better classification performance.

Also, we have presented machine learning methods frequently used in MI-based BCIs for the classification task, such as linear discriminant analysis, decision trees, gradient boosting, and support vector machines.

In chapter 3 we have introduced the concept of artificial neural networks and their architectures. We have selected Adam as an optimization algorithm for our work as it fixes vanishing learning rate and high variance problems and at the same time it progress faster comparing to other algorithms.

After analyzing of the existing solutions for the EEG signal processing, we have selected CNN architectures and their adaptations for raw EEG signals classification such as Deep Convolutional Network, Shallow Convolutional Network, EEGNetfor the following implementation and investigations.

In chapter 4 to implement and validate proposed methods, we have presented 3 different datasets, which were used in our experiments: motor imagery 2-class dataset, motor imagery 8-class dataset containing combined movements, and median nerve stimulation dataset. The latest dataset consists of conditions that allow us to compare two discriminative tasks: standard MI vs. rest and newly proposed MNS vs. MI+MNS task.

We have described participants, experimental tasks, time schemes, EEG data, preprocessing, feature extraction approaches, and cross-validation strategies that were used to obtain results for all datasets.

In chapter 5 we have represented results of classification using standard and deep learning techniques. Firstly, we have introduced a study about motor imagery quality estimation using outlier detection before BCI training.

Secondly, we have presented an approach for combined movement motor imagery classification which includes feature selection and classification using an architecture with three binary parallel-connected classifiers for detecting three basic movements and further detection of combinations of these movements.

Finally, motor imagery detection using deep learning methods on two different discriminating tasks (median nerve stimulation (MNS) vs. MI+MNS and MI vs. rest) was investigated. We have represented adaptation of deep learning architecture EEGNet that could better handle mid and high-frequency components from raw EEG data. Also, we have investigated the influence of parameters, such as frequency band, electrode layout and dropout date on the classification performance.

In chapter 6 we have analyzed results, considered limitations of proposed solutions, and discussed ways of solving them in future work. We have presented perspectives of proposed methods in the field of MI-based BCI. Also we have discussed the possibility of proposed setup application in clinical conditions with anesthetics for accessing the intraoperative awareness.

Chapter Conclusion finalizes the work and summarizes the application of deep learning techniques for MI-based BCI.

Chapter 1

Brain-Computer Interfaces (BCI)

Contents

1.1 BCI definition and architecture

Brain-Computer Interfaces (BCIs) are systems created to exchange information between the brain and physical electronic external device (for example, a computer). In unidirectional BCI, external devices can either receive signals from the brain or send signals to it (for example, mimicking the retina of the eye when restoring vision with an electronic implant). Bidirectional BCI allows the brain and external devices to exchange information in both directions [59]. Nowadays using BCI it is possible to control computers, robots [60, 61, 62], wheelchairs [63, 64], and prostheses [65, 66] but for simple tasks (Figure 1.1). Moreover, it is widely spread in different fields, like rehabilitation of stroke patients [1, 2, 3, 4], treatment [5, 6], mental state monitoring [7, 8, 9, 10, 11] and even in the entertainment [12, 13, 14, 15, 16]. BCI technology is a potentially powerful communication and control option between users and systems [67].

BCI could have different architectures according to the purpose of the system. But usually, all BCIs consists of five main modules, namely, input signal acquisition module, signal preprocessing module, feature extraction, classification, and feature translation (Figure 1.2).

The signal acquisition module is responsible for signal recording from the brain. It could be implemented using invasive or non-invasive procedures i.e. using intracranial or extracranial sensors.

The signal preprocessing module includes artifact removal (such as eye-blink artifacts), sampling, and filtering in required frequency bands.

The feature extraction module analyzes brain signals obtained previously. At this stage features, suitable are identified using mathematical methods. This allows defining useful information (power changes, peak amplitudes, etc) for classification. At this stage dimensionality reduction and feature selection occurs.

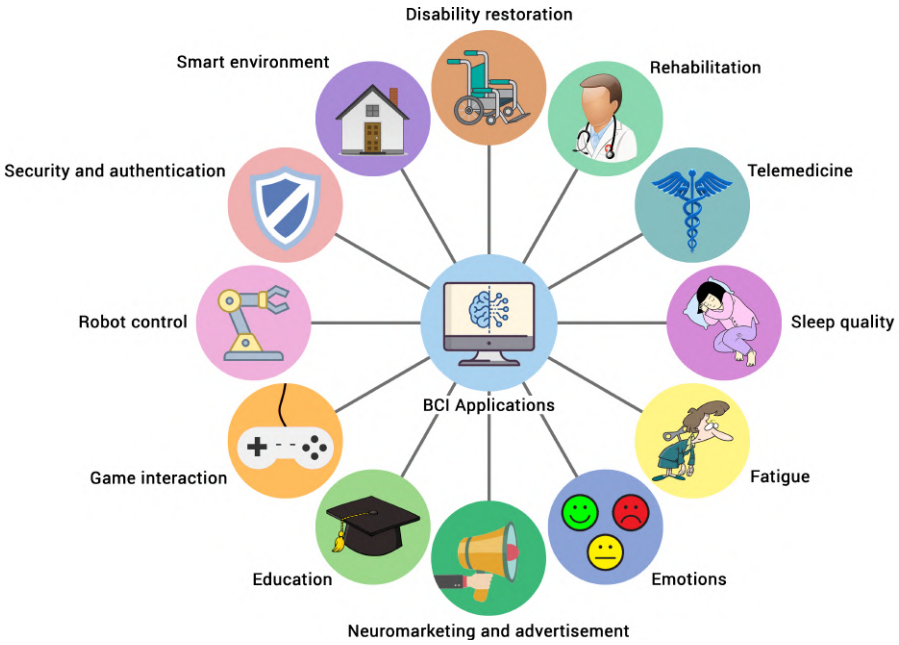


Figure 1.1: BCI applications.

The goal of the classification module is to map feature vectors obtained on the previous stage to the correct class among N. Here a class corresponds to a brain condition that the system detects and passes to the output module to generate different commands.

Translation module is responsible for translation classification results into the commands for device interaction which may consist of device control or commands transmission. The number of previously considered classes corresponds to several control commands. These actions could be used to control computers, wheelchairs, robotics, or produce commands for rehabilitation, treatment, mental state monitoring devices, and video games.

Besides, BCIs could include an additional feedback module to provide users with information about commands or brain activity. It could be helpful during training sessions to assist users in providing them information about errors or the quality of their mental actions. The aim is to improve the learning stage of the device and increase performance. Occasionally BCIs can have two feedback modules: one after the feature extraction part which responsible for the quality of the motor task performance and brain activity in general and other feedback after the translation module.

Figure 1.2 represents the general architecture of BCI, including modules listed above.

1.2 Neurophysiologcal aspects of BCI

1.2.1 Signals for BCI

The electrical activity of the brain is the main source of information for the BCI. There are many techniques for recording brain signals: electroencephalography (EEG), electrocorticography (ECoG), magnetoencephalography (MEG), functional near-infrared spectroscopy (fNIRS), functional magnetic resonance imaging (fMRI), and intracranial microelectrode array (MEA) recording. All of these techniques have their advantages and disadvantages according to scale level, noise, cost of technology, and invasiveness [68]. Further, these techniques are presented in more detail.

1.2.1.1 Electrocorticography (ECoG)

Electrocorticography (ECoG) is an intracranial method recording brain potentials using electrodes applied directly on the surface of the cortex. This method has a high spatial resolution with interelectrode spacing near 1

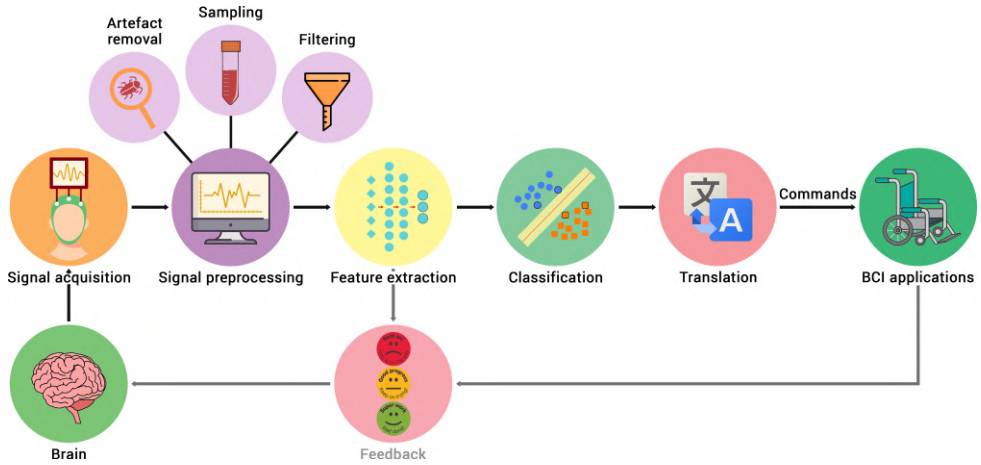


Figure 1.2: General BCI architecture. There are five main modules that every BCI consists of, namely, input signal acquisition module, signal preprocessing module, feature extraction, classification, and translation. Besides some BCIs could include feedback modules.

mm [69] and temporal resolution about 1kH, as well as a high signal-to-noise ratio (SNR). The main and extremely important disadvantage of this method is the need to perform surgery following postoperative recovery. Due to the invasiveness of this method, it makes its potential application inconvenient and limited to patients with strong disabilities.

1.2.1.2 Magnetoencephalography (MEG)

Magnetoencephalography is a functional neuroimaging method for mapping brain activity by a recording of magnetic fields produced by electric currents occurring in the neurons of the brain. The advantages of this method are high spatial (millimeters) and temporal (milliseconds) resolution [70]. But the high cost of equipment and the additional requirements, such as a magnetically shielded room from external magnetic signals, including the Earth's magnetic field make applications of this technology limited and not suited for many BCI cases.

1.2.1.3 Functional Magnetic Resonance Imaging (fMRI)

This method is based on the fact that cerebral blood flow and neuronal activation are interrelated. Changes in blood flow (hemodynamic response) are associated with the consumption of energy by brain cells. When an area of the brain is used, blood flow to that area also increases. So functional MRI measures brain activity by detecting changes in blood flow. This method has a high spatial resolution (millimeters), but at the same time it has several disadvantages: first of all, the equipment is expensive and cumbersome and this technique also has a low temporal resolution, because changes in the oxygenation of the cells are not instant and it takes from 1 up to 8 seconds.

1.2.1.4 Functional Near Infrared Spectroscopy (fNIRS)

Functional Near-Infrared Spectroscopy (fNIRS) is a non-invasive technique that measures cerebral hemodynamic responses using near-infrared light. Cerebral hemodynamic responses correspond to activation or deactivation of the parts of the brain. This method capable to track changes both in oxy- and deoxyhemoglobin concentration. The advantage of fNIRS is its spatial resolution ($1cm^2$) [71] which is suitable for the BCI domain and portability of the system. A Head cap with sensors makes this technique comfortable for users. fNIRS could be used for BCI to monitor the depth of anesthesia.

1.2.1.5 Electroencephalography (EEG)

Electroencephalography (EEG) is a non-invasive technique, which measures the electrical activity of the brain. EEG signal is a complex oscillating electrical process that is the result of electrical summation and filtration of elementary processes occurring in the neurons of the brain.

Table 1.1: Advantages and disadvantages of different signal recording techniques

	Advantages	Disadvantages
ECoG	High spatial and temporal resolution, High SNR	Invasive
MEG	High spatial and temporal resolution, High SNR, Non-invasive	Big room, Expensive, High degree of immobility
fMRI	High spatial resolution, Non-invasive	Low temporal resolution, Expensive, High degree of immobility
fNIRS	Non-invasive, Portable	Low temporal resolution
EEG	Non-invasive, High temporal resolution, Portable, Cheap	Low SNR

Frequency, amplitude, and location are the main characteristics of the EEG waveform. The following methods are used to analyze EEG signals: frequency domain, time-frequency, and non-linear techniques like correlation dimension, largest Lyapunov exponent, Hurst exponent, different entropies, fractal dimension, higher-order spectra, phase space plots, and recurrence plots.

EEG is widely used in the diagnosis of various diseases or disorders of the brain. This method is cheap, safe, and easy to use, that's why it is widely spread in the diagnosis of Parkinson's and Alzheimer's disease, epilepsy, panic attacks, hysteria, drug poisoning, and other disorders of the brain. Moreover, it has a high temporal resolution (milliseconds). EEG recordings use a head cap with electrodes located all over the scalp or over specific regions of the head.

The main disadvantages of EEG are low noise-to-signal ratio and relatively low spatial resolution ($1cm^2$).

Now EEG systems become more and more comfortable for users due to developments in areas of dry electrodes and wireless EEG head caps that allow full mobility for the user. A comparison of the main advantages and disadvantages of the signal recording techniques are presented in the table 1.1.

1.2.2 Brain rhythms

Brain rhythm is a dynamic process characterized by simultaneous activation of a specific population of neurons, or synchronization between local oscillations of two or more populations of neurons [72]. The synchronized activity of large populations of neurons is the main mechanism for the formation of rhythms in the human brain. Rhythms play an important role in the transmission and processing of information in the central nervous system (sleep, memory, attention, spatial navigation, etc.)

Brain rhythms can be divided into alpha-, beta-, theta- and delta-rhythms (Figure 1.3). Each of these rhythms is a special type of cortical activity characterized by brain waves and is associated with states of consciousness such as anxiety, rest, dreaming, or sleep [72]. Brain activity changes continuously. Therefore, our mental state reflects the mix of different waves of brain activity at different localizations.

Brain rhythms are characterized by the largest power of wave activity. For example, a person with closed eyes may have a high power of alpha and low power of beta waves in the visual cortex because it does not process visual information. On the other hand, when the eyes are open, the power of the alpha waves decreases sharply, and the beta waves increase as a result of processing the visual information coming to the visual part of the cortex.

Examples of EEG waves are given below and are shown in Figure 1.3.

- Delta waves (< 4 Hz): include all EEG waves with a frequency below 4 Hz and often have amplitude 2-4 times higher than most other types of brain waves. They occur during deep rest, dreamless sleep, frontally in adults, posteriorly in children, and with severe brain damage [74].

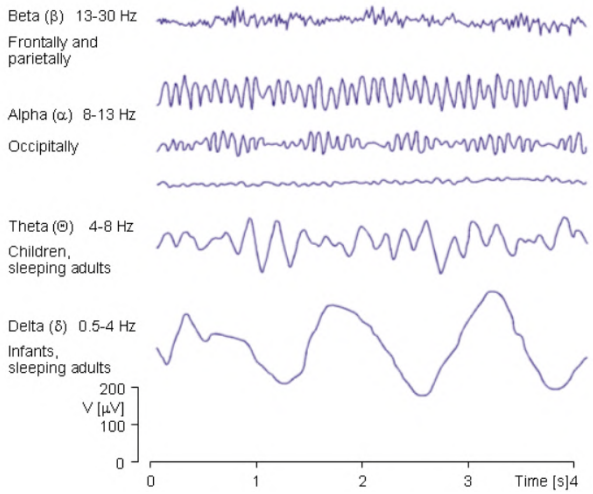


Figure 1.3: Examples of EEG waves characterizing brain rhythms (from [73]).

- Theta waves (4 - 8 Hz): This rhythm is usually observed during deep relax state, tiredness, sleepiness, REM (i.e., Rapid Eye Movement) sleep, and could be caused by emotional stress, alcohol consumption and fatigue [75]. Theta waves are divided into two types: the hippocampal theta waves and cortical theta waves.

- Alpha waves (8 - 13 Hz): are rhythmic oscillations with a frequency within 8-13 Hz, which are detected in the EEG of almost all healthy adults at rest state, during daydream and calm. These waves are most intense in the occipital lobe, but they can also be observed in the parietal and frontal areas of the scalp. Their value is usually about 50 μV. Alpha waves disappear during deep sleep. Alpha activity varies throughout the day according to our natural biorhythms, so this factor must be taken into account when performing an EEG analysis.

- Mu waves (8 - 12 Hz): It is a specific rhythm associated with motor activity and observed mostly under the primary motor cortex. Its peculiarity is that Mu waves decrease during the preparation and the execution of the movement and this is called Event-Related Desynchronization (ERD) [76]. While at rest state it increases and called Event-Related Synchronization (ERS) [77].

- Beta waves (13 - 30 Hz): When a person's attention is focused on a certain type of mental activity, alpha waves are replaced by asynchronous, high-frequency, but low-amplitude beta waves. Beta waves oscillations with a frequency from 13 to 30 Hz appeared during the alert, active thinking, anxiety, panic attack, focus, concentration [74]. This rhythm is most strongly expressed in frontal areas, but various types of intensive mental activity sharply amplify and extend to other areas of the brain. Thus, the power of beta-rhythm increases with mental stress or emotional arousal. Their amplitude is 4-5 times smaller than the amplitude of alpha waves and is 5-30 μV.

- Gamma waves (> 30 Hz): These waves are responsible for perception and consciousness. Often appeared during the combination of two senses in somatosensory cortex [74]. Also, their presence may be caused by electromyographic artifacts [78].

1.2.3 Brain rhythms localization

The cerebral cortex is an accessible source of information about motor and executive behavior, thus it represents an area of interest to researchers in the BCI domain. Interacting through synaptic connections, neurons form electrical impulses that control the activity of the entire body [79].

The left hemisphere of the human brain is responsible for the right half of the body movement, the right one for the left. Right and left hemispheres have different functions. Visual signals go to the visual cortex (in the occipital lobe), tactile to the somatosensory cortex (in the parietal lobe), olfactory signals to the olfactory cortex on the base

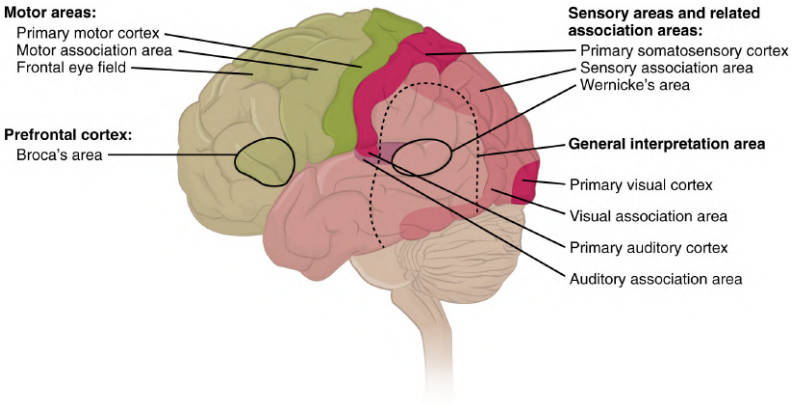


Figure 1.4: Types of cortical areas of the human brain[80].

of the frontal lobe and medial aspect of the temporal lobe, etc. In the associative areas of the cortex, the integration of sensory signals of different types takes place [79].

The motor areas of the cortex (the primary motor cortex and other areas of the frontal lobes showed in Figure 1.4) are responsible for regulating movement.

The brain has the property of plasticity. If one of its parts is affected, other parts after a while can compensate for its loss of function. Brain plasticity also plays a role in learning new skills.

The frontal lobe represents the most interest for the BCI domain. It is the largest lobe and is located at the front of each hemisphere of the brain. The frontal lobe is covered by the frontal cortex. The frontal cortex includes the premotor cortex, and the primary motor cortex responsible for movements. The front part of the frontal lobe is covered by the prefrontal cortex.

1.2.3.1 Prefrontal cortex

The prefrontal cortex is the part of the cerebral cortex which covers the front part of the frontal lobe.

This area of the brain is involved in the planning of complex cognitive behavior, personality manifestation, decision-making, and regulation of social behavior [81]. The main purpose of this area of the brain is considered as the coordination of thoughts and actions in correspondence with internal goals [82].

The main function of the prefrontal cortex is an executive function.

1.2.3.2 Primary motor cortex

One of the main functions of the primary motor cortex is to control voluntary discrete movements performed by muscle groups.

The primary motor cortex contains the topographic representation of various areas of the body upside down, from the legs (top of the hemisphere) to the mouth (bottom of the hemisphere) (Figure 1.5). The primary motor cortex of each hemisphere controls muscles on the opposite side of the body. The surface area of the regions of the primary motor cortex that control various parts of the body is proportional not to the physical surface of these parts, but to the number and degree of complexity of their motor responses. Because of this, about 75% of the main volume of the primary motor cortex is represented by insignificant (by weight), but performing subtle non-stereotypical movements of the muscles of the hands, fingers, facial muscles of the face, lips, tongue [83, 84, 85].

Along with the primary, there is a secondary motor cortex, which includes the premotor cortex and supplementary motor area.

1.2.3.3 Electrode placement

The way the electrodes are placed on the scalp to obtain an EEG is called a derivation scheme. The standard method for covering different areas of the brain is known as the International 10-20 system [87]. Figure 1.6 represents the electrode placement over the scalp. The main idea of this system is the uniform coverage of the

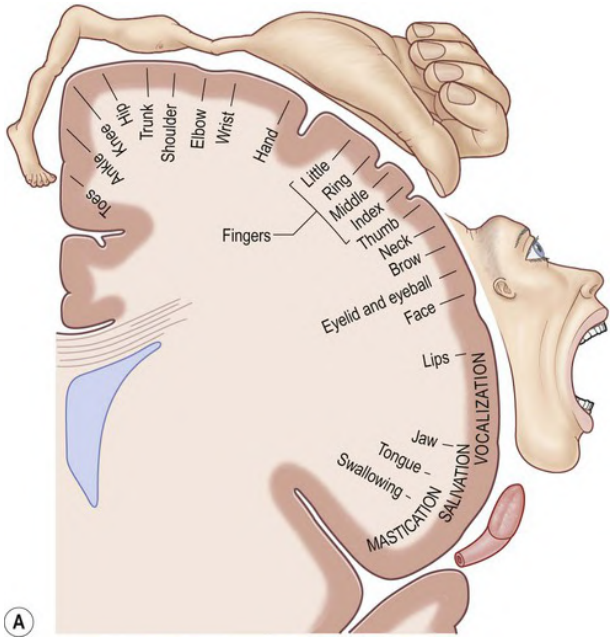


Figure 1.5: The motor strip contains an somatotopic representation of the contralateral half of the body, from the toes to the face and tongue [86].

cortex based on the proportional distance separating the electrodes from each other. For this purpose, nasion and inion are taken as references, of which five aligned electrodes are placed at 10%, 20%, 20%, 20%, 20%, and 10% of the distance separating both points.

On the International 10-20 system electrodes *C3*, *Cz* and *C4* are located above the areas of the cerebral cortex, which are responsible for the motor activity of the right hand, legs, and left hand, respectively.

There are also other layouts to cover a scalp, such as the ABC layout represented in Figure 1.7. ABC layouts partly overlap with the international 10-20 system.

1.2.4 Motor imagery

Motor imagery (MI) is a mental process during which a person imitates or repeats a certain movement without execution. At the same time, the person feels its performance [88]. It is widely used for sports training as a mental practice of movement, for neurological rehabilitation, as well as for studies of the structure of unconscious processes that precede the performance of movement [89]. MI represents the result of conscious access to the content of the intention of a movement [90], which is usually performed unconsciously during movement preparation [91].

MI has two different components, namely the visual imagery (VI) and the kinesthetic motor imagery (KMI) [92]. VI activated predominantly the occipital regions and the superior parietal lobules [93], whereas KMI generates an event-related desynchronization and an event-related synchronization in the contralateral sensorimotor area, which is similar to the one observed during the preparation of a real movement (RM) [49].

MI is widely used in the BCI domain and it is the most common discriminating task. MI-based BCIs are used in the recovery of the motor capacity after neurological damage such as rehabilitation for stroke patients [94] or monitoring consciousness during general anesthesia [95]. For example, MI training is a promising approach in facilitating paretic limb recovery [96].

Imaginary movements are used to train brain-computer interfaces, but it is often impossible to assess the quality of a subject's performance before training the BCI. The process of performing imaginary movement for the user is not an easy task and requires prior training. It could be tiring, boring, and hard to concentrate on the subject. The condition when BCI users fail to reach competence in using a BCI during a standard training process is called BCI illiteracy [97, 98]. Therefore, the classification accuracy for some subjects is frequently many times lower than for other subjects. But the reason why is not well-known.

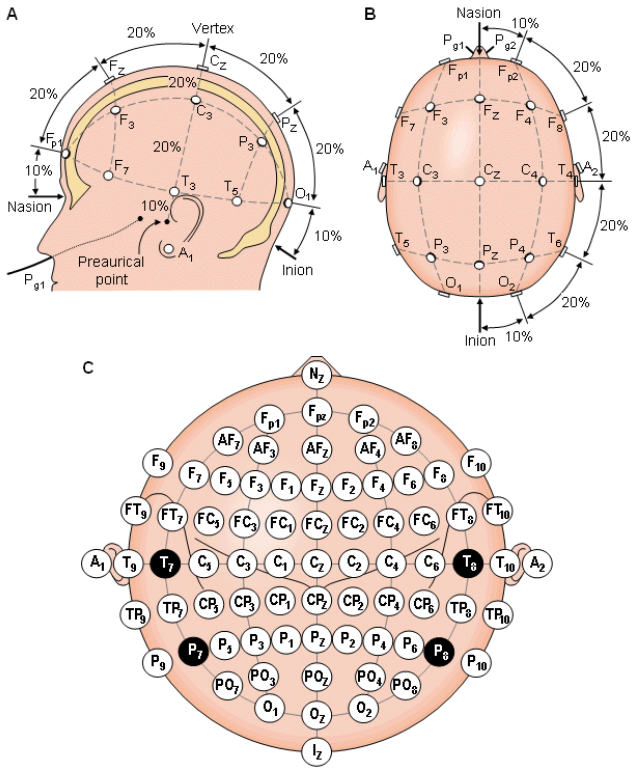


Figure 1.6: The international 10-20 system seen from (A) left and (B) above the head. A = Ear lobe, C = central, Pg = nasopharyngeal, P = parietal, F = frontal, Fp = frontal polar, O = occipital. (C) Location and nomenclature of the intermediate 10% electrodes, as standardized by the American Electroencephalographic Society [73].

All existing methods could be improved if there is a method of assessing the quality of subjects MI performance. Training data could be filtered to make models generalize better, and therefore increase the accuracy of detecting imaginary movements on new data. We will propose a method based on outlier detection to estimate the quality of imaginary movements before model training (section 5.1).

1.2.5 Event-related synchronization and desynchronization

Dynamic of motor activity is observed by the index called Event-related synchronization/desynchronization (ERD/ERS%). Event-Related Desynchronization (ERD) is the suppression of oscillations in the alpha and beta rhythms that takes place in the sensorimotor cortex during the movement [51]. ERD could be observed not only during a real and imaginary movement itself but also before them, during the preparation of movement [99, 100]. Event-Related Synchronization (ERS) is an opposite occurrence of ERD which is characterized by increasing of power in alpha and beta bands just after the motor task [77].

More precisely, compared to a resting state taken before motor imagery, several power modulations are observed in the alpha and in the beta bands of the EEG signal measured over the sensorimotor area corresponding to the body part involved in the motor imagery.

At the beginning of the motor task, there is a gradual power decrease (ERD) in the alpha and in the beta bands (Figure 1.8). Then, as seen from Figure 1.8 a low power level is maintained during the movement.

Finally, from 300 to 500 milliseconds after the end of the motor imagery, there is an ERS or post-movement beta rebound with a duration of about one second. Although several studies showed an activity uniquely in the contralateral area [52], other studies showed that ERD and ERS are also in the ipsilateral area [55].

Detailed algorithm for computing ERD/ERS% will be presented in the section 2.2.2.1

The emergence of ERD and ERS patterns during and after a MI has been intensively studied in the BCI domain [101] to define detectable commands for the system.

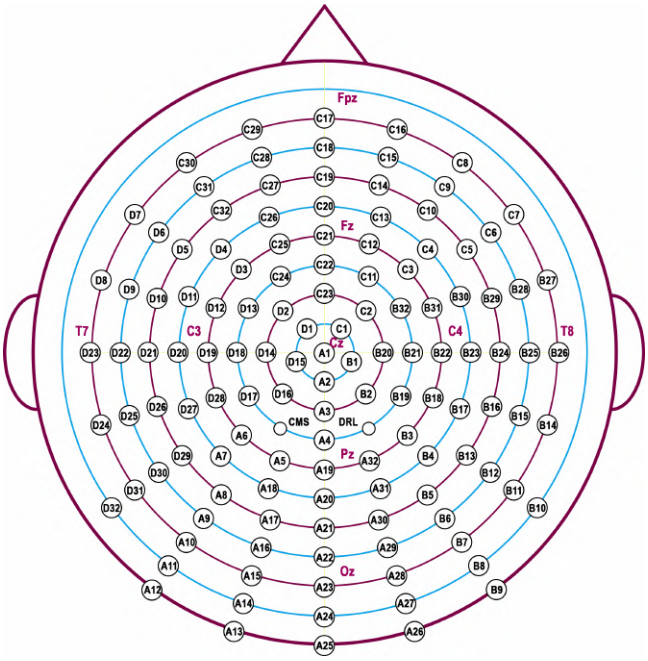


Figure 1.7: EEG electrode placement according to the ABC layout.

1.3 BCI application for general anesthesia monitoring

General anesthesia is an artificially induced reversible depression of the central nervous system with a switch-off of consciousness, suppression of sensitivity (primarily pain) and reflex reactions, and a decrease in skeletal muscle tone. All this occurs when one or more general anesthetics are administered, the optimal dose and combination of which is selected by the anesthesiologist, taking into account the individual characteristics of a patient and depending on the type of medical procedure [102, 103]. Anesthesia drugs inhibit interneuronal (synaptic) transmission of excitation in the central nervous system. At the same time, the transmission of afferent impulses is disrupted [104, 102].

The main goal of anesthesia is to slow down the body's reactions to surgery, primarily the sensation of pain. At the same time, drug sleep, with which the concept of "anesthesia" is most often associated, is only one component of anesthesia. When carrying out anesthesia, it is also important to suppress or significantly reduce the severity of

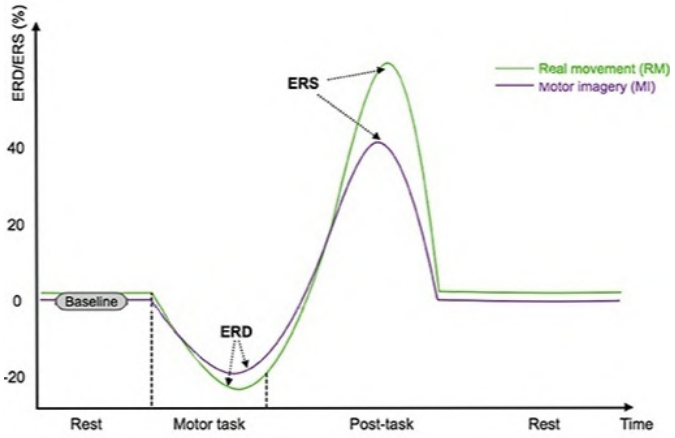


Figure 1.8: Example of ERD and ERS occurrence during real movement and motor imagery (adopted from [20]).

the autonomic (automatic) reactions of the body to surgical trauma, which are manifested by an increase in heart rate, an increase in blood pressure, and other phenomena that can occur even when consciousness is switched off. This suppression of autonomic responses is called pain relief or analgesia. Another component of anesthesia is muscle relaxation, which is necessary to ensure normal conditions for the work of surgeons [103].

1.3.1 Accidental Awareness during General Anesthesia (AAGA)

Accidental Awareness during General Anesthesia (AAGA) is an unexpected awakening during a surgical procedure under general anesthesia [30]. This happens when general anesthesia is not deep enough to compensate for all surgical stimuli related to the surgical procedure [31]. While the precise number of incidents is still debated, there is no doubt about the negative consequences for a patient experiencing an AAGA. Indeed, it can lead to physical suffering or psychological damage called post-traumatic stress disorder (PTSD) [32, 33]. People with explicit memories of PTSD may have long-term post-operative effects including anxiety, irritability, fear of death, sleep disturbances, and depressive syndromes [34]. In cases of implicit memories, the consequences are not fully established, but all evidence suggests that the patient may suffer from it [35, 30]. AAGAs also generate a high level of anxiety for anesthesiologists and are one of the three main causes of legal attacks by patients against the hospital [36], with a significant economic impact.

Prevention of AAGA during surgery is a challenge because technologies used by anesthesiologists are not able to predict all patients' recovery. Currently, there are two ways to monitor the depth of general anesthesia: (i) by observing the clinical features (i.e. heart rate, blood pressure, ciliary reflex, sweating), which may be objective or subjective in nature, (ii) or by analyzing indices from the electroencephalographic signal at the frontal cortex (i.e. *Bispectral Index*, *Patient State Index*, *Entropy*) [37]. Unfortunately, both clinical signs and brain indices based on frontal lobe activity are not satisfactory enough to successfully assess the depth of general anesthesia and prevent AAGA [38, 39, 36]. In particular, these cerebral indices do not ensure that consciousness during general anesthesia can be detected [40, 41]. Indeed, some studies have failed to demonstrate the superiority of these monitors over physical monitoring [42, 30] and others have demonstrated the unreliability of these techniques [43].

1.3.2 BCI for AAGA detection

In 2016, Blokland et al. [95] firstly proposed an approach to use BCI for detection of AAGA. The idea of such a BCI is to detect the intention of moving from the patient which usually preceded an AAGA. In this work, Blokland et al. investigated the effect of propofol on the EEG signals recorded over the motor cortex. In this investigation, subjects were not completely anesthetized and were asked to perform movements according to sound beeps while an increasing dosage of anesthetic was administered to them. This first approach shows that the BCI domain could contribute to the issue of AAGA. However, the study conducted by Blokland et al. was based on synchronous active communication, i.e., the voluntary subject was explicitly asked to perform a movement during the experiment after a visual and audible signal, which does not realistically reflect the conditions during intraoperative awareness. Therefore, a strategy must be found to design a passive BCI whose task would be to detect the intention of movement of an AAGA victim [20].

NeuroRhythms team from LORIA laboratory proposed concept of an innovative passive brain-computer interface (BCI) based on detection of movement intentions to prevent AAGA [44, 45, 20, 28]. Indeed, during AAGA patients typically try to move to alert the medical staff during an AAGA, but discover that they are unable to move [46, 33]. The proposed solution incorporates a Median Nerve Stimulation (MNS) causing specific modulations of activity in the sensorimotor cortex [47], which are confirmed to be altered by an intention of movement [20, 48]. Experimental design of such a median nerve stimulation based BCI presented in section 4.3.

This new MNS-based BCI provides a solution for two major challenges: (i) be able to detect the intention to move of a person who is a victim of AAGA without any time marker, which is equivalent to continuously analyzing the EEG signal with limited indications concerning the temporal phases to be investigated. A second challenge is, therefore (ii) to obtain appropriate accuracy with an asynchronous BCI, which guarantees the reliability of a device that can be used in a medical context for which the motor intention is self-paced. Results show that this may provide a foundation for an innovative BCI that would allow the detection of AAGA [28, 20].

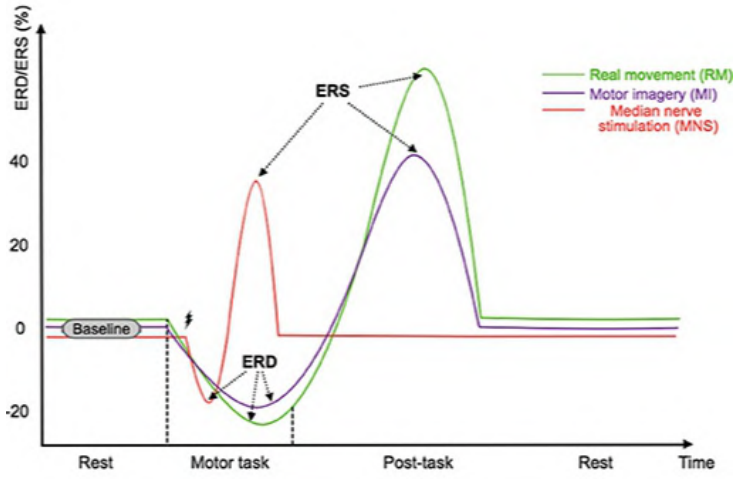


Figure 1.9: Example of ERD and ERS occurrence during real movement, motor imagery and median nerve stimulation (from [20]).

1.3.3 Median nerve stimulation (MNS) in BCI

Electrical stimulation of the median nerve is a safe, inexpensive, non-invasive technique that is commonly used for neuroresuscitation of coma patients [105, 106] and to alleviate symptoms of nausea and vomiting [107].

As classification performance of standard MI vs. rest discriminating task remains low in the BCI domain there is a requirement for improvement. MNS could be one of the possible triggers for BCIs [20]. Of course, the use of this technique reduces the user's mobility. But MNS can be used for BCI that does not require mobility, such as BCI for AAGA detection.

1.3.4 Influence of MNS on ERD/ERS

It was observed that there is a difference in oscillations in the sensorimotor cortex during motor activity depending on whether the median nerve of a person was stimulated or not [20].

While, during a movement or motor imagery, a minimal power level is maintained in both alpha and beta bands [49], called event-related desynchronization (ERD). Then, 300–500 ms after the end of the task, there is an increase of power referred to as an event-related-synchronization (ERS) in the beta band, also known as post-movement beta rebound (Figure 1.9), lasting 1 s [47, 50]. Concurrently, in the alpha band, the power returns to a baseline after several seconds. Typically, these patterns occur in the motor cortex, in the contralateral hemisphere [51, 52, 53] but can appear bilaterally [54, 55, 56].

More interestingly, a MNS induces an ERD during the stimulation while an ERS appears after the stimulation [57, 48, 47] but the ERS is strongly abolished if a MNS is performed during an intention of movement [20, 45, 48]. This interesting process could make the detection of AAGA with a passive BCI possible. Indeed, we can imagine a routine system where the patient would be stimulated at the median nerve, and the analysis of ERD and ERS modulations of the motor cortex would be used to find out if the patient has an intention to move [20].

Results in chapter 5 show that a MI + MNS vs. MNS classification results have better accuracy than a MI vs. rest classification, suggesting that MNS could be used as a trigger.

1.4 Conclusion

In this chapter, the basic concepts of brain-computer interfaces were represented. Signals which are used in the BCI domain were described and the EEG technique was selected as a non-invasive, mobile, cheap, and easy to use method.

Also in this chapter were discussed neurophysiological concepts of brain rhythms and their localization. Motor imagery as a frequently used trigger for BCIs was represented as well as event-related synchronization and

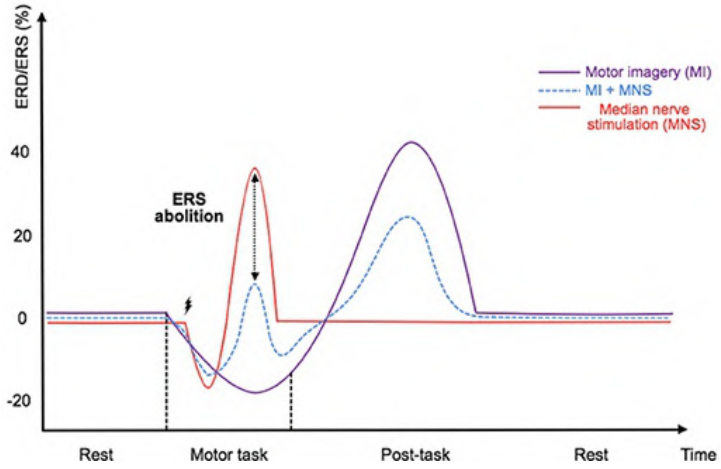


Figure 1.10: Example of ERD and ERS occurrence during motor imagery, motor imagery during median nerve stimulation and median nerve stimulation itself (from [20]).

desynchronization which takes place in the brain rhythms during motor imagery and motor tasks.

This thesis will present how to improve it using classic machine learning and deep learning techniques to create a reliable device that can be used in hospitals.

Firstly, in chapter 2 we will present classic machine learning methods for feature extraction and classification in the MI-based BCI domain.

Secondly, we will consider deep learning techniques as universal feature extractor combined together with the classification stage for raw EEG signals in chapter 3.

All the results obtained in this thesis will be carried out on real EEG data described in chapter 4.

Results of machine learning and deep learning techniques application to improve MI detection from the EEG signals will be represented in chapter 5.

The problem of accidental awareness during general anesthesia was posed during this chapter and current techniques for its detection were discussed. Moreover, the concept of the novice method based on BCI and median nerve stimulation for detection of accidental awareness during general anesthesia was presented. Although the use of median nerve stimulation is promising, it still requires more detailed investigations, particularly concerning which frequency bands and electrode layouts can be used to maximize the detection of accidental awareness during general anesthesia. These investigations will be presented in section 5.3.

Another challenge for this kind of BCI remains to be used in clinical practice. We will discuss it further in section 6.4.

Chapter 2

Classic machine learning methods for motor imagery detection

Contents

Machine learning is a subsection of artificial intelligence that includes a set of methods used to create complex models and algorithms. These models allow researchers, engineers, and analysts to produce reliable solutions and results and reveal the hidden dependence in data through training [108].

The use of machine learning in the field of BCI is important because machine learning techniques can detect brain activities and reveal complex processes in the human brain, recognize subject profiles, identify, assess, classify events, all of these allow to speed up and automate many processes in BCI. Due to the rapid development of

technology, particularly in the field of parallel computing, cloud technology and computing capabilities usage of machine learning has increased.

The whole process includes important steps such as data preprocessing, feature extraction, feature selection, model training, prediction. Some of these parts used in our projects will be described in this chapter.

2.1 Preprocessing

After the acquisition, EEG signals are preprocessed. Typically, the signals pass through several stages of preprocessing: spatial filtering, frequency filtering, and artifact removal.

2.1.1 Spatial filtering

Spatial filtering is done to highlight the specific information at one location compared to the neighborhood. To accomplish this, the mean or weighted mean of the signals from the other electrodes is subtracted from the target signal.

There are a few different methods of spatial filtering for large electrode setups (Figure 2.1). A ground electrode for EEG recordings is often placed on the forehead, but could be placed anywhere else on the body. Usually, receiving a signal, each electrode referenced to an electrode behind one ear or both, this is called an "ear reference" (Figure 2.1a). Common average reference (CAR) refers to a filter for which the mean of the signals from all electrodes is taken as a baseline (Figure 2.1b). A bipolar filter is a subtraction one signal from another (not represented in Figure 2.1a). The Laplacian filter approximates the second spatial derivative of the signal. For a small Laplacian filter, the distance-weighted mean of the four nearest electrodes is calculated and used as a reference (Figure 2.1c). For a large Laplacian filter, the next-nearest neighbors are used [109] (Figure 2.1d).

2.1.1.1 Common average reference (CAR)

Common average reference (CAR) is the most often used spatial filter [110, 111]. In CAR, the potential at each electrode is measured with respect to the average of all electrodes [111]. The CAR defined by the following equation 2.1:

$$V_i^{CAR} = V_i - \frac{1}{n}\sum_{j=1}^{n} V_j, \tag{2.1}$$

where V_i is an amplitude of each site (electrode) and n - number of sites.

Bad electrodes should be detected (for example using a threshold on amplitude or impedance) and removed from the processing stage before CAR because in this method each electrode contributes to others.

2.1.1.2 Laplacian filter

Laplacian filter is obtained by subtracting the sum of the weighted potential of the surrounding electrodes from the considered electrode potential where the weight is electrode distance-dependent [112, 113].

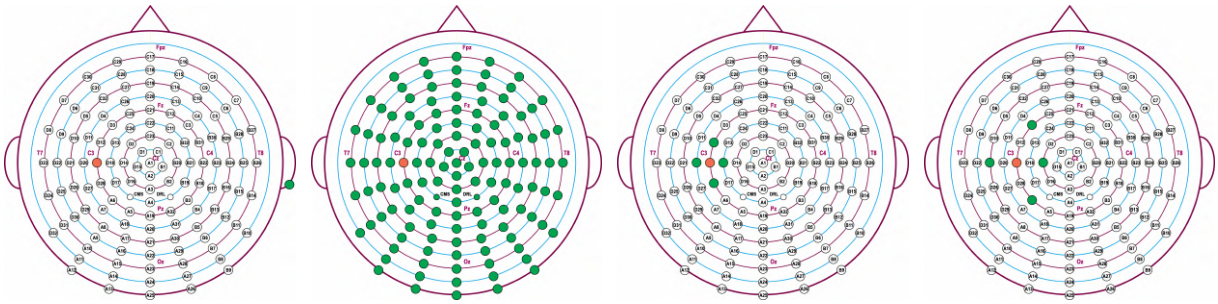
The small and large Laplacian filters are calculated according to equations 2.2 and 2.3:

$$V_i^{Lap} = V_i - \sum_{j}^{n} g_{ij} V_j, \tag{2.2}$$

with

$$g_{ij} = (d_{ij}\sum_{j}^{n}\frac{1}{d_{ij}})^{-1} \tag{2.3}$$

where V is a signal value, Lap means Laplacian, i is an index that corresponds to the considered electrode, j is an index of the nearest neighbor electrodes and n is the total number of electrodes, d_{ij} is the distance between electrodes i and j.



(a) Ear reference **(b)** Common Average Reference **(c)** Small Laplacian **(d)** Large Laplacian

Figure 2.1: Different spatial filters. Mean or weighted mean of green electrodes would be subtracted from the red electrode for each spatial filter presented on the figure.

Laplacian filters are usually used with electrode layouts that include a large number of electrodes covering the scalp as a grid. Nearest neighbors selected as North, South, East, and West electrodes relative to the considered electrode [112].

2.1.2 Filtering in the frequency domain

Filtering in the frequency domain is one of the main steps in signal processing. It helps to separate useful data from noise.

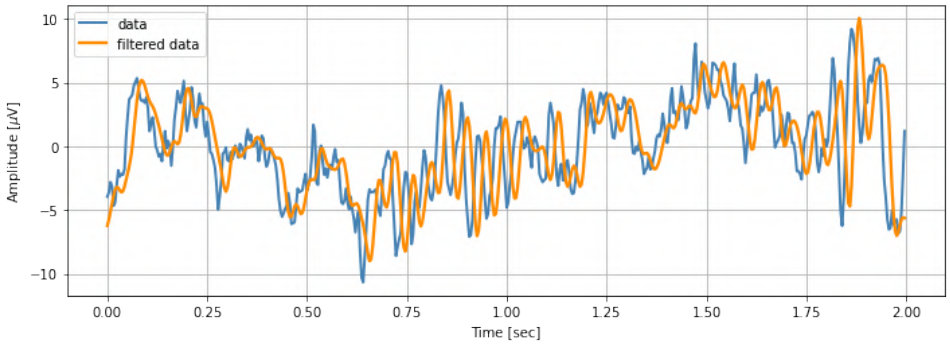
In this section we used EEG signal from the dataset described in detail in section 4.2 to show the examples of applications for different types of filters in BCI domain. For this purpose we used raw EEG signal of right hand motor imagery from channel C3 of subject 2. We used 5th order Butterworth filters of different types with different cutoff frequency ranges. Figure 2.2 represents raw and filtered signals. Filters are divided into four main types according to the frequency response:

- low-pass filter (LPF) has a relatively high transmission rate at frequencies that are below the cutoff frequency, and relatively low at other frequencies. Therefore, such filters keep the value of the spectral components of the input signal at low frequencies. LPF does transmit low-frequency signals and does not transmit high-frequency signals. LPF is used in BCI for downsampling, as a first step to reduce high-frequency signal components before decimation. Also, LPF can be considered as a part of band-pass filtering widely used in the BCI domain. Filtered signal using LPF filter with cutoff frequency 30 Hz showed on Figure 2.2a;

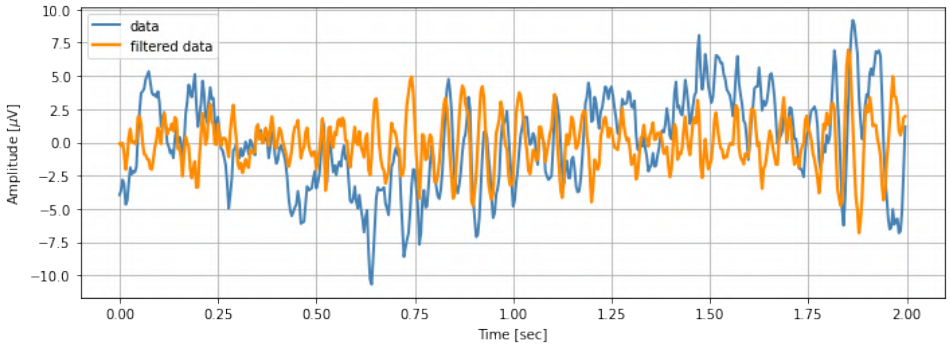
- high-pass filter (HPF) has a high transmission rate at frequencies that are higher than the cutoff frequency, and low transmission rate at frequencies from zero to the cutoff frequency f_p. Therefore, they pass without changes part of the spectrum of the input signal, which is at frequencies from the cutoff frequency f_p to $+\infty$ and do not pass the components of the signal spectrum from 0 to f_p. HPF transmits high-frequency signals. HPF is used in BCI as a part of a band-pass filter. Filtered signal using HPF filter with cutoff frequency 8 Hz showed on Figure 2.2b;

- band-pass filter (BPF) has a high transmission factor only in a certain frequency band between frequencies f_{p1} and f_{p2}. Outside this range, the transmission rate is small. Therefore, only those spectral components of the signal that are in the bandwidth of the filter will pass to the output, and those that lie from zero to f_{p1} and from f_{p2} to $+\infty$ will not pass.

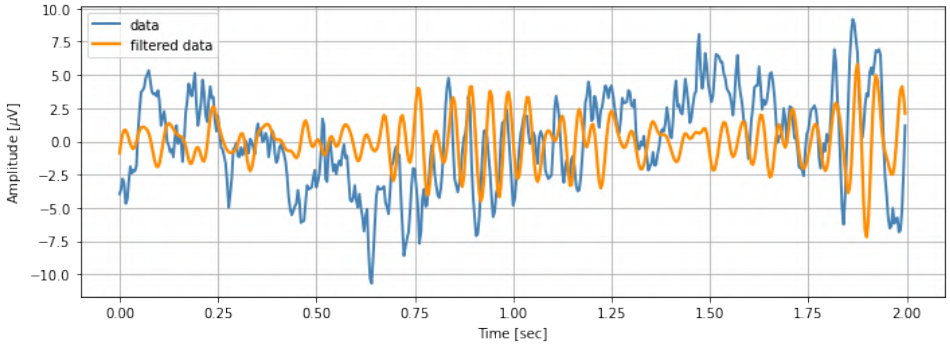
  BPF is widely used in the BCI domain to obtain frequency ranges that respond to the specific brain rhythms. For example, signals usually filtered between 8-13 Hz to obtain activity only in alpha waves, 13-30 Hz to obtain beta waves, or 8-30 Hz range to obtain alpha+beta waves (see section 1.2.2). The order of the filter is one of the characteristics of the filter and it is equal to the number of poles of the transfer function. In our studies, we filtered EEG-signals using 5th order Butterworth band-pass filter. The main feature of Butterworth filter [114] is that the frequency response is maximally flat (i.e. has no ripples) in the pass-band and rolls off towards zero in the stop-band [115]. Raw and filtered signal shown on Figure 2.2c;



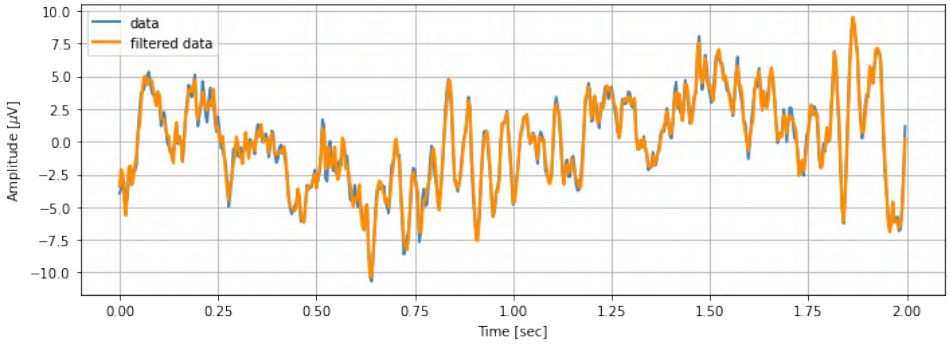
(a) Raw and filtered signal using LPF (30 Hz)



(b) Raw and filtered signal using HPF (8 Hz)



(c) Raw and filtered signal using BPF (8-30 Hz)



(d) Raw and filtered signal using BSF (49-51 Hz)

Figure 2.2: Example of raw and filtered signal of right hand MI from channel C3 using 5th order Butterworth filters. (a) LPF with cutoff frequency 30 Hz. (b) HPF with cutoff frequency 8 Hz (c) BPF with cutoff frequency range in alpha+beta band (8-30 Hz) (d) BSF with cutoff frequency range in (49-51 Hz).

- band-stop filter (BSF) is inverse to BPF: it has a large transmission factor outside the frequency range from f_{p1} to f_{p2}. The transmission factor is large starting from zero and up to f_{p1} frequency, as well as from f_{p2} to $+\infty$. Therefore, when passing a signal through such a filter, the components lying between f_{p1} and f_{p2} will be removed from its spectrum, and all other components will pass to the filter output without changes. A Band-stop filter is also known as a notch filter or band-reject filter.

  In particular, such filters (called anti-hum filters) are used to eliminate network interference when it is necessary to cut from the signal spectrum only the spectral component in a narrow range of 50 Hz or 60 Hz depending on the electric power system of the country. For countries using 60 Hz power lines: low frequency: 59 Hz, middle frequency: 60 Hz, high frequency: 61 Hz. This means that the filter passes all frequencies, except for the range of 59–61 Hz. This would be used to filter out the mains hum from the 60 Hz power line, though its higher harmonics could still be present. For countries where power transmission is at 50 Hz, the filter would have a 49–51 Hz range. Filtered signal using BSF filter with cutoff frequency (49-51 Hz) showed on Figure 2.2d;

2.1.3 Artifact removing

Signal artifacts contaminate significantly the quality of EEG data. Artifacts mainly can be grouped as environmental noise, subject's mistakes during experiments, and physiological artifacts.

Artifact removing in terms of EEG preprocessing often refers to the eye-blink removing. Furthermore, the environment artifacts and experiment error, which come from external factors, are classified as extrinsic artifacts, whereas the physiological from the body itself (e.g., eye blink, muscle activity, heartbeat) can be categorized as intrinsic artifacts. The environment artifacts can be eliminated by a simple notch filter due to the frequency of such artifacts are inconsistent with desired signals [116].

Eye movement and eye blink are the main reason for ocular artifacts. They generate a significant number of artifacts in EEG recordings. Electrooculogram (EOG) and/or eye-tracking are usually recorded to track eye movement and help to manage this type of artifact. Ocular artifacts have a low frequency (< 8 Hz), so they could be removed by a band-pass filter.

Muscle artifacts have a much larger frequency range (0 Hz to >200 Hz). They appeared because of subject talks, sniffs, swallows [117] and measured by electromyogram (EMG). Independent Component Analysis (ICA) is reported as a good method to remove this type of artifacts [118, 119].

Another type of artifact is cardiac artifacts. These artifacts usually appear when some electrode is placed over or near the blood vessel. Such artifacts are also called pulse artifacts and their frequency is near 1.2 Hz. Removal of these artifacts could be easier using the reference waveform of electrocardiogram (ECG).

There are a lot of methods used for artifact removal, such as Blind Source Separation, Wavelet transform, Principal Component Analysis, Independent Component Analysis, Canonical Correlation Analysis, Sparse Decomposition Methods, and others (see [116] for a review).

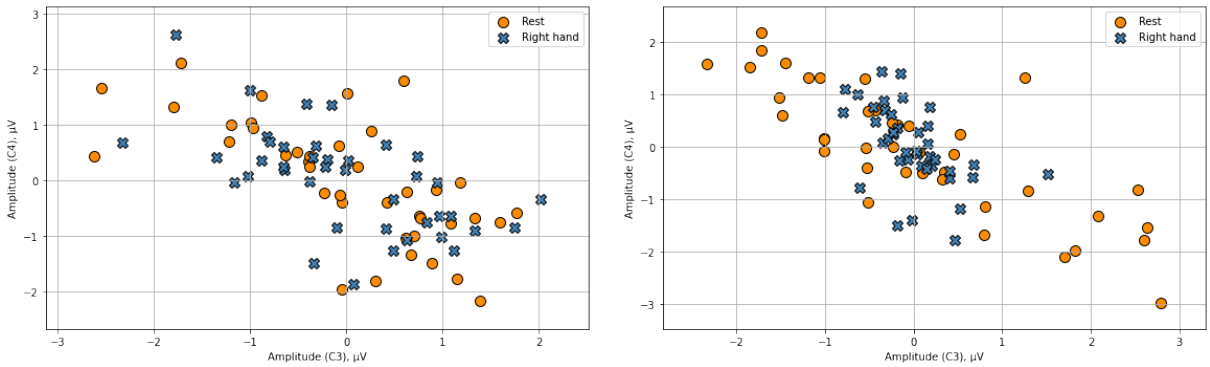
2.2 Feature extraction

Feature extraction is an important part of data preprocessing which has a strong influence on final results. In machine learning feature extraction starts from an initial set of measured variables and creates derived variables (features) intended to be more informative and non-redundant, facilitating the subsequent learning and generalization step.

Feature extraction reduces the number of resources needed to describe a large data set. One of the main problems when performing the analysis of complex data is the number of involved variables. Analysis with a large number of variables in the general case requires a large amount of memory and computing power, and it can lead to overfitting of classification, and poor generalization to new samples. Feature extraction is a general term for methods of constructing such combinations of variables to solve these problems while maintaining sufficient accuracy of the data description.

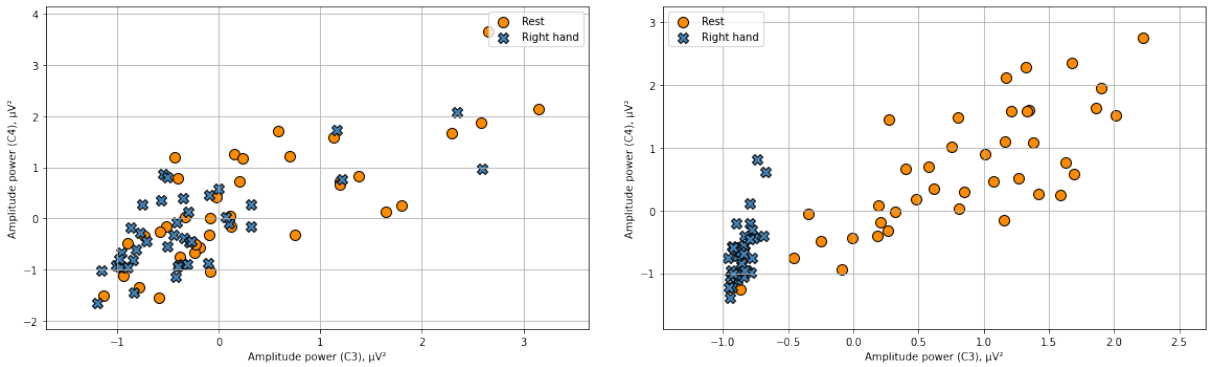
Not all feature extraction methods reduce dimensionality, some of them increase data dimension to improve class separability, others enhance signal-to-noise ratio without changing dimensionality [120].

As features for classification in BCI domain different statistics of EEG signals, power features of different frequency bands (such as ERS/ERD features), covariance matrices, etc. are usually used.



(a) Data from subject 1 **(b)** Data from subject 2

Figure 2.3: Trials representation according to average amplitude feature extraction technique. Two-classes problem with right hand MI and rest. EEG data is filtered in the alpha+beta (8-30 Hz) frequency range and averaged over time samples. Electrodes C3 and C4 are considered. In these examples, we used EEG data from two different subjects: (a) subject 1; (b) subject 2.



(a) Data from subject 1 **(b)** Data from subject 2

Figure 2.4: Trials representation according to average of signal power. Two-classes problem with right hand MI and rest. EEG data is filtered in the alpha+beta (8-30 Hz) frequency range, squared and then averaged over time samples. Electrodes C3 and C4 are considered. In these examples, we used EEG data from two different subjects: (a) subject 1; (b) subject 2.

In this section, we will compare different methods for feature extraction in the MI-based BCI domain. We will use a 2-classes problem from the dataset (section 4.2) discriminating right hand MI vs. rest as an example. Electrodes C3 and C4 are considered. For these examples, we will use data recorded from subjects 1 and 2 to show difference not only between different algorithms but also between different subjects. Firstly, let's consider EEG data, filtered in alpha+beta (8-30 Hz) frequency range and averaged over time samples duration of 2 seconds, half a second after the trigger corresponded to the start of MI. Trials representation according to such average amplitude feature extraction technique showed in Figure 2.3.

As it could be seen from Figure 2.3 classes are hardly separable using just average amplitude as a feature for both subjects.

To maximize separation between classes EEG signals could be squared before averaging (it is equal to computing signal power, this will be discussed in detail in section 2.2.2). Trials representation using the mean of signal power showed in Figure 2.4.

We can see, that in contrast to Figure 2.3b in Figure 2.4b, the classes are well separable for subject 2. However, classes for subject 1 (Figure 2.4a) remain heavily separable using this method.

2.2.1 Time-frequency analysis

2.2.1.1 Windowed Fourier transform

Spectral analysis with window functions (Fourier transform with window, window processing, windowing) is commonly used to reduce the presence of unwanted components in the signal spectrum.

To do this, before calculating the Fourier transform, the signal $x[n]$ is multiplied by some weight (window) function $w(t)$, where n is a number of samples in the signal and t is a time. Window function $w(t)$ must fall (usually to zero) at the ends of the segment in which we consider the signal. In the middle of the segment, it must have constant values.

As a result of multiplying the window by the signal, we obtain the area of the signal that is located on the time interval which is placed in the window $w(t)$. This area at the edges is slightly distorted, but in the middle of the interval coincides with the original signal. In addition, the signal values at the end and at the beginning of the interval are equal to each other (zero), which is required to get rid of the Gibbs effect [121].

The expression for the direct discrete windowed Fourier transform in this case will look like:

$$c_w[m] = \sum_{n=0}^{N-1} x[n]w[n]e^{-2\pi j \frac{m}{N} n}, \tag{2.4}$$

where, $m = 0, \ldots, N-1$ and $e^{\frac{j2\pi}{N}}$ is a primitive Nth root of 1.

Two-dimensional frequency-time representation is often useful for analyzing signals with a variable spectrum. It is necessary to be able to know both the spectral composition and the duration of the presence of components at different frequencies in the signal.

The mathematical expression for the spectral-temporal decomposition of a signal is called the Short-Time Fourier Transform. The calculation of the spectral-temporal representation of the signals is as follows. Initially, the window function $w(t)$ is placed in the neighborhood of the point $t = 0$. The Fourier spectrum of the product $x(t)w(t)$ is then calculated. Then the window function is shifted by a certain step τ_1, the function $w(t - \tau_1)$ is formed and the spectrum product $x(t)w(t-\tau_1)$ is calculated. Then the procedure is repeated for all possible offsets of the window function. The expressions of the spectral-temporal decomposition of discrete signals have the form [121]:

$$c_w[m, k] = \sum_{n=0}^{N-1} x[n]w[n-k]e^{-2\pi j \frac{m}{N} n}, \tag{2.5}$$

where $w(t), w[n]$ is window function; and τ, k is parameter of window function offset in time.

It transforms an array $x[n]$ with a d-dimensional vector of indices $n = (n_1, n_2 \ldots, n_d)$ by a set of d nested summations (over $n_i = 0 \ldots N_i - 1$ for each i), where the division $\frac{n}{N}$, defined as $\frac{n}{N} = \left(\frac{n_1}{N_1}, \ldots, \frac{n_d}{N_d}\right)$, is performed element-wise.

Representation of the power distribution of spectral components, which reflects the change in the spectrum over time, is called a spectrogram:

$$c_w[m, k] = \left| \sum_{n=0}^{N-1} x[n]w[n-k]e^{-2\pi j \frac{m}{N} n} \right|^2. \tag{2.6}$$

Therefore, it is necessary to apply time-frequency transformations for signal analysis, which allows observing the change of the spectral composition of the signal in time. To perform time-frequency analysis, the window type, window duration, window overlap, and offset step must be specified.

2.2.1.2 Wavelet transform

A wavelet transform is a tool that cuts up data or functions or operators into different frequency components, and then studies each component with a resolution matched to its scale [122]. For several decades Wavelet transform has become very popular when it comes to analysis, de-noising, and compression of signals and images [121].

So wavelet analysis is the decomposition of signals into components of different shapes and durations. Like any decomposition, the wavelet decomposition of the analyzed signal $x(t)$ aims to present it as a series - a weighted finite or infinite sum of some components:

$$x(t) = \sum_n c_n \psi_n(t), \tag{2.7}$$

where c_n the decomposition coefficient, which corresponds to the contribution of the corresponding component in the signal; ψ_n function from some set. In the case of wavelets, this set does not have to be an orthogonal basis, but can also be a semi-orthogonal, non-orthogonal basis, or dictionary of functions.

Let scale $a \in \mathbb{R}, a \neq 0$ and offset parameter $\tau \in \mathbb{R}$, using a rule:

$$\psi_{a,\tau}(t) = \left|a\right|^{-\frac{1}{2}} \psi\left(\frac{t-\tau}{a}\right) \tag{2.8}$$

we can get a set of wavelet functions $\left\{\psi_{a,\tau}(t)\right\}$ that are scalable (stretched or compressed depending on the size of the scaling factor) and shifted by argument (depending on the size τ) copies of a single generating mother wavelet function $\psi(t)$.

Graphs of some mother wavelets are shown in Figure 2.5.

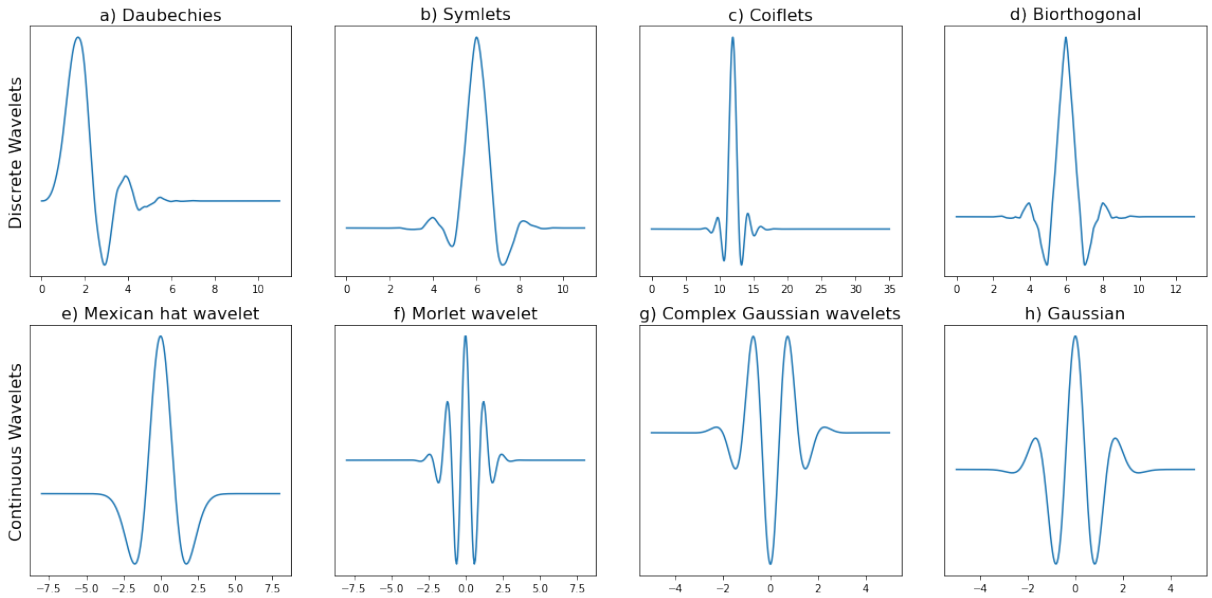


Figure 2.5: Descrete and continuous mother wavelet functions: a) Daubechie wavelet of the 6th order, b) Symlet wavelet of the 6th order, c) Coiflet wavelet of the 6th order, d) Biorthogonal wavelet $bior2.6$, e) Mexican hat wavelet, f) Morlet wavelet, g) Complex Gaussian wavelet of the 6th order, h) Gaussian wavelet of the 6th order.

The goal of signal wavelet analysis is to decompose the analyzed signal into the sum of scaled wavelet functions (2.7) [121].

The values of the corresponding decomposition coefficients are defined as the scalar products of the signal for each scaled and shifted wavelet function from the set $\left\{\psi_{a,\tau}(t)\right\}$. The Continuous Wavelet Transform (CWT) of continuous signal is presented by the following expression:

$$W_\psi(a,\tau) = \langle x(t), \psi_{a,\tau}(t)\rangle = \int_{-\infty}^{+\infty} x(t), \psi^*_{a,\tau}(t)dt = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t), \psi^*\left(\frac{t-\tau}{a}\right)dt, \tag{2.9}$$

$a, \tau \in \mathbb{R}, a \neq 0$,
where $\psi^*_{a,\tau}$ complex-conjugate mother function;
$W_\psi(a,\tau)$ - wavelet decomposition of the signal $x(t)$.

It is seen that $W_\psi(a,\tau)$ is a function of two variables: scale a and shift τ. This function shows the value of the amplitude with which the decomposition 2.7 includes each wavelet for all possible values a and τ. Such a decomposition reflects the function of one variable $x(t) \in L^2(\mathbb{R})$ into the function of two variables $W_\psi(a,\tau) \in L^2(\mathbb{R}\times\mathbb{R})$. Function (2.9) is the distribution of the signal for all possible compressions and displacements of one mother wavelet function [123].

In order to reduce the number of coefficients, parameters a and τ of wavelet functions are discretized [122, 123]. Parameters are limited, first, only to the positive values of the scale factor a. To discretize the scale parameter a is selected

$$a = a_0^m, m \in Z, a_0 > 0, a_0 \neq 1, \tag{2.10}$$

where a_0 is a fixed scaling step. To discretize the parameter of the shift τ of the mother wavelet along the time axis, values that are multiples to some fixed-step τ_0 ($\tau_0 > 0$) are chosen.

The most common case is when $a_0 = 2$ and $\tau_0 = 1$ so $\tau = na$. In this case, the family of wavelet functions will be obtained by the expression:

$$\psi_{m,n}(t) = 2^{-\frac{m}{2}}\psi\left(2^{-m}t - n\right), \qquad m,n \in \mathbb{Z}. \tag{2.11}$$

The resulting wavelets are given on the dyadic grid, ie the step of changing the scale values is a power of two, and the time offset is a multiple of two as well. The wavelet transform of the signal for this choice of coefficients is called the dyadic wavelet transform (DWT) [122, 123].

In this case, instead of a continuous wavelet transform as a continuous function of two arguments (scale and offset), we have a finite number of values:

$$D_{m,n} = \langle x(t), \psi_{m,n}(t)\rangle = \int_{-\infty}^{+\infty} x(t)\psi_{m,n}(t)dt \tag{2.12}$$

Equation (2.12) determine the direct orthogonal dyadic wavelet transform of a continuous signal.

It should be noted that the concept of "discrete" or "dyadic" in the wavelet transform refers only to the method of selecting the scaling factors, which will not now change continuously but should take only certain values. It is possible to speak about both CWT, and about DWT of continuous or discrete signals [121].

If the signal is discrete, then the scaling factors can still take any value to conduct its CWT. The offset can also be arbitrary. But it is necessary to replace the integrals with sums in the expressions for direct and inverse CWT .

To perform DWT of the discrete signal, all of the above is also true, but the scale and offset are dyadic [121].

2.2.2 Band power features

One of the most widely used methods in EEG signal processing is based on extracting band power features. These features represent the average power of a signal in a specific frequency ranges.

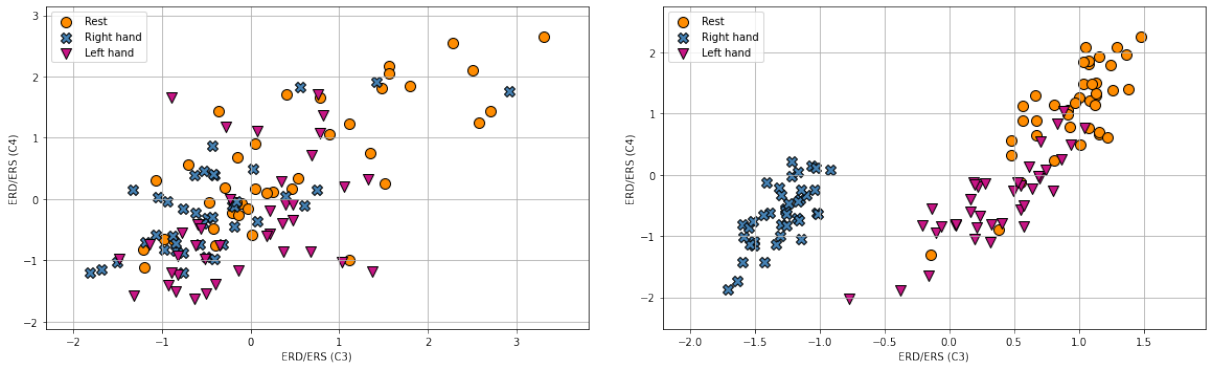
This method consists of signal decomposition into functionally different frequency ranges, such as delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-100 Hz).

This involves the decomposition of the EEG signal into frequency components, which is usually obtained using a Fourier transform. The most popular algorithm to compute the Fourier transform is the Fast Fourier Transform (FFT), which returns, for each frequency bin, a complex number from which amplitude and phase of the signal at that a specific frequency can be extracted. Magnitude-squared of the FFT is used to obtain the power spectral density.

The average band power consists of computing a single number that summarizes the contribution of a given frequency band to the overall power of the signal. This could be particularly useful to summarize a particular aspect of the data.

2.2.2.1 Event-related synchronization and desynchronization

A common method for feature extraction in BCI is event-related desynchronization/synchronization (ERD/ERS). It is used to evaluate the modulation produced by the tasks. ERD/ERS can be computed using the band power method [49]. Using this measure it possible to see decreasing and increasing oscillations in alpha and beta bands



(a) Data from subject 1

(b) Data from subject 2

Figure 2.6: Motor imagery of the right hand, left hand and rest class are represented. Logarithmic ERD/ERS features are computed for EEG channels C3 and C4, which are located above areas of the brain responsible for right and left hand activity respectively. In these examples, we used EEG data from two different subjects: (a) subject 1; (b) subject 2.

which take place in localized regions over sensorimotor cortices. Event-related desynchronization (ERD) can be observed during and even before sensory and cognitive processing and motor behavior of the brain [99, 100]. The opposite phenomenon event-related synchronization (ERS) is characterized by increasing power in alpha and beta bands accurate after imaginary or real movement. It was described in more detail in the section 1.2.5.

The filtered signal is squared for each trial and averaged over trials. Then it is smoothed using a 2-second sliding window with a 150-250 ms shifting step. We have chosen a specific sliding window because the nature of the real and imagined movement, as well as the components ERD/ERS that underline them, require a short window. Finally, the averaged power computed for each window was subtracted and then divided by the averaged power of a baseline corresponding 2 seconds before each trial. This transformation was multiplied by 100 to obtain percentages. This process can be summarized by the following equation:

$$ERD/ERS\% = \frac{\overline{x^2} - \overline{BL^2}}{\overline{BL^2}} \times 100 \ , \tag{2.13}$$

where $\overline{x^2}$ is the average of the squared signal over all trials and samples of the studied window, $\overline{BL^2}$ is the mean of a baseline segment taken at the beginning of the corresponding trial, and ERD/ERS% is the percentage of the oscillatory power estimated for each step of the sliding window. It is done for all channels separately.

ERD and ERS are difficult to observe from the EEG signal. Indeed, an EEG signal expresses the combination of activities from several neuronal sources. One of the most effective and accurate techniques used to extract events is the average technique [124]. We decided to use this technique to represent the modulation of the power of the Alpha and Beta rhythms.

Logarithmic ERD/ERS also could be used as feature, it can be presented by the following equation:

$$ERD/ERS = \log\left(\overline{x^2} + 1\right) \tag{2.14}$$

We introduce a 3-classes problem using motor imageries of the right hand, left hand, and rest class from the dataset (section 4.2). We have used EEG data for subjects 1 and 2 in this example. Figure 2.6 shows logarithmic ERD/ERS features, computed for EEG channels C3 and C4, which are located above areas of the brain responsible for right and left hand activity respectively.

ERD is observed during and immediately after the motor imagery task and better represented in the alpha band, while ERS is observed after the task and better expressed in the beta band. That's why features filtered in alpha bands (ERD) were calculated during 0.5-6 seconds of each trial, and features filtered in beta bands (ERS) were calculated beginning with 7 and till 12 sec.

2.2.3 Spatial analysis

Spatial analysis is another group of frequently used techniques for feature extraction and dimensional reduction in the MI-based BCI domain. In this section, we will introduce covariance matrices as one of the ways of

representing EEG signals. Covariance matrices are used in the common spatial pattern method, its adaptations, and in Riemannian geometry-based methods. We will present methods like common spatial pattern and filter bank common spatial pattern also along this section. These techniques are standard benchmark methods for feature extraction in BCI.

2.2.3.1 Covariance matrix

A covariance matrix is a matrix made up of pairwise covariances of elements of one or two random vectors. Let $\mathrm{x} \in \mathbb{R}^n$ be a multidimensional random vector containing scalar random variables.

$$\mathrm{x} = \begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \\ \vdots \\ \mathrm{x}_n \end{bmatrix} \tag{2.15}$$

The covariance matrix $C_{i,j}$ of a random vector x is a square symmetric positive semi-definite matrix, on the diagonal of which the variances of the vector components are located, and the off-diagonal elements in the (i, j) positions are the covariances between the ith and jth elements of x.

$$C_{i,j} = cov(\mathrm{x}_i, \mathrm{x}_j) = E\Big[(\mathrm{x}_i - \mu_i)(\mathrm{x}_j - \mu_j)^\top\Big], \tag{2.16}$$

where $\mu_i = E[x_i]$ is the expected value of the ith element of vector x. Thus the equation 2.16 takes the form:

$$C = \begin{bmatrix} E\Big[(\mathrm{x}_1 - \mu_1)(\mathrm{x}_1 - \mu_1)\Big] & E\Big[(\mathrm{x}_1 - \mu_1)(\mathrm{x}_2 - \mu_2)\Big] & \cdots & E\Big[(\mathrm{x}_1 - \mu_1)(\mathrm{x}_n - \mu_n)\Big] \\ E\Big[(\mathrm{x}_2 - \mu_2)(\mathrm{x}_1 - \mu_1)\Big] & E\Big[(\mathrm{x}_2 - \mu_2)(\mathrm{x}_2 - \mu_2)\Big] & \cdots & E\Big[(\mathrm{x}_2 - \mu_2)(\mathrm{x}_n - \mu_n)\Big] \\ \vdots & \vdots & \ddots & \vdots \\ E\Big[(\mathrm{x}_n - \mu_n)(\mathrm{x}_1 - \mu_1)\Big] & E\Big[(\mathrm{x}_n - \mu_n)(\mathrm{x}_2 - \mu_2)\Big] & \cdots & E\Big[(\mathrm{x}_n - \mu_n)(\mathrm{x}_n - \mu_n)\Big] \end{bmatrix} \tag{2.17}$$

As every covariance matrix is positive semi-definite, it means that their eigenvalues are non-negative.

The covariance matrix of a random vector is a multidimensional analog of the variance of a random variable for random vectors. The covariance matrix of two random vectors is a multidimensional analog of the covariance between two random variables.

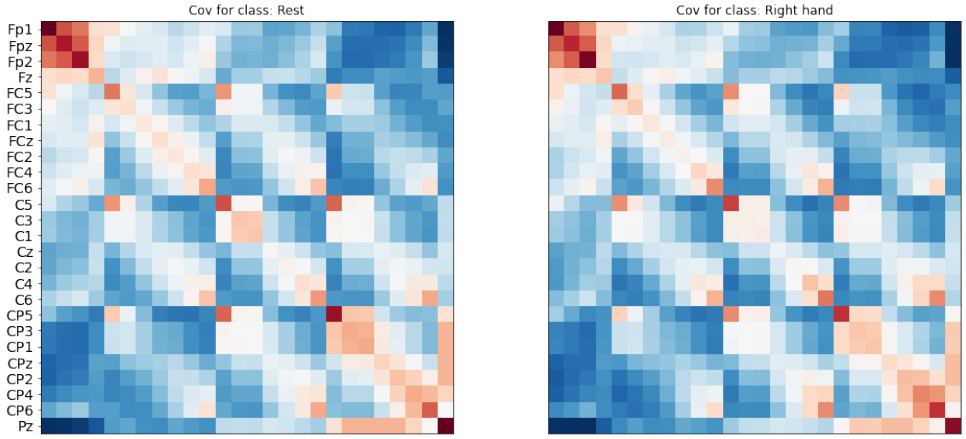
In the case of a normally distributed random vector, the covariance matrix together with the expected value (average) of this vector completely determine its distribution.

Figures 2.7a and 2.7b represents covariance matrices for two different subjects. For these examples we used EEG data from subject 1 (Figure 2.7a) and subject 2 (Figure 2.7b) from the dataset described in section 4.2. Data for two classes (rest and right hand MI) were used. We have calculated covariance matrices from filtered (8-30 Hz) EEG signals containing 26 channels. From Figure 2.7b we could see that trials from different classes has different structure of covariance matrices for subject 2. But for subject 1 difference is not so notable (Figure 2.7a).

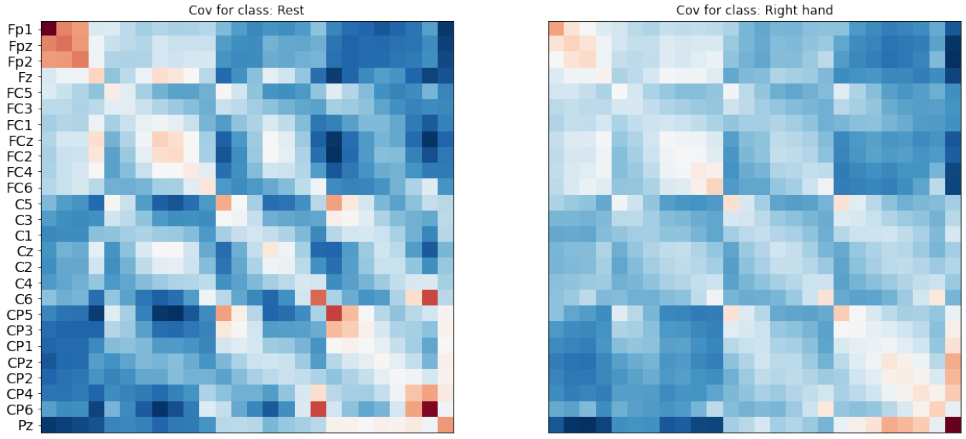
2.2.3.2 Common Spatial Pattern (CSP)

An algorithm called Common Spatial Pattern (CSP) is the most commonly used technique to extract features from EEG signals in the BCI domain. This algorithm generates a series of spatial filters that applies to decompose multi-dimensional data into a set of uncorrelated components [19]. These filters aim to extract features that simultaneously maximize the variance of one class while minimizing the variance of the other one [125, 19].

CSP is a supervised decomposition technique for feature extraction in a two-class decoding problem for multichannel signals. The CSP algorithm considers the whole set of electrodes that are placed during the recording session, which normally includes those covering the sensorimotor cortex. The number of resulting filters equals the number of EEG channels and, in most of the cases, at least two pairs (i.e., four filters) are selected. All trials have to be previously filtered between the selected frequency range.



(a) Data from subject 1



(b) Data from subject 2

Figure 2.7: Covariance matrices for two classes (rest and right hand MI) representing EEG data (26 channels). Signals are previously filtered in the alpha+beta (8-30 Hz) frequency range. In these examples, we used EEG data from two different subjects: (a) subject 1; (b) subject 2.

CSP is very useful for motor imagery BCIs (based on oscillatory activity). Usually, the signal is previously filtered in alpha and beta bands before CSP application typically between 8-30 Hz [125].

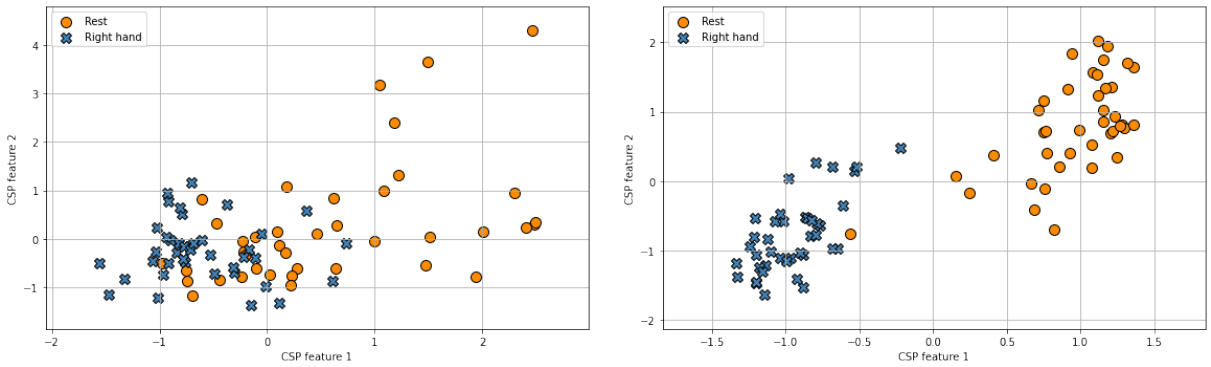
First step to compute CSP is to find covariance matrices (see Section 2.2.3.1) of filtered signals X:

$$C_i^k = \frac{1}{t-1} X_i^k (X_i^k)^\top \tag{2.18}$$

where C_i^k is the spatial covariance matrix for a band-pass filtered trial i of signal $X_i^k \in \mathbb{R}^{n \times t}$ from class k (with samples t as columns and channels n as rows); with $^\top$ representing the transpose. Then means of the covariance matrices $\overline{C^1}$ and $\overline{C^2}$ are computed by averaging each of these matrices over corresponding trials. Note that, the original paper describes a step for covariance matrix trace normalization to eliminate magnitude variations in the EEG between subjects [126]. But it is not applied in more recent works [19, 127]. Then CSP analysis is given by the simultaneous diagonalization of two covariance matrices. In other words, we should find a single invertible matrix W such that $W^\top \overline{C^k} W$ is a diagonal matrix for every $\overline{C^k}$. Technically, this can easily be computed by solving the following generalized eigenvalue decomposition problem [19]:

$$\overline{C^1} W = \lambda \overline{C^2} W \tag{2.19}$$

where $\overline{C^1}$ and $\overline{C^2}$ are n-by-n covariance matrices, λ is the diagonal matrix of generalized eigenvalues of $\overline{C^1}$



(a) Data from subject 1

(b) Data from subject 2

Figure 2.8: Trials representation according to CSP feature extraction technique. Two-classes problem with right hand MI and rest. EEG data (26 channels) from the dataset 4.2 is filtered in the alpha+beta (8-30 Hz) frequency range, and decomposed to 2 CSP components. In these examples, we used EEG data from two different subjects: (a) subject 1; (b) subject 2.

and W is a matrix whose columns are eigenvectors. These column vectors of W represent spatial filters, and the number of columns equals the number of channels in the EEG signal. The spatially filtered signal X^k_{CSP} can be obtained from the original EEG signal X^k as

$$X^k_{CSP} = W^\top X^k \tag{2.20}$$

Note that the CSP filters are paired, so, after being sorted according to the values of λ, only the first m and the last m columns of W are selected. Typically, 6 filters (i.e., 3 pairs) are used corresponding to the 3 largest and 3 lowest eigenvalues. But not all of the achieved filters are relevant for the following classification. Once selected filters W are obtained, a CSP features f for band-pass EEG signal X can be computed as follows [128]:

$$f = log(WXX^\top W^\top) = log(WCW^\top) = log(var(WX)) \tag{2.21}$$

CSP requires more channels than spatial filters such as Bipolar or Laplacian, however in practice it usually achieved higher classification performance [125].

Figure 2.8 shows the application of CSP we have done on the EEG data from the dataset (section 4.2) with 2-class problem (right hand MI vs. rest). The same data from subjects 1 and 2 as considered earlier for Figure 2.3. From Figure 2.8 we could see that rest and right hand MI classes are better separated than before using just average (Figure 2.3) or average of power (Figure 2.4) for both subjects. However, for subject 2 (Figure 2.8b) we can more easily separate both classes than for subject 1 (Figure 2.8a).

The CSP algorithm has numerous advantages: it leads to high classification performance, it is flexible, computationally efficient, and simple to implement. This makes CSP one of the most popular approaches for the oscillatory activity-based BCI [19, 128]. Though despite all these advantages CSP has flaws and limitations. Notably it has been shown that CSP is non-robust to noise, to non-stationarities and prone to over-fitting when small training set is available [127, 129, 130]. Also, CSP identifies only the important spatial information but does not consider the spectral one. Moreover, the CSP algorithm has a limitation of classes number. It is applicable only to discriminate between two classes. That's why this algorithm can not be used to solve multi-class problems. Though there are some multi-class extensions [127].

CSP algorithm requires parameters to setups, such as frequency range for band-pass filtering, the time interval of the signal trial to use for decomposition, and the number of CSP filters to use. Most commonly for motor intention detection tasks in BCI domain frequency range that includes both alpha and beta bands (8-30 Hz) is used [125]. Time segment usually is taken after 0.5-1 second after the cue, and 2-3 pairs of filters (4-6 filters) are taken into account [18].

2.2.3.3 Filter Bank Common Spatial Pattern (FBCSP)

The goal of the Filter Bank Common Spatial Pattern (FBCSP) algorithm is to improve the standard CSP method taken into account specific parameters and profiles of the subjects. This is achieved by automatic selection of the

most discriminative subject-specific frequency range for band-pass filtering [131].

The FBCSP algorithm includes 4 different stages:

1. Band-Pass Filtering.

   In this stage, a filter bank is used to decompose EEG signals into different frequency bands. According to the original algorithm [131] nine 4 Hz wide band-pass filters are used: 4-8, 8-12, ..., 36-40 Hz. These filters totally cover the frequency range from 4 to 40 Hz which corresponds to theta, alpha, beta, and low delta brain rhythms. The number and range of band-pass filters could be changed for a specific task.

2. Spatial Filtering.

   On the second stage CSP algorithm is used to employ spatial filtering. CSP filters apply to the signals filtered in the 9 frequency ranges obtained in the previous step. Thus the equation 2.20 takes the form:

   $$X_{b,i}^{FBCSP} = W_b^\top X_{b,i}, \tag{2.22}$$

   where $X_{b,i}$ is an i-th trial of the EEG signal filtered in the b-th frequency band. $X_{b,i}^{FBCSP} \in \mathbb{R}^{n \times t}$ denotes $X_{b,i}$ after spatial filtering, $W_b \in \mathbb{R}^{n \times n}$ is a CSP filters coefficient matrix, where n is a number of EEG channels, t is a number of time samples per channel and $\top$ denotes transpose operator.

   Then transformation matrix W_b with optimal coefficients for discriminating 2 classes of EEG data is computed by solving eigenvalue decomposition problem (equation 2.19) as showed in section 2.2.3.2.

   The FBCSP feature vector for the i-th trial is then formed as follows:

   $$v_i = [v_{1,i}, v_{2,i}, ..., v_{9,i}], \tag{2.23}$$

   where $v_i \in \mathbb{R}^{1 \times (9*2m)}$, $i = 1, 2, ..., n$; n denotes the total number of trials in the data; m is number of CSP filters pairs which corresponds to the first m and the last m columns of W_b.

3. Feature selection.

   On the third stage a feature selection technique applied to select discriminative CSP features from equation 2.23. Different feature selection algorithms can be used in this step.

4. Classification.

   In the final stage classification algorithm is applied to the selected CSP features from the previous step. Here there are also many choices, such as linear discriminant analysis algorithm or support vector machine classifier. Both methods will be represented later in this chapter.

2.2.4 Riemannian geometry

As covariance matrices are symmetric positive-definite, this means that they belong to a Riemannian manifold from where their classification can be directly reached. Square real symmetric positive-definite (SPD) matrices have the characteristic of being always diagonalizable with strictly real positive eigenvalues, whose space represents a differentiable Riemannian manifold M. There are several advantages of using this manifold that can be easily applied in feature extraction and classification [23].

In differential geometry, a N-dimensional smooth manifold is a topological space that is locally similar to the Euclidean space and has a globally defined differential structure. A smooth Riemannian manifold (or Riemannian space) M is a real smooth manifold equipped with an inner product on the tangent space $T_\Omega M$ defined at each point Ω that varies smoothly from point to point. The tangent space $T_\Omega M$ at point Ω is the Euclidean vector space containing the tangent vectors to all curves on M passing through Ω (Figure 2.11) [132].

Riemannian geometry has a specific measure of the distance between positive semi-definite matrices. This measure will be used to find the closest distance between the covariance matrix describing the window under study and the averaged covariance matrix of each class, e.g., rest and motor imagery.

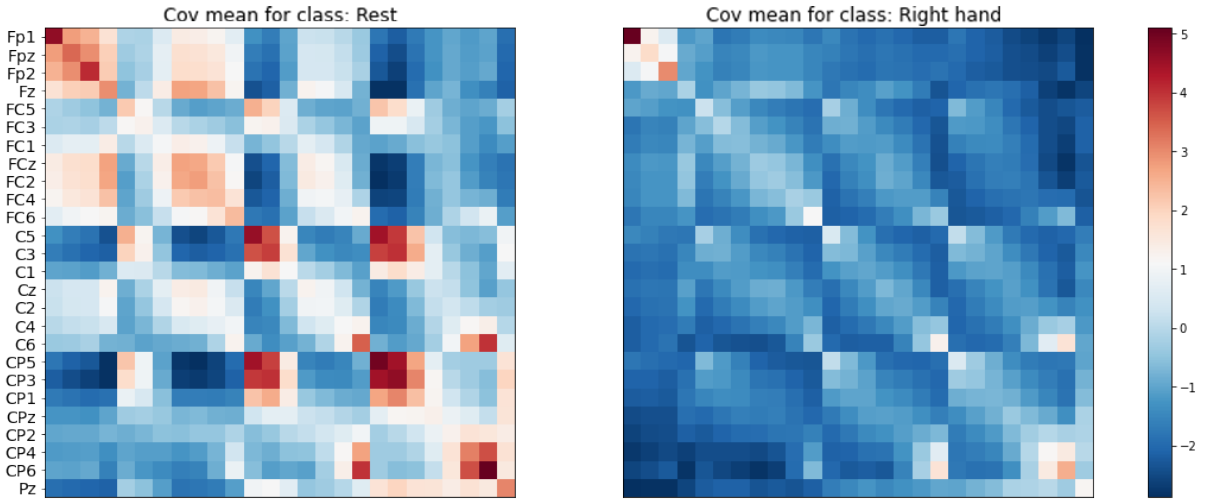


Figure 2.9: Centers of mass for two classes (rest and right hand MI) representing EEG data (26 channels). Signals are preliminary filtered in alpha+beta (8-30 Hz) frequency range.

2.2.4.1 Riemannian distance

The minimum length curve connecting two points in the manifold M called the geodesic $\gamma(t)$ and the distance of this curve is Riemannian distance. It is one of the basic metrics in the Riemannian Geometry, also called Affine Invariant Riemannian Metric (AIRM), and it computes as follows:

$$d_R(C_1, C_2) = ||\log C_1^{-1} C_2||_F = ||\log C_1^{-\frac{1}{2}} C_2 C_1^{-\frac{1}{2}}||_F, \tag{2.24}$$

where C_i are covariance matrices (2.18) distance between which is computed, λ_i are eigenvalues of $C_1^{-1}C_2$ and F denotes Frechet norm [132, 21].

2.2.4.2 Riemannian centre of mass

To be able to compare distances between particular trials and certain classes we can compute Riemannian centre of mass for each class in Riemannian space. Riemannian (Frechet) mean or Riemannian centre of mass is computed using equation 2.25. Let's say it will be point Ω (Figure 2.11) and SPD covariance matrix in this point is denoted as $\overline{C}$.

$$\overline{C} = \arg\min_C (\frac{1}{I} \sum_{i=1}^{I} d_R^2(C, C_i)), \tag{2.25}$$

where I is a number of trials of the observed class.

Figure 2.9 shows centers of mass for two classes (rest and right hand MI) of subject 2. We have considered 26-channel EEG data from the dataset described in section 4.2.

Centers of mass for the same subject but only for two EEG channels (C3 and C4) represented in Figure 2.10.

In both cases, it is clear that each class yields a different structure of the covariance matrix.

2.2.4.3 Tangent space

To be able to use linear classifiers such as LDA or SVM it is necessary to have a feature vector. The next algorithm shows how to get a feature vector from Riemannian space projecting covariances matrices on the tangent space of the current covariance matrix. Consider a point Ω on M and construct the tangent space $T_\Omega M$ on it. Now take a tangent vector V departing from Ω, which is our reference point. There exists one and only one geodesic on the manifold starting at Ω that corresponds to V; think of rolling the plane (tangent space) on the surface (manifold) in such a way that the vector always touches the surface. The endpoint on M is Φ. We see that the geodesics on

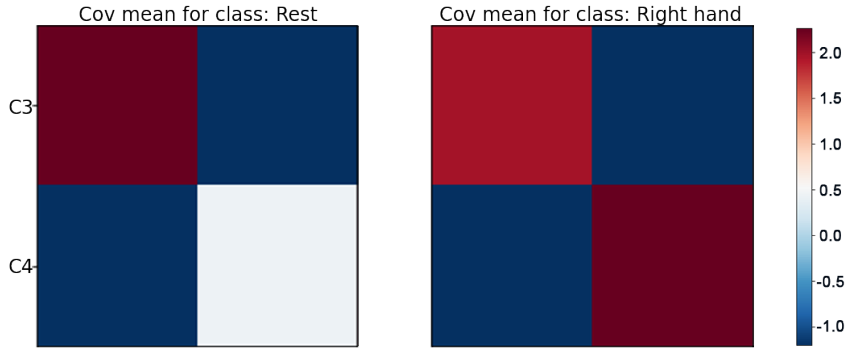


Figure 2.10: Centers of mass for two classes (rest and right hand MI) representing EEG data (from channels C3 and C4). Signals are previously filtered in alpha+beta (8-30 Hz) frequency range.

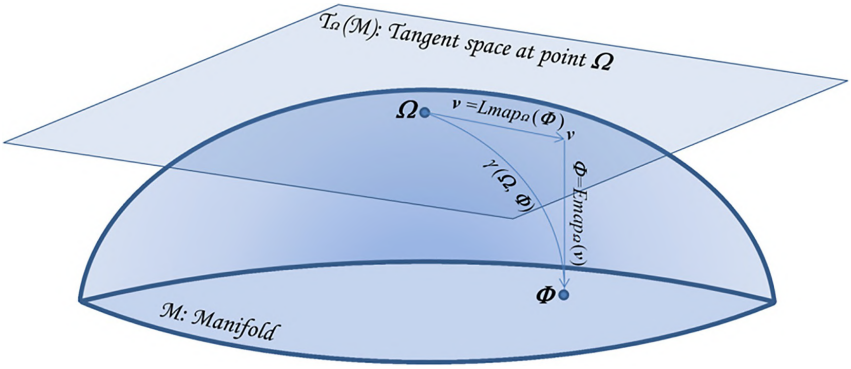


Figure 2.11: The Manifold and the tangent space at a point. Consider a point Ω on M and construct the tangent space $T_\Omega M$ on it. Now take a tangent vector V departing from Ω, which is our reference point. There exists one and only one geodesic on the manifold starting at Ω that corresponds to V; think of rolling the plane (tangent space) on the surface (manifold) in such a way that the vector always touches the surface. The endpoint on M is Φ. We see that the geodesics on M through Ω are transformed into straight lines and the distances along all geodesics are preserved (from [132, 133]).

M through Ω are transformed into straight lines and the distances along all geodesics are preserved (this is true in the neighborhood of Ω) [132, 133].

In this case the shortest curve between two points Ω and Φ on M is called the geodesic $\gamma(\Omega \rightarrow \Phi)$ and computes as follows:

$$\gamma(\Omega \rightarrow \Phi) = \Omega^{\frac{1}{2}}(\Omega^{-\frac{1}{2}}\Phi\Omega^{-\frac{1}{2}})^{\beta}\Omega^{\frac{1}{2}}, \beta \in [0,1], \tag{2.26}$$

where β is the arc length parameter. When $\beta = 0$ we are at Ω, when $\beta = 1$ we are at Φ and when $\beta = 0.5$ we are at the geometric mean of the two points[132].

2.2.4.4 The exponential and logarithmic maps

The exponential and logarithmic maps are shown graphically in Figure 2.11. The function that maps a vector $V \in T_\Omega M$ to the point $\Phi \in M$ following the geodesic starting at Ω, is named the exponential map and denoted by $\Phi = Emap_\Omega(V)$ [132]. It is defined as:

$$\Phi = Emap_\Omega(V) = \Omega^{\frac{1}{2}} exp(\Omega^{-\frac{1}{2}} V \Omega^{-\frac{1}{2}})\Omega^{\frac{1}{2}} \tag{2.27}$$

The inverse operation is the function mapping the geodesic relying Ω to Φ back to the tangent vector $V \in T_\Omega M$. It is named the logarithmic map and denoted $V = Lmap_\Omega(\Phi)$ [132]. It is defined as:

$$V = Lmap_\Omega(\Phi) = \Omega^{\frac{1}{2}} ln(\Omega^{-\frac{1}{2}}\Phi\Omega^{-\frac{1}{2}})\Omega^{\frac{1}{2}} \tag{2.28}$$

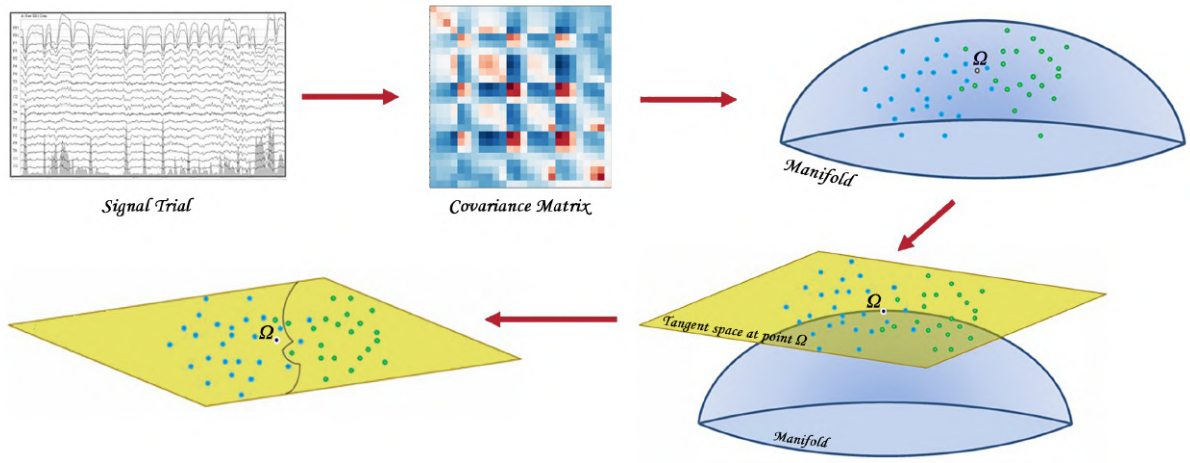


Figure 2.12: Classification on the tangent space (from [134]).

So each covariance matrix (trial) could be projected to tangent space at some point Ω using equation 2.29. Then S_i can be vectorized to turns it into Euclidean R^n space where the basis is orthonormal. This transformation also preserves the distances between the computed center of mass C (identity on the tangent space) and each covariance matrix C_i (but not the distances between covariance matrices). After finishing these manipulations linear classifier (as LDA, LR or SVM) can be applied (Figure 2.12). Vectorized S_i can be used as an input to these classifiers.

So taking into account equation 2.28, tangent space projection of C_i at $\overline{C}$ will be defined as:

$$S_i = \overline{C}^{\frac{1}{2}} ln(\overline{C}^{-\frac{1}{2}} C_i \overline{C}^{-\frac{1}{2}}) \overline{C}^{\frac{1}{2}} \tag{2.29}$$

And according to equation 2.27 to projecting S_i back to the manifold next equation should be used:

$$C_i = \overline{C}^{\frac{1}{2}} exp(\overline{C}^{-\frac{1}{2}} S_i \overline{C}^{-\frac{1}{2}}) \overline{C}^{\frac{1}{2}} \tag{2.30}$$

Using equation 2.29 trials and centers of mass (represented by SPD matrices) could be projected on the tangent space.

Figure 2.13 shows trials and centers of mass for 2-channel EEG data (previously introduced in Figure 2.10) represented in 3-dimensional tangent space. We used EEG data from 2 channels (C3 and C4) of subject 2 from the dataset described in section 4.2. We have previously filtered signals in alpha+beta (8-30 Hz) frequency range. Now we could apply linear classifiers in this space to separate classes.

2.3 Feature selection

When the input of the algorithm is too large to be processed and is suspected to be redundant (for example, the same measurements in meters and feet, or repeatability in images represented by pixels), then they can be converted to a shortened set of features (also called a feature vector). Determining a subset of initial features is called feature selection. It is expected that the selected features contain relevant information from the input data so that the desired task can be performed using this reduced representation instead of the complete primary data.

Feature selection, also known as variable selection, attribute selection, or variable subset selection, is the process of selecting a subset of relevant features for use in model construction. Feature selection techniques are used for four reasons:

- simplification of models to make them easier to interpret [135];
- shorter training times;
- to avoid the curse of dimensionality;

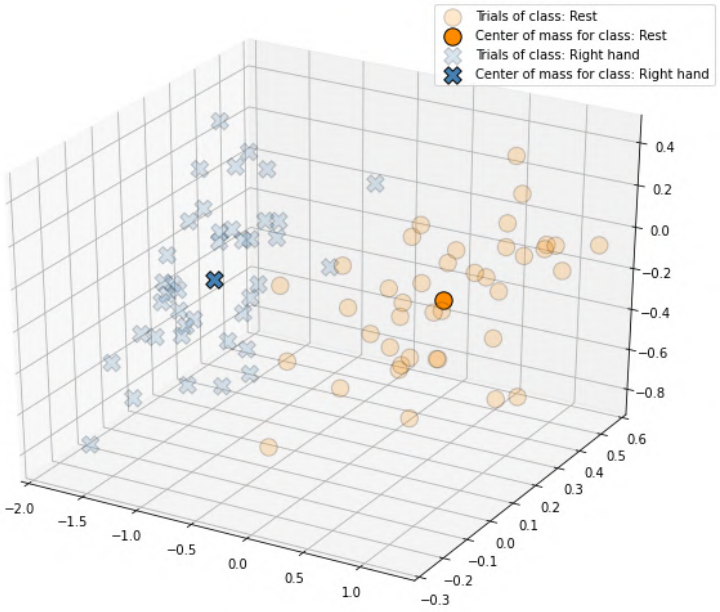


Figure 2.13: Centers of mass and trials for two classes (rest and right hand MI) represented in tangent space. We used EEG data from 2 channels (C3 and C4). We have previously filtered signals in alpha+beta (8-30 Hz) frequency range.

- enhanced generalization by reducing overfitting [136] (formally, reduction of variance)

In the thesis we used the method which accepts a threshold parameter and selects the features whose importance (defined by the coefficients) are above this threshold. For this purpose, we select a meta-transformer SelectFromModel, realized in Python library Scikit-learn for feature selection in section 5.2.3. This algorithm could be used along with any estimator that has a coefficients or feature importance's attribute after fitting. The features are considered unimportant and removed if the corresponding coefficients or feature importance's values are below the provided threshold parameter.

In section 5.2.3 we selected a threshold value equals to $1e-5$. A threshold used to demarcate features: it kept features whose importance is greater or equal compared to the threshold while the others are discarded.

In the thesis, we use linear classifiers with L1 penalized parameter, so many estimated coefficients equal to zero. In this case, SelectFromModel selects the non-zero coefficients to reduce the dimensionality of the data.

We will apply feature selection to investigate variability between bands in section 5.2.3. It can help to understand which features (bands) are more useful for the classification.

2.4 Outlier detection

Outlier or anomaly detection is the identification of elements that do not match the expected other elements or patterns of the dataset. This technique widely used in detecting fraud among banking operations, text errors, structural defects, and medical problems [137].

Four outlier detection methods were investigated in this work: local outlier factor [138], Isolation forest [139], one-class SVM [140], and elliptic envelope method (or robust covariance) [141]. Figure 2.14 presents the application of these methods on ERD/ERS features extracted from the right hand MI class. For the illustrative purpose in this section, we computed results only on EEG signals for subject 2 from the dataset described in section 4.1.

One common way to detect outliers is to assume that the data comes from a known distribution (e.g., a Gaussian distribution). From this assumption, the shape of the data distribution density is determined, and remote observations that go far beyond the shape of the distribution density are defined as outliers.

Local Outlier Factor (LOF) is an unsupervised algorithm for outlier detection [138]. It assigns some score to each observed sample. This score measures the local deviation of the density of a given sample with a respect to

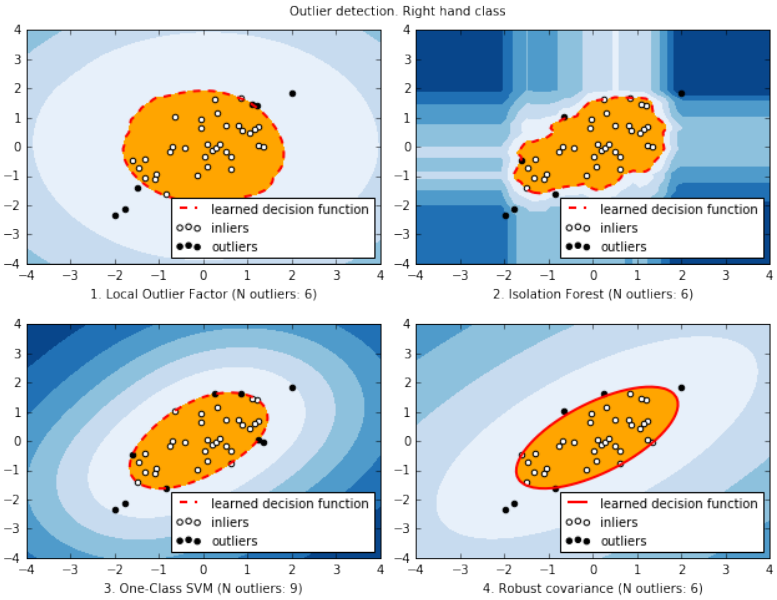


Figure 2.14: Results of outlier detection for right hand MI using different methods: local outlier factor, Isolation forest, one-class SVM, robust covariance.

its neighbors and is called the Local Outlier Factor. K-nearest neighbors estimate the local density and determine the degree of isolation of an object from its neighborhood. Then LOF compares the local densities of a sample and its neighbors detecting samples that have a significantly lower density than others. These samples are considered outliers.

The elliptical envelope method (or robust covariance) [141] determines the covariance estimate of the data so that the central data points are in the ellipse, ignoring points (outliers) outside the ellipse (see Figure 2.14.d).

One-class implementation of the Support Vector Machine (SVM) method, on the contrary, works better when the data distribution is not strongly Gaussian because this algorithm can closely approximate the distribution of the data set. The one-class SVM method detects the soft boundary of a given data set and classifies whether new observations belong inside the boundary or not [140]. To determine the boundary, it is needed to select the core and the parameter responsible for the probable saturation of the data with outliers. A RBF (radial basis function) kernel can be used (see Figure 2.14.c).

The Isolation Forest (IF) algorithm [139] distinguishes (isolates) observations by randomly selecting a feature and a split value between the maximum and minimum values of this feature. As recursive separation can be represented as a tree, the number of splits required to allocate observations is equivalent to the path length from the root node to the end node. Then the estimation is calculated as the path length to isolate the observation. The length of this path, averaged over all such random trees, determines the decision function. Random separation creates noticeably shorter paths for anomalies. Thus, when the forest of random trees often reveals shorter path lengths for the same observations, they are probably anomalies (see Figure 2.14.b). There are few parameters to set the Isolation Forest algorithm such as the number of base estimators in the ensemble and maximum features parameter, which defines the number of features to draw from the data to train each base estimator. Also, the contamination parameter defines the proportion of outliers in the dataset as well as the saturation parameter for One-class SVM.

Isolation Forest with the number of trees equal to 100 and the maximum number of features to train each base estimator chosen equal to five.

2.4.1 Improving the model performance using outlier detection as quality index of imaginary movements estimation

To improve BCI training and classification performance we propose to detect outliers in the data and exclude them before training.

For each class, outliers are detected in the training set and removed using different saturation parameter. Saturation is a parameter responsible for the percentage of data that will be selected as outliers. Outlier detection is based on one of the clustering algorithms presented in section 2.4.

After outlier detection logistic regression is used as a classification method. For each outlier saturation parameter, classification performance is computed. The parameter with the maximum classification accuracy is selected. This is followed by the learning of outlier detectors (clustering models) with the best parameter selected at the previous step for each class. We trained outlier detectors on the training set, saved them and then used for cleaning the test set.

Since the true class of the test data is unknown, it is not known which of the anomaly detectors to use. Therefore, we have proposed to use the method of logical disjunction: for an observation to be marked as an anomaly, it must be marked as an anomaly by each outlier detector. After removing the anomalies, classifiers predict classes on the cleared test set.

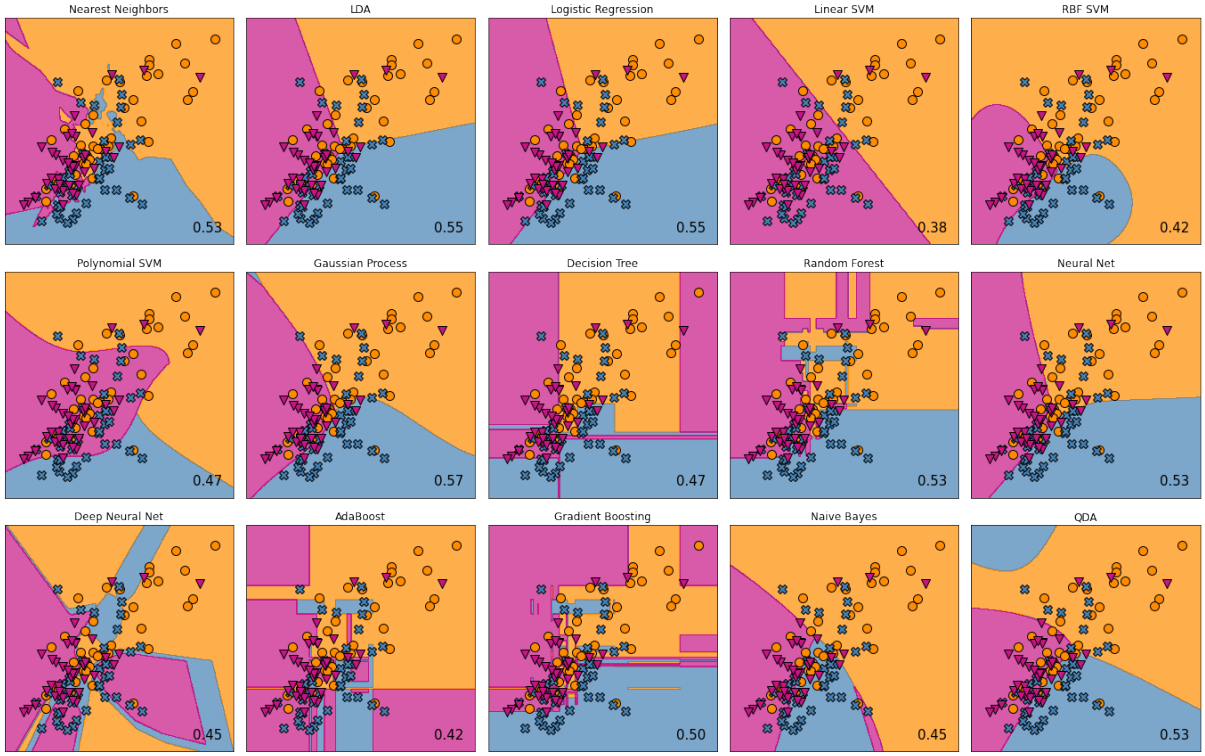
Elliptical envelope and local outlier factor methods were prone to overfitting and showed very bad generalization on our high dimensional data. As a result, the models trained with these methods removed about 90% of the data from the test set, considering them as outliers. Therefore in section 5.1, we will present results obtained using Isolation Forest and One-class SVM algorithms because these methods obtained the best performances on our datasets.

2.5 Classification

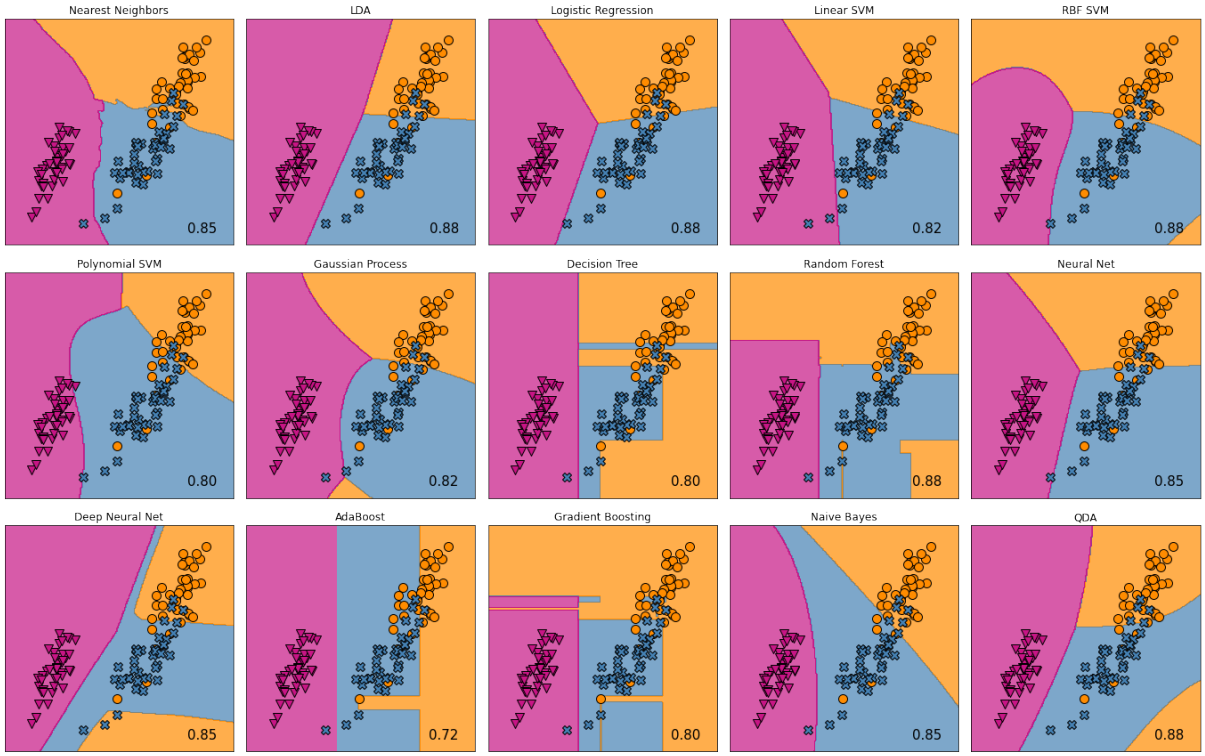
Classification is an instance of supervised learning. It is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations whose category membership is known.

Many different classification methods exists (see Figure 2.15). Some of classification techniques that were used for our problems will be described below. Here again, for illustrative purpose, we will consider EEG signals for subjects 1 and 2 from the dataset (section 4.2). We used data for three different classes: right hand MI, left hand MI and rest. We filtered signals previously in the alpha+beta (8-30 Hz) frequency range and ERD/ERS features from two channels C3 and C4 were computed.

From Figure 2.15a we could see that all presented classifiers straggle separating 3 classes for subject 1 having average classification accuracy around 50-55 %. At the same time, for subject 2, classes are well separated by many classification algorithms (Figure 2.15b) having classification accuracy around 82-88 %. The best classifiers are linear ones, such as linear discriminant analysis and logistic regression, support vector machines with RBF kernel, and neural networks. However, low results of classification for subject 1 and the huge difference between results for two subjects shows that another approach should be found to improve both feature extraction and classification stages. For this purpose method based on parallel connection of binary classifiers will be presented (section 2.5.6) together with classic machine learning (sections 2.5.1-2.5.4) and Riemannian geometry approaches (section 2.2.4.3, 2.5.5) along this section. Moreover, in chapter 3, we will consider deep learning methods such as convolutional neural networks for direct feature extraction and classification from raw EEG signals.

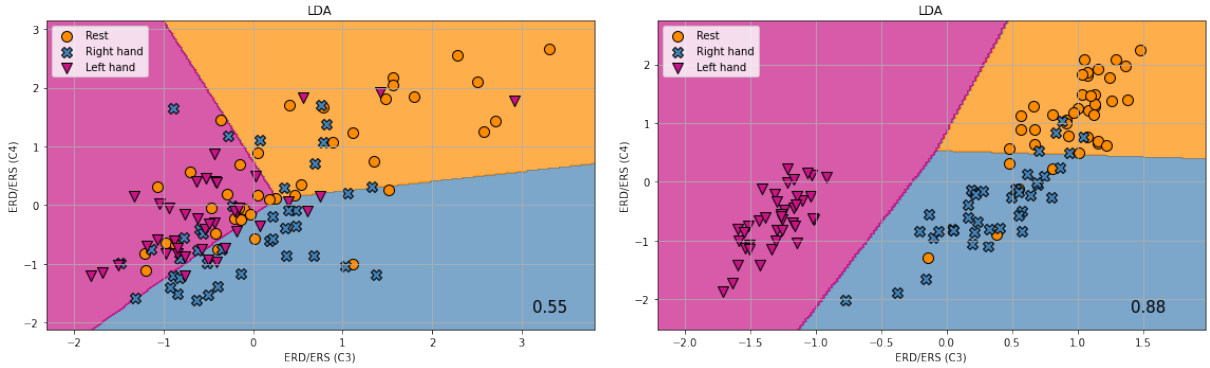


(a) Data from subject 1



(b) Data from subject 2

Figure 2.15: Decision boundary of different classifiers for 3 classes: right hand MI, left hand MI and rest. Classification decision based on ERD/ERS features from two channels, C3 and C4, acquired above motor cortex. In these examples, we used EEG data from two different subjects: (a) subject 1; (b) subject 2. Here rest class marked in orange, right hand MI marked in blue and left hand MI marked in purple.



(a) Data from subject 1

(b) Data from subject 2

Figure 2.16: Decision boundary of LDA classifier for 3 classes: right hand MI, left hand MI and rest. Classification decision based on ERD/ERS features from two channels C3 and C4, acquired above motor cortex. In these examples, we used EEG data from two different subjects: (a) subject 1; (b) subject 2.

2.5.1 Linear discriminant analysis

Linear discriminant analysis (LDA) is a generalization of Fisher's linear discriminant, a method used in statistics, pattern recognition, and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

LDA is a classifier with a linear decision boundary, generated by fitting class conditional densities to the data and using Bayes' rule. The model fits a Gaussian density to each class, assuming that all classes share the same covariance matrix.

In this thesis, we also used shrinkage (or regularized) linear discriminant analysis (sLDA) [142] which is a version of LDA with a regularisation parameter. This strategy commonly used when it is deals with small sample data set.

For illustrative purpose, we consider EEG signals for subjects 1 and 2 from the dataset (section 4.2). Decision boundary of LDA classifier is represented on the Figure 2.16 for 3 classes: right hand MI, left hand MI, and rest. Classification decision based on 2 ERD/ERS features from channels C3 and C4 computed as described earlier in section 2.2.2.1

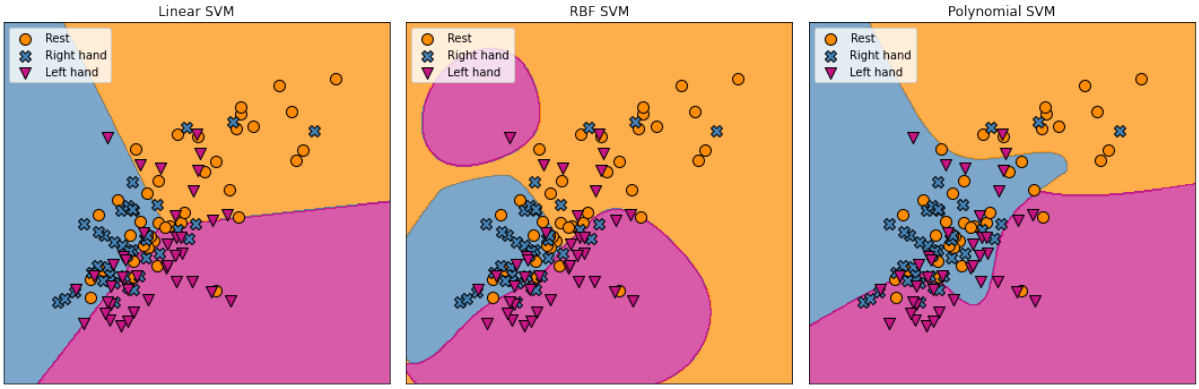
2.5.2 Support vector machine

Support vector machine (SVM) is a supervised learning model with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

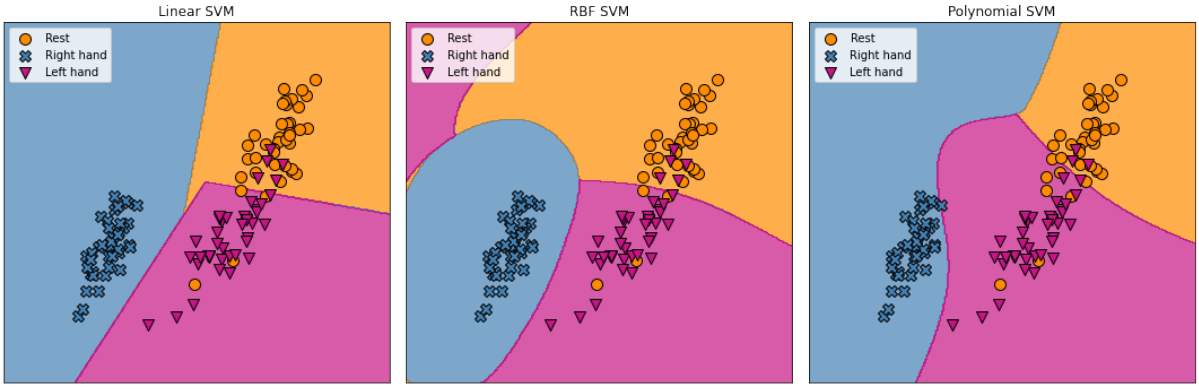
In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

The kernel function can be any of the following:

- Linear;
- Polynomial;
- Radial Basis Function (RBF);
- Sigmoid.



(a) Data from subject 1



(b) Data from subject 2

Figure 2.17: SVM classifiers with different kernels (linear, RBF and polynomial (3rd degree) for 3 classes: right hand MI, left hand MI and rest. Classification decision based on ERD/ERS features from two channels C3 and C4, acquired above motor cortex. In these examples, we used EEG data from two different subjects: (a) subject 1; (b) subject 2.

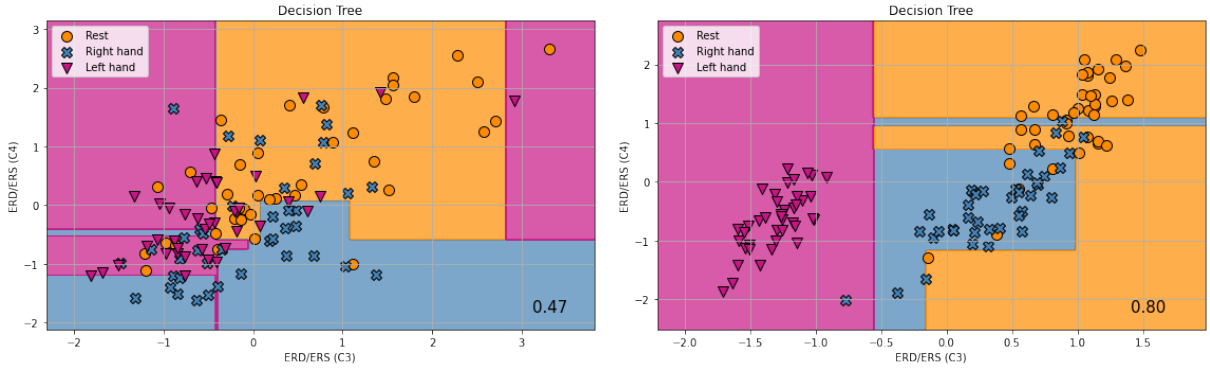
One of the main parameters of SVMs is a regularization parameter C. This parameter controls slack variables. In this thesis we choose C equals 0.01.

In the thesis, we used linear, polynomial, and RBF kernels. SVM classifiers with different kernels (linear, RBF and polynomial (3rd degree)) are represented on Figure 2.17. For illustrative purpose, we consider EEG signals for subjects 1 and 2 from the dataset (section 4.2). Classification decision based on ERD/ERS features obtained from two channels C3 and C4, acquired above motor cortex as described in section 2.2.2.1.

2.5.3 Decision tree classifier

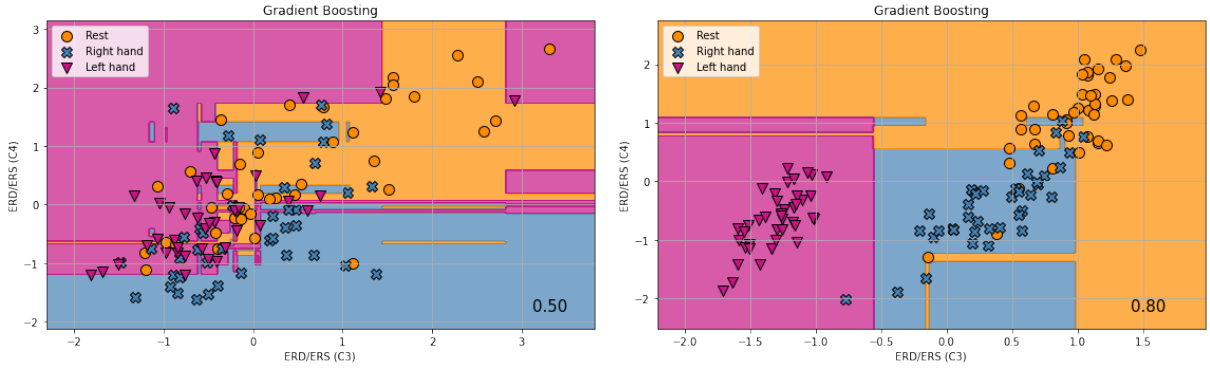
Decision tree learning [143] uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It is one of the predictive modeling approaches used in statistics and machine learning. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.

For illustrative purpose, we consider EEG signals for subjects 1 and 2 from the dataset (section 4.2). Example of decision boundary for decision tree classifier for 3 class problem (right hand MI, left hand MI, and rest) is represented on Figure 2.18. Classification decision based on ERD/ERS features obtained from two channels C3 and C4, acquired above motor cortex as described in section 2.2.2.1.



(a) Data from subject 1 **(b)** Data from subject 2

Figure 2.18: Results of decision tree classifier for 3 classes: right hand MI, left hand MI and rest. Classification decision based on ERD/ERS features obtained from two channels C3 and C4, acquired above motor cortex as described in section 2.2.2.1 In these examples, we used EEG data from two different subjects: (a) subject 1; (b) subject 2.



(a) Data from subject 1 **(b)** Data from subject 2

Figure 2.19: Results of gradient boosting classifier for 3 classes: right hand MI, left hand MI and rest. Classification decision based on ERD/ERS features obtained from two channels C3 and C4, acquired above motor cortex as described in section 2.2.2.1 In these examples, we used EEG data from two different subjects: (a) subject 1; (b) subject 2.

2.5.4 Gradient Boosting

Gradient Boosting (GB) is a supervised ML method for regression and classification problems that creates a prediction model as an ensemble of "weak learners". Weak learners are simple, typically linear prediction models. For gradient boosting usually used decision tree prediction models. GB algorithm builds the model in stages like other boosting methods and generalizes them to optimize an arbitrary differentiable loss function [144, 145].

For illustrative purpose, we consider EEG signals for subjects 1 and 2 from the dataset (section 4.2). Example of decision boundary for gradient boosting classifier represented on Figure 2.19 for 3 classes: right hand MI, left hand MI and rest. Classification decision based on ERD/ERS features obtained from two channels C3 and C4, acquired above motor cortex as described in section 2.2.2.1.

2.5.5 The Minimum Distance to Riemannian Mean Algorithm

As was mention in section 2.2.4 covariance matrices of EEG signals belongs to the Riemannian manifold. So one of the approaches is to use distance-based classification directly in the Riemannian manifold without projecting it to tangent space and turning to the Euclidean space [21, 23] (as was described in the section 2.2.4.1).

This method represented on the Figure 2.20.

At first it is necessary to compute the Riemannian Center of Mass of each class [146]:

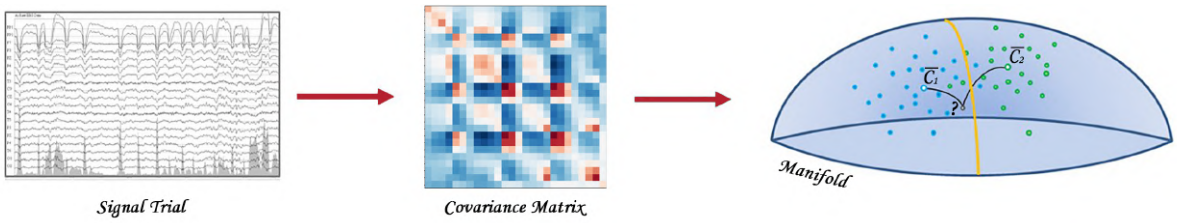


Figure 2.20: Classification on the Riemannian manifold using distance based approach [134].

Table 2.1: Codes for different states of binary classifier decoder

Code	State
0	Rest
1	Left hand
2	Right hand
3	Both hands
4	Feet
5	Left hand and feet
6	Right hand and feet
7	Both hands and feet

$$\overline{C^k} = \arg\min_{C}(\frac{1}{I}\sum_{i=1}^{I} d_R^2(C, C_i^k)) \tag{2.31}$$

Than for the new observations distances to each centers of mass are computed. The minimum distance defines the classification result:

$$class = \arg\min_{k} d_R(\overline{C^k}, C_{new}) \tag{2.32}$$

2.5.6 Proposed solution: a parallel connection of binary classifiers

As usual binary classifier can distinguish only two classes, it is not possible to use them for multi-class tasks. But in the task with combined movements, we proposed a solution of 3 binary classifiers that were connected in a parallel way.

Initially, we had a dataset with 8 classes (7 imaginary movements and rest). There are left hand, right hand, feet, and combinations of these movements: left hand and right hand, left hand and feet, right hand and feet, both hands and feet.

So each of 3 binary classifiers predict presence of only one motor imagery (right hand / left hand / feet) in the signal. In the end of the classification we have combination of 3 states (True or False). Then to decode this we used following equation 2.33:

$$PredictedLabel = 2^0 \times pred^1 + 2^1 \times pred^2 + 2^2 \times pred^3 \ , \tag{2.33}$$

where $pred^1$ - predicted label ('0' or '1') of the first classifier, which corresponds to the left hand activity; $pred^2$ - prediction of the right hand activity; $pred^3$ - prediction of the feet activity.

The code for each of the combined movements is presented in Table 2.1.

2.6 Cross-validation

Cross-validation (CV) is a model validation technique for estimating how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is to predict, and it is necessary to

assess how accurately the predictive model will work in practice. In the prediction task, models typically provide a set of known data on which training is done (a training data set) and a set of unknown data on which the model is tested (called validation data set or test data set).

The purpose of cross-validation is to test the model's ability to predict new data that was not used in its definition to find problems such as overfitting and selection bias and to give an idea of how this model will be generalized to unknown data for a real problem.

One round of cross-validation involves partitioning a sample of data into complementary subsets, performing an analysis on one of the subsets (called a training set), and validating the results on another subset (called validation or test set). To reduce variability, most methods perform several rounds of cross-validation using different partitions. Then results are averaged over the CV rounds to assess the predicted performance of the model.

Testing the quality of machine learning models should be carried out on examples that did not participate in its learning. The number of test samples should be the greater, the higher the quality of training is. If model errors have a probability close to one billion, then a billion test examples are needed to confirm this probability. It turns out that testing well-trained models becomes a very difficult task.

Cross-validation strategies that were used in this work will be reported later in chapter 4.

2.7 Statistical analysis

2.7.1 Wilcoxon signed-rank test

We performed a Wilcoxon signed-rank test (two-sided) to find the significant difference in accuracy obtained by different models. We report different significance levels ($p-value < 0.05, p-value < 0.01, p-value < 0.001$) in the plots.

The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test used to compare the differences between two samples of paired or independent measurements in terms of the level of any quantitative characteristic, measured on a continuous or ordinal scale [147].

The method consists of the comparison of absolute values of expression shifts in one direction or another. To do this, first, all the absolute values of the shifts are ranked, and then the ranks are summed up. If shifts in one direction or another occur randomly, then the sums of their ranks will be approximately equal. If the shifting intensity in one direction is greater, then the sum of the ranks of the absolute values of shifts in the opposite direction will be significantly lower than it could be with random changes.

One of the main advantages of this test is that the Wilcoxon test can be confidently used with a sample size of up to 25 elements [1] [2]. This is because, with a larger number of observations, the distribution of the values of this criterion is rapidly approaching normal. Therefore, in the case of large samples, it is advisable to transform the Wilcoxon test into a z-score [148].

2.8 Topographies

To analyze the differences between conditions, we performed an event-related spectral perturbation (ERSP) analysis for different frequency bands using EEGLAB.

Brain topography allowed us to display the possible changes over different electrodes on the scalp to localize which part of the brain was involved when the subject performed the requested task. In particular, it allowed us to understand how different conditions can be discriminated and which time parameters we can choose to guide the classification.

A surrogate permutation test (4,000 permutations) from the EEGLAB toolbox was used to validate differences in terms of localization of these ERSPs.

In addition to this analysis, we applied a false discovery rate (FDR) correction test to clarify how the FDR was controlled for multiple comparisons. This test consists of repetitively shuffling values between conditions and recomputing the measure of interest using the shuffled data. It was performed by drawing data samples without replacement and is considered suitable to show the difference between different conditions [149].

[1] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wilcoxon.html
[2] https://data.library.virginia.edu/the-wilcoxon-rank-sum-test/

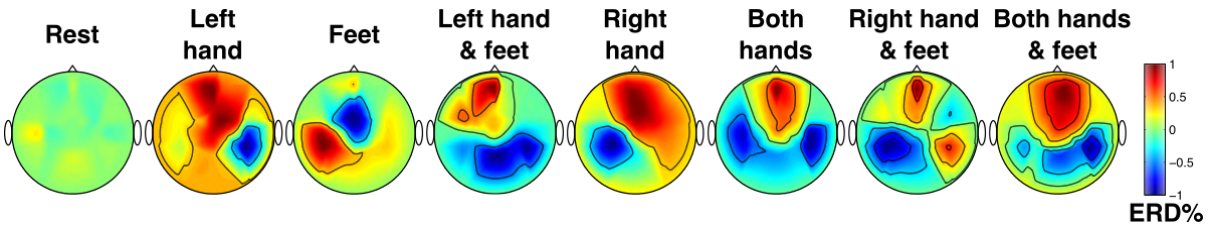


Figure 2.21: Distribution of the relative oscillatory power along the sensorimotor cortex during the course of the different motor tasks for subjet 2. ERD/ERS% values were obtained for all electrodes within the mu frequency range and averaged over the 4-second period starting one second after the task cue was presented. Negative values indicate ERD% modulation, which is characteristic of a motor task execution that appears over the contralateral hemisphere of the body part used in the process (broadly, electrode C4 for the left hand, C_z for feet, and C3 for right hand). Note that each topographic map is independently normalized to enhance the corresponding patterns [150].

Figure 2.21 shows a topographic representation for subject 2 from the dataset described in section 4.2 along with all conditions of the ERD/ERS% values obtained for all electrodes and averaged over the 4-second period starting one second after the task cue was presented. It can be observed that the lowest values appear on the regions associated with the limbs that are engaged at each motor task (broadly, right hand over electrode C3, feet over electrode C_z, and left hand over electrode C4).

2.9 Conclusion

MI-based BCIs is a domain that requires improvements and a lot of researchers working on new algorithms and methods for feature extraction and classification. In this chapter, we have presented algorithms frequently used in preprocessing, feature extraction, and classification modules of BCI. We have introduced time-frequency analysis, band power features, spatial analysis, and Riemannian geometry-based methods for feature extraction in MI-based BCI. Nevertheless common spatial pattern (CSP) algorithm is still the most popular method for feature extraction from EEG signal. The CSP algorithm finds spatial filters to maximize the variance of the filtered signal for one class and minimize it for another class at the same time. It has numerous advantages: it leads to high classification performance, it is flexible, computationally efficient, and simple to implement. Though despite all advantages CSP algorithm has limitations: it identifies only the important spatial information, but not the spectral one; frequency bands should be selected manually. To solve this issue, a filter bank common spatial pattern (FBCSP) algorithm has been proposed . The FBCSP algorithm uses several small different frequency bands (usually 4 Hz length) in parallel, which led to better classification performance. Other common approaches for the extraction of useful information from EEG data are algorithms based on Riemannian geometry (such as tangent space (TS) algorithm and minimum distance to Riemannian mean (MDRM) algorithm . This approaches usually outperforms the standard FBCSP+LDA method on classification sensory-motor rhythm (SMR) data.

We have introduced classical machine learning methods such as linear discriminant analysis (LDA), support vector machines (SVMs), gradient boosting, and decision tree algorithms that have been commonly used in MI-based BCIs for the classification task. Also, we have proposed a method based on three parallel-connected binary classifiers for the detection of seven MI combined movements and rest. Results of this proposed solution will be represented in section 5.2.2.

In this chapter, we also presented feature selection and outlier detection techniques that could be used in MI-based BCIs. We proposed a method based on outlier detection to improve classification performance and estimate the quality of imaginary movements before training the BCI.

In further chapter 3 we will present deep learning methods, such as convolutional neural networks for the direct feature extraction from raw EEG signals with the following classification.

Then, in chapter 4, we will present datasets with motor imagery data that were used in our investigations.

Classification performance results of algorithms represented in this chapter together with their comparison will be introduced in chapter 5.

Chapter 3

Deep neural networks for motor imagery detection

Contents

Artificial neural networks (ANNs) are machine learning methods, inspired by neural networks in the biological brain, based on an oriented graph where nodes are artificial neurons and oriented arcs are artificial connections [151]. Each connection between artificial neurons can transmit a signal (or value) from one neuron to another. The biological synaptic modulation is reproduced by a connection weight. The artificial neuron that receives the values from some other neurons via the weighted connection can process them and then transfer an output value to the following artificial neurons via a transfer function (or activation function) [152]. Weights and functions that calculate arousal can be learned from samples [153]. The output of each artificial neuron is calculated by a nonlinear function of the sum of its inputs. Artificial neurons and connections usually have a weight that adjusts by training. Weight increases or decreases the signal strength of the connection. Artificial neurons are usually arranged

in layers. Different layers can perform different types of transformations of their inputs. The signals pass from the first (input) to the last (output) layer, possibly after passing through the layers several times [151, 154, 155].

ANNs have been used in a variety of domains including computer vision, speech recognition, machine translation, natural language processing, and medical diagnosis [156].

A deep neural network (DNN) is an artificial neural network that organizes neurons in multiple layers between the input and output layers [157]. Additional layers enable the composition of features from first layers, potentially modeling complex data with fewer units than a similar ANN.

ANNs can be used for classification, segmentation and other tasks presented in section 2.5. But due to the complex nature of the models, numerous variants, and the fact that ANN can be used as a universal approximator of functions performing both processing and classification steps on the raw data forced us to devote a separate chapter to deep neural networks.

There are many types of artificial neural networks such as Feedforward Neural Networks, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) cells, Deep Belief Networks (DBNs). All of these artificial neural networks are different in their architecture and intended tasks (e.g. classification, object detection or segmentation).

ANN architecture design is determined depending on the problem being solved. But a lot of ANN architectures could not be used (or their application is not efficient) to properly solve BCI problems due to their complex structure and a small amount of data usually available for training in this domain. In the following sections, only the types of ANN which were used in our study will be reviewed.

3.1 Basics of artificial neural networks

There are many neural networks but they always consist of the same components: neurons, connections, weights, biases, and functions.

A neuron j at time t, receiving the input $p_j(t)$ from preceding neurons, consists of the following components [153]:

- activation $a_j(t)$, which depends on the discrete time parameter. Activation is a weighted sum of all the inputs, weighted by the weights of the connections from the inputs to the neuron;
- optional bias b_j, which learnt by the learning function;
- activation function f (usually nonlinear), which calculates the new activation at a given time $t+1$ with $a_j(t)$, b_j and network input $p_j(t)$, resulting in the relation:

$$a_j(t+1) = f(a_j(t), p_j(t), b_j); \tag{3.1}$$

- and output function f_{out}, which calculates the activation output for the last layer

$$y_j(t) = f_{out}(a_j(t)). \tag{3.2}$$

The input neuron has no predecessor but serves as an input interface for the entire network. The initial inputs are external data, such as image and signal values. Similarly, an output neuron has no successor and therefore serves as an output interface for the entire network [153]. The outputs forecast the value of the dependent variable accomplishing the task, such as classification or object recognition in an image.

In artificial neural networks, the activation function determines the output signal, which is determined by the input signal or a set of input signals. However, only nonlinear activation functions allow such networks to solve non-trivial problems using a small number of nodes. In artificial neural networks, this function is also called transfer function [154].

In the case of using linear or binary functions, it is needed to use many neurons for computations. In addition, networks built using such activation functions have unstable convergence, since the excitation of priority inputs of neurons tends to an unlimited increase since this function is not normalizable [155, 154]. Problems mentioned above can be solved with the nonlinear activation functions (for example Sigmoid function).

There are many activation functions that are used for ANNs. Most common activation functions are: Logistic (Sigmoid) function, Hyperbolic tangent (tanh), Rectified Linear Unit (ReLU) [158], Leaky Rectified Linear Unit (Leaky ReLU) [159], Exponential Linear Unit (ELU) [160], Scaled Exponential Linear Unit (SELU) [161], Maxout unit [162] and Softmax function [163]. Some of these activation functions are shown in the Figure 3.1.

For the classification task, the last layer is usually the logistic function for the binary classification problem and the softmax function for multi-class classification. Whereas for the hidden layers a sigmoid function (logistic function or hyperbolic tangent) is traditionally being chosen. But currently used a lot of different activation functions, with ReLU and it's varieties being common. Some desirable mathematical properties of the activation function:

- Non-linearity

  If the activation function is non-linear, it can be proven that the two-level neural network will be a universal function approximator [164]. The identity activation function does not satisfy this property. If multiple layers use the identity activation function, the entire network is equivalent to a single-layer model.

- Continuous Differentiability

  This property is desirable to provide gradient descent optimization techniques. For example, ReLU is not continuously differentiable and has some problems with gradient descent optimization, but remains a valid option. The binary step activation function is not differentiable at point 0 and its derivative is 0 at all other points, so gradient descent methods do not give any success for it [165].

- Range of values

  If the set of values of the activation function is limited, the gradient-based learning methods are more stable, since the representations of the standards significantly affect only a limited set of link weights. If the range of values is infinite, training is generally more efficient, since reference representations significantly affect most weights. In the latter case, a lower learning rate is usually required.

- Monotonicity

  If the activation function is monotonic, the error surface associated with a sibling model is guaranteed to be convex [166].

- Smooth functions with monotonic derivative

  It is shown that in some cases they provide a higher degree of generalization [154].

- Approximates the identity function near the origin

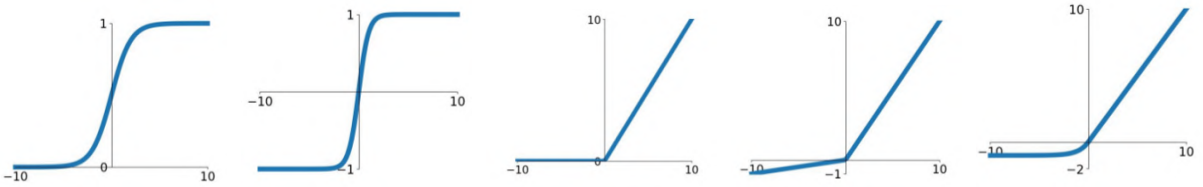
  If the activation functions have this property, the neural network will train efficiently if its weights are initialized with small random values. If the activation function does not approximate the identity around the origin, it is needed to be careful when initializing the weights [167]

In this thesis we used ReLU as an activation function for the hidden layers and softmax function for the output layer. Rectified linear unit (ReLU) is an activation function defined as the positive part of its argument [158]:

$$f(x) = \max(0, x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3.3}$$

ReLU has a range $[0, +\infty)$ and can provide better training for deep neural networks, compared to widely used activation functions, such as logistic sigmoid and its more practical analog, hyperbolic tangent [168]. It was demonstrated for the first time in 2011 [169]. And soon in 2017, ReLU was the most popular activation function [170].

The softmax function is a generalization of the logistic function to multiple dimensions. This activation function takes as input a vector x of J real numbers and normalizes it into a probability distribution consisting of J probabilities proportional to the exponentials of the input numbers. The softmax function has a range $(0, 1)$[163] and defined by the formula:



(a) Sigmoid **(b)** tanh **(c)** ReLU **(d)** Leaky ReLU **(e)** ELU

Figure 3.1: Activation functions

$$f_i(x) = \frac{e^{x_i}}{\sum_{j=1}^{J} e^{x_i}} \tag{3.4}$$

for $i = 1, ..., J$ For instance, j could be iterating through the number of kernels of the previous neural network layer while i iterates through the number of kernels of the current layer [163].

3.2 Training of artificial neural networks

In the learning process, the network views the training set in a certain order. The order of viewing can be sequential, random, etc. Some networks for unsupervised learning (for example, Hopfield networks) view the sample only once. Others (such as Kohonen networks), as well as networks for supervised learning, view the training set many times. One complete pass through the training set called the learning epoch [151]. In the case of supervised learning, the set of initial data is divided into two parts - the training data and test data; the principle of separation can be arbitrary. Training data is provided to the network for training, and test data is used to calculate the network error. Thus, if the error on the test set decreases, the network does perform a generalization. If the error on the training data continues to decrease, but the error on the test data increases, then the network has stopped generalizing and is simply "remembering" the training data. This phenomenon is called network overfitting. In such cases, training is usually terminated [154].

Other problems may occur during the training process, such as vanishing gradient problem or the network could fall into the local minimum [154, 155].

3.2.1 Optimisation algorithms for ANN

Optimization algorithms are techniques and methods used to modify the attributes of the neural network (for example weights and learning rates), to reduce losses, and to provide the most accurate results possible.

In this thesis different optimization strategies were studied. Three variants of gradient descent (such as standard Gradient Descent (GD) Stochastic Gradient Descent (SGD) [171, 172] and mini-batch GD) have been reviewed, among which mini-batch gradient descent is the most popular and has fewer disadvantages [173]. Recently more algorithms for optimization appeared, such as Adagrad [174], Adadelta [175], RMSprop (proposed by Geoff Hinton in Lecture 6e of the Coursera Class [3]), and Adam [176].

We choose Adam for the implementation in this thesis due to its speed and stability [176, 173].

Adam [176] (Adaptive Moment Estimation) works with momentums of the first and second order. The intuition behind Adam is that it doesn't run so fast just to jump over the minimum, it decreases a velocity gradually to search carefully. In addition to keeping the exponentially decaying mean of the previously squared gradients like AdaDelta, Adam also stores the exponentially decaying mean of the previous gradients $\hat{m}_t$.

Values $\hat{m}_t$ and $\hat{v}_t$ are the first and the second moment, which is the mean and uncentered variance of the gradients, respectively [173].

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{3.5}$$

[3] http://www.cs.toronto.edu/ tijmen/csc321/slides/lecture_slides_lec6.pdf

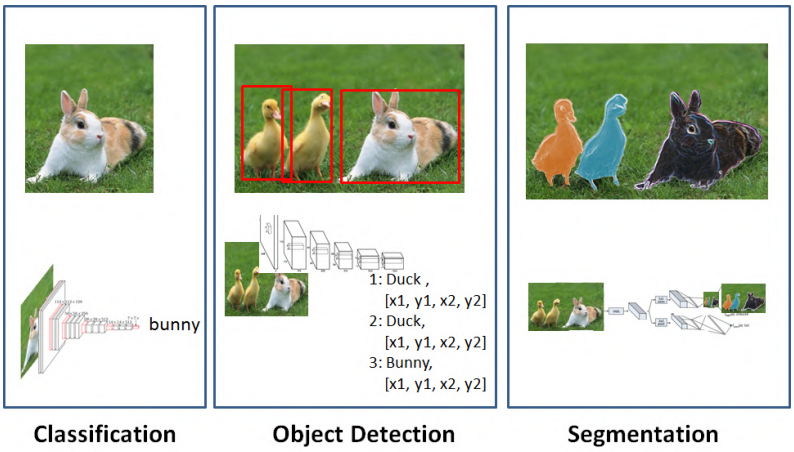


Figure 3.2: CNN applications (from [181]).

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{3.6}$$

Here the average of $\hat{m}_t$ and $\hat{v}_t$ has taken, so $E[m(t)]$ can be equal to $E[g(t)]$, where $E[f(x)]$ is the expected value of $f(x)$.

Then $\hat{m}_t$ and $\hat{v}_t$ used to update the parameters, as in Adadelta and RMSprop, which gives Adam update rule:

$$\Theta_{t+1} = \Theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t \tag{3.7}$$

In this work we used default values of β_1 equal to 0.9, and β_2 equal to 0.999, and 10^{-8} for ϵ (as proposed by authors [176]).

Adam has several advantages, this method is fast and converges quickly [176] and it fixes vanishing learning rate and high variance [173]. In the works [176, 177, 173] it is shown that Adam makes progress faster in terms of epochs number comparing to other algorithms. But at the same time it is computationally expensive.

3.3 Convolutional Neural Networks (CNN)

Convolutional neural network (CNN) is a special architecture of artificial neural networks proposed by Yann Lecun in 1988 [178] and aimed at efficient pattern recognition.

The idea behind convolutional neural networks is an alternation of convolution layers and pooling layers. The structure of the network is unidirectional, fundamentally multilayer. For training, standard methods are used, most often the backpropagation method. Any activation function could be chosen, but commonly the activation function is a ReLU layer [179].

The network architecture got its name due to the presence of the convolution operation, the point of which is that each fragment of the image is multiplied by the convolution matrix (kernel) element by element, and the result is summed up and written into a similar position in the output image [180].

The operation of a convolutional neural network usually performs a transition from specific features of an image to more abstract details, and then to even more abstract details up to the isolation of high-level features. At the same time, the network self-adjusts and develops the very necessary hierarchy of abstract features (feature maps), filtering unimportant details and highlighting important ones [180].

In fact, the features generated by a complex CNN are usually difficult to interpret, so in practice, it is not particularly recommended to try to understand the contents of these features, instead, it is recommended to improve the structure and architecture of the network itself to obtain better results [179].

CNNs are good in classification, object detection and segmentation [179] of the images and signals (Figure 3.2).

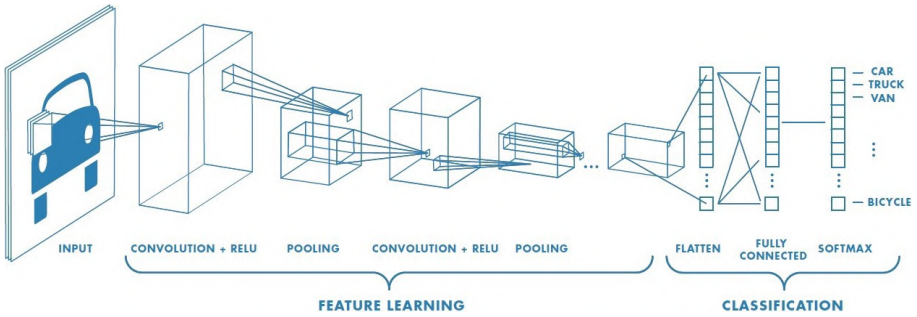


Figure 3.3: CNN architecture (from [182]).

3.3.1 General architecture of CNN

In CNN, the convolution operation uses only a limited matrix of small weights, which is moved over the entire processed layer (at the very beginning directly over the input image or signal), forming after each shift an activation signal for the neuron of the next layer with a similar position.

The same weight matrix (convolution kernel) is used for different neurons of the layer. It is interpreted as a graphic encoding of some feature, for example, the presence of an oblique line at a certain angle. Then the next layer, resulting from the convolution operation with such kernel, shows the presence of this feature in the processed layer and its coordinates, forming a feature map. Naturally, in CNN, there is a set of different kernels that encodes image elements. These convolution kernels are not set manually in advance but are formed independently during the training of the network [180].

Each kernel generates its own instance of a feature map, making the neural network multichannel (many independent feature maps on one layer). It should be noted that when a kernel is iterated over a layer, it usually moves not for a full step (the size of this matrix), but a short distance (one or two neurons (pixels)).

Subsampling or pooling is the operation of combining, performs a dimension reduction of the generated feature maps. In CNN architecture information about the presence of the desired feature is more important than accurate knowledge of its coordinates, therefore, the maximum (or average) is selected from several neighboring neurons of the feature map and is saved as one neuron of the condensed feature map of lower dimension. Due to this operation, in addition to speeding up further calculations, the network becomes more invariant to the scale of the input image [179].

Let's consider a typical structure of CNN in more details (see Figure 3.3). The network consists of a large number of layers. After the input layer (input image), the signal passes through a series of convolutional layers, in which the actual convolution and pooling alternate. Usually, after passing through several layers, the feature map degenerates into a feature vector. At the output of the convolutional layers, several fully connected layers are additionally installed, at the input of which flatten feature maps are fed.

3.3.2 Convolution layer

The convolution layer is the main block of CNN. The convolution layer includes its own kernel (filter) for each channel, the convolution kernel processes the previous layer fragment by fragment (summing the results of the element-wise product for each fragment). The weights of the convolution kernel are unknown and are set during training.

A feature of the convolutional layer is the relatively small number of parameters set during training. So for example, if the original image has a dimension of $12 \times 12 \times 3$ pixels (this means 432 input neurons), and the convolutional layer uses filters with a 5x5 pixel kernel with an output of 1 channel, then only 25 kernel weights are determined during the training process. However, because we have 3 channels image our kernel needs to have also 3 channels. That's why we actually have $5 \times 5 \times 3 = 75$ learnable parameters, which is significantly less than the number of required parameters of a fully connected neural network [183].

Convolutional layer with regularized weights over fewer parameters allows avoiding the vanishing gradient and exploding gradient problems that can be observed during backpropagation in FNN [184].

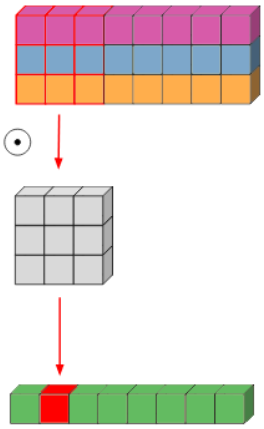


Figure 3.4: Convolutional layer with 1 channel output (adapted from [185]).

The result of each convolution goes to the activation function, which is a non-linear function. The activation layer is usually logically combined with the convolution layer.

Figure 3.4 shows the example of the convolutional layer with one channel output. Consider a signal with 8 samples and 3 channels as an input. In this case, there exists only one kernel $3 \times 1 \times 3$. And resulting feature map has a size $8 \times 1 \times 1$ in the case with the same padding and a stride of one.

3.3.3 Same and valid padding

Padding is an addition of pixels to the edge of the image. Adding padding allows creating very deep models so that feature maps don't fizzle out[179].

The same padding calculates and adds the padding required to the input image (or feature map) to control that the output has the same shape as the input.

The same padding tries to add columns evenly to the left and right, but if the number of columns to add is odd, it will add an extra column to the right (same logic applies vertically: there might be an extra row of zeros at the bottom) [179].

With the same padding, if a stride of one is used, the layer's output will have the same spatial dimensions as its inputs.

With valid padding, there are no augmented input fields and the layer uses only valid input data. That's why some number of columns or rows could be dropped in this case. Valid padding discards only the right-most columns (or the bottom-most rows) [179].

3.3.4 Depthwise convolutional layer

Depthwise convolution is a type of convolutional layer. With regular convolution performed on multiple input channels (Figure 3.4), the filter has a depth equal to the input and allows to mix channels freely to generate each element in the output. Depthwise convolution performs other way - each channel is kept separately [186, 187]. The Figure 3.5 showed how it works [185, 183].

Depthwise convolution has three steps:

- Divide the input data into channels and divide the kernel into channels (the input and the kernel have the same number of channels).
- For each of the channels, convolve the input with the appropriate kernel, creating an output feature maps.
- Stack the output kernels together.

Note that the channel multiplier depth parameter of the depthwise convolutional layer lets us compute multiple outputs for each input channel [188, 185]. Example in Figure 3.5 has multiplier depth equals to one. But it can be

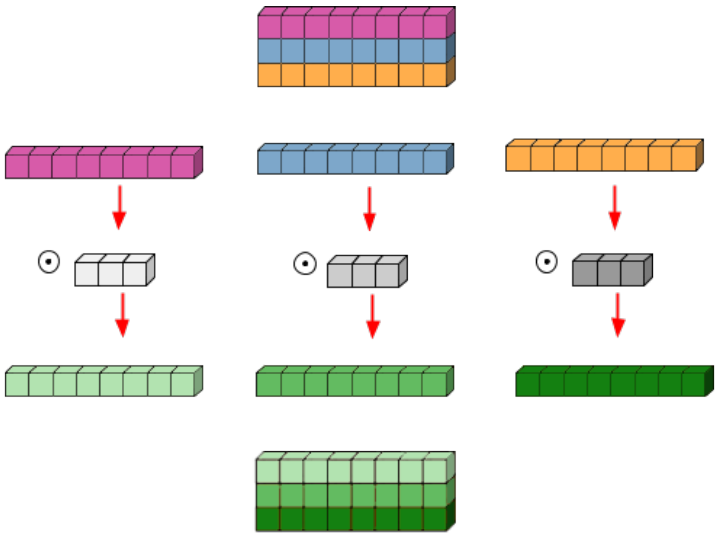


Figure 3.5: Depthwise convolutional layer with 1 channel output (adapted from [185]).

$n \geq 2$. In this case, n different kernels for each input channel will be created, allowing to compute more variations of the feature maps. The output channel number will be n times greater than the input channel number.

3.3.5 Depthwise separable convolutional layer

Depthwise separable convolutional [186, 187] is a composition of two convolutions: depthwise convolution and following pointwise convolution. In Figure 3.6 it could be seen that after the depthwise convolution is completed, an additional step is performed: 1×1 spatial convolution across channels. Pointwise convolution kernel iterates through every single point. This kernel has a depth of however many channels the input image has; in our example, 3 (Figure 3.6) [183].

Similarly to depthwise convolution depthwise separable convolution has the parameter "channel multiplier depth" which allows computing several outputs channels for one input channel [185, 186].

Depthwise and separable convolutions are efficient alternatives to standard convolution layers, which reduces the number of required weights and computations [186, 187, 183, 185] and thus are less prone to overfitting.

3.3.6 Pooling layers

A pooling layer is nonlinear downsampling of a feature map, while a group of samples is compressed to one sample. The maximum function [189] is most commonly used in this case. The transformations involve non-intersecting rectangles or squares, each of which is compressed into one sample, and the sample with the maximum value is selected. The pooling operation can significantly reduce the feature map size. Pooling filters unnecessary data and prevents overfitting. The pooling layer is usually inserted after the convolution layer, and before the next convolution, layer [179].

In addition to pooling with a maximum function, average pooling could be used. In this case, the average of the samples in the square is used for the output [179]. Average and max-pooling showed in Figure 3.7.

3.3.7 Fully-connected layer

After several passes of convolution and pooling layers, the system rebuilds from a concrete representation of the signal to more abstract feature maps. Usually, after each convolutional layer, the number of channels increases, and the signal dimension decreases in each channel. In the end, there remains a large set of channels storing a small amount of data (even one parameter), which are interpreted as the most abstract features identified from the original data [179].

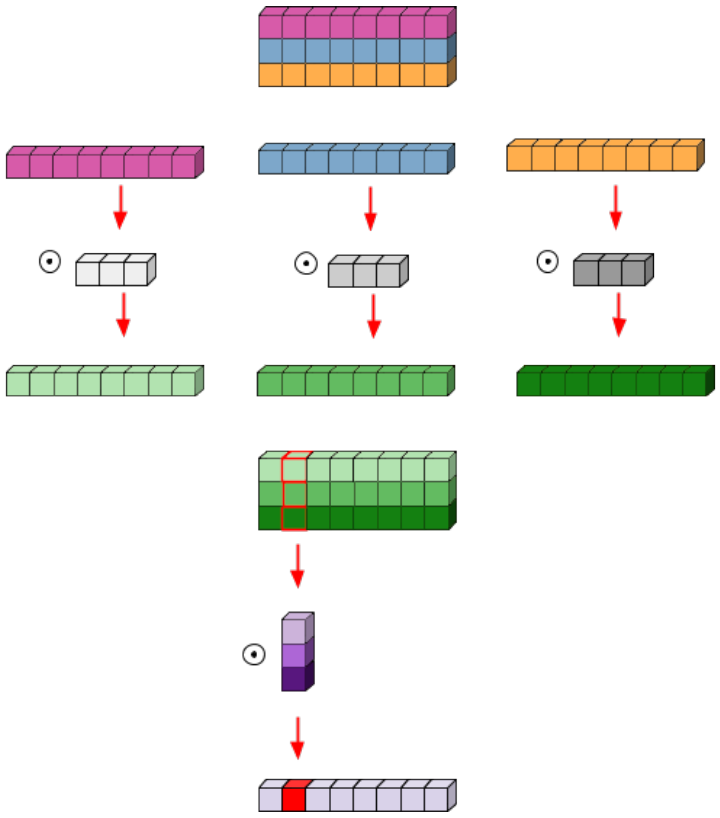


Figure 3.6: Depthwise separable convolutional layer with 1 channel output (adapted from [185]).

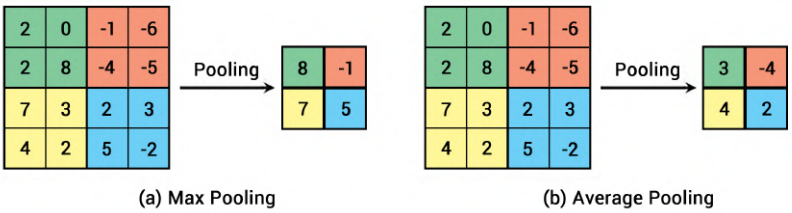


Figure 3.7: Examples of max and average pooling.

This data is combined and transmitted to a standard fully connected neural network, which can also consist of several layers. In this case, fully connected layers have a relatively small dimension (relative to the number of samples in the original image).

Fully connected layers could have different activation functions. ReLU activation function is commonly used. But the last layer of the ANN uses softmax activation function to predict several classes [184].

3.4 Overfitting and regularization techniques

Overfitting is one of the main problems of ANN training. It consists of the following: the model explains well only examples from the training set, adapting to training examples, instead of learning to classify examples that were not involved in training and fail to capture statistical processes which described the data, thereafter losing the model's ability to generalize.

In general, this problem requires separate treatment. This is especially important for cases where only a very limited number of training samples are available as in the case of the BCI domain. One of the most simple methods to learn on limited data is called early stopping [190]. Usage of this method often ensures that the

network generalizes well examples outside of the training set. This is achieved by finding the moment when the model stops learning and starts overfitting, usually checking the model on a validation dataset.

Regularization is another method used to avoid overfitting. There are regularization techniques that could be applied directly to learnable parameters inside layers of ANN and also particular regularization layers, such as dropout [191] or batch normalization layers [192]. These techniques could be used separately or collectively because they regularize different parts of ANN. So in this thesis to prevent overfitting, we decided to use L1 and L2 regularization for CNN kernels; dropout before fully connected layer and between convolutional layers; and batch normalization layer after convolution layer and before the activation layer. We will introduce these techniques further in current section.

3.4.1 L1 and L2 regularization

The first regularizer allows for applying penalties to layer parameters or layer activity during optimization. These penalties are summed up in a loss function that the network optimizes.

Regularizers can be of three types:

- Regularizer to apply a penalty on the layer's kernel
- Regularizer to apply a penalty on the layer's bias
- Regularizer to apply a penalty on the layer's output

In this case regularization terms are "constraints" that the optimization algorithm must "adhere to" when minimizing the loss function, in addition to the need to minimize the error between true y and predicted $\hat{y}$. Loss function without regularization looks as follows:

$$Loss = Error(y, \hat{y}) \tag{3.8}$$

Regularizers could apply different penalties:

- A regularizer that applies a L1 regularization penalty:

$$Loss = Error(y, \hat{y}) + \lambda_1 \sum_{i=1}^{N} |w_i| \tag{3.9}$$

- A regularizer that applies a L2 regularization penalty:

$$Loss = Error(y, \hat{y}) + \lambda_2 \sum_{i=1}^{N} w_i^2 \tag{3.10}$$

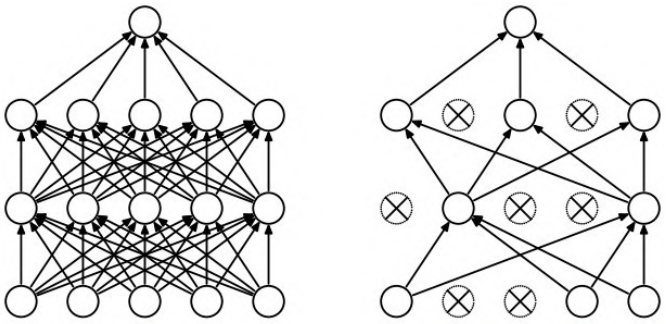
- Regularizer that applies both L1 and L2 penalties:

$$Loss = Error(y, \hat{y}) + \lambda_1 \sum_{i=1}^{N} |w_i| + \lambda_2 \sum_{i=1}^{N} w_i^2 \tag{3.11}$$

3.4.2 Dropout layer

Dropout is a method for regularization of ANN, designed to reduce overfitting of the network by preventing complex coadaptation of individual neurons to training data during learning [191].

Dropout characterizes the exclusion of a certain percentage (for example, 30% or 50%) of random neurons at different epochs during the training of a neural network. This is a very efficient way of averaging models within a neural network. As a result, more trained neurons receive bigger weight in the network [193]. This technique significantly decreases the learning time, increase the quality of learning on training data, and also improves the quality of model predictions on new test data [191].



(a) Standard Neural Net (b) After applying dropout.

Figure 3.8: Graphical representation of the Dropout method (from [191]). On the left is the neural network before dropout was applied to it, on the right is the same network after dropout. The network shown on the left is used for testing, after parameters training.

The main idea behind dropout is to train an ensemble of several ANNs instead of training one ANN and then average the results.

Networks for training are obtained by dropping out neurons with probability p, so the probability that a neuron will remain in the network is $1 - p$. Dropping out a neuron means that for any input and with any parameters, it returns 0.

The excluded neurons do not contribute to the learning process at any stage of the backpropagation algorithm.

Consider $h(x) = xW + b$ to be a linear projection of the input d_i-dimensional vector x on the d_h-dimensional space of output values; and activation function $a(h)$. Applying Dropout at the training stage can be represented as a modified activation function:

$$f(h) = D \odot a(h) \tag{3.12}$$

where $D = (X_1, ..., X_{d_x})$ - d_x-dimensional vector of random variables X_i distributed according to Bernoulli's principle.

X_i has the following probability distribution:

$$f(k, p) = \begin{cases} p, & if k = 1 \\ 1 - p, & if k = 0 \end{cases}, \tag{3.13}$$

where k are all possible output values.

Obviously, this random variable fits perfectly if Dropout is applied to one neuron. Indeed, the neuron is turned off with probability $p = P(k = 1)$, otherwise, it is left on.

Let's look at applying Dropout to the i-th neuron:

$$O_i = X_i a(\sum_{k=1}^{d_i} w_k x_k + b) = \begin{cases} a(\sum\limits_{k=1}^{d_i} w_k x_k + b), & if X_i = 1 \\ 0, & if X_i = 0 \end{cases}, \tag{3.14}$$

where $P(X_i = 0) = p$.

Since at the training stage the neuron remains in the network (is not excluded) with probability $1 - p$, at the testing stage we need to emulate the behavior of the ensemble of neural networks used in training. For this, the authors propose at the testing stage to multiply the activation function by the coefficient $q = 1 - p$. Thereby,

At the training stage:

$$O_i = X_i a(\sum_{k=1}^{d_i} w_k x_k + b) \tag{3.15}$$

At the testing stage:

$$O_i = q a(\sum_{k=1}^{d_i} w_k x_k + b) \tag{3.16}$$

3.4.3 Batch normalization layer

Batch normalization is a method that improves performance and stabilizes ANN through normalization of the input layer by re-centering and re-scaling [192].

Normalization of the input layer of the neural network is usually done by scaling the data supplied to the activation functions. For example, when there are features with values from 0 to 1 and some features with values from 1 to 1000, then they need to be normalized to speed up learning. Data normalization can be performed in the hidden layers of neural networks, which is what the batch normalization method does.

Batch normalization reduces internal covariate shift [192]. Covariate shift is a situation when the distributions of feature values in the training and test sets have different parameters (mathematical expectation, variance, etc.).

Consider values x from batch $B = x_1, ..., x_m$; adjustable parameters γ, β; constant ϵ for computational stability.

The batch normalization algorithm can be represented as follows:

1. Determine batch expected value:

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{3.17}$$

2. Determine batch variance:

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2 \tag{3.18}$$

3. Compute normalization:

$$\hat{x_i} = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \tag{3.19}$$

4. Compute compression and shift:

$$BN_{\gamma,\beta}(x_i) = \gamma\hat{x_i} + \beta \tag{3.20}$$

Besides, using batch normalization has several additional useful properties:

- faster convergence of models is achieved [192], despite performing additional calculations;
- batch normalization allows each layer of the network to learn more independently of other layers;
- it becomes possible to use a higher learning rate since batch normalization ensures that the outputs of the neural network nodes will not have too large or small values [192];
- batch normalization is also a regularization mechanism: this method introduces some noise into the outputs of hidden layer nodes, similar to the dropout method;
- models become less sensitive to initial weight initialization [192].

3.5 Review of deep learning for motor imagery detection in BCI

With the growth of computing powers interest in artificial neural networks was renewed recently. Every year models become more complex and demonstrate better results what is proved by ImageNet Large Scale Visual Recognition Challenge [194]. In 2011 classification error rate was 25%. Deep CNN AlexNet [195] achieved winning top-5 test error rate of 15.3%, compared to runners-up's 26% in 2012. AlexNet was consisted of 8 layers and had 650 000 neurons. From that time AlexNet was outperformed numerous times and in 2017, 29 of 38

competing teams had greater than 95% accuracy. With increasing accuracy, the performance number of layers also increased, and in 2017 winning SENet [196] had 154 layers[4].

Naturally, the growth of deep learning popularity did not leave aside the BCI community. However, there is a problem with a very limited amount of labeled data in the BCI domain. A large amount of data and the development of computing power allowed and played a significant role in the success of deep learning techniques and ANNs in the areas of computer vision and natural language processing. A small amount of data limited the application of deep learning in the BCI domain.

Below is an overview of the application of deep learning in motor imagery based BCIs. This section presents the data from the 71 reviewed articles, Table 3.1 summarizes all the obtained data. We selected all articles (including conference papers) written in English which we were able to find as of November 29, 2020, using next key words: artificial neural network, deep learning, motor imagery, BCI, EEG. We used ieeexplore.ieee.org, scholar.google.com and google.com in our research.

In the reviewed studies different datasets were used, some of them are available, such as BCI Competition datasets [197, 198, 199, 200], EEG motor movement/imagery dataset (eegmmidb) dataset [201] from Physionet [202], upper limb movement (ULM) dataset [203], High-Gamma Dataset (HGD) [26], and some of the reviewed studies used private datasets.

Feature extraction methods, deep learning classifier architectures, and optimization strategies that have been used by the reviewed articles are presented in Table 3.1.

There are different deep learning architectures used in MI-based BCI domain. Most reviewed studies used CNN architectures (62.2%), while portion of following LSTM (8.5%), RNN (7.3%) and DBN (6.1%) is significantly lower. The use of the remaining architectures, such as CapsNet [204, 205], Stacked Autoencoder (SAE) [206, 207], Autoencoder (AE) [208, 209], Variational Autoencoder (VAE) [210], Gated Recurrent Unit (GRU) [211] and Extreme Learning Machine (ELM) [212] even lower (only 10.8% in total). Figure 3.9 represents proportion use of the DNN architectures across all the reviewed articles.

Some architectures contains stacked layers of different types, such as CNN + Stacked Autoencoders (SAE) [206], CNN + RNN [213] and RNN + Autoencoder (AE) + XGBoost [208, 209].

The reviewed studies used different activation functions in the proposed architectures of neural networks. The employed activation functions in the reviewed articles were as follows: 53% ReLU, 19.7% ELU, 9.1% Sigmoid, and 6.1% SELU. The proposed architectures in the remaining studies used the tanh and linear activation functions as well as S-shaped Rectified Linear Unit (SReLU) and Leaky ReLU. Figure 3.10 represents the proportion use of the different activation functions across all the reviewed articles where it was mentioned.

During the training procedure, the parameters of the neural networks updates using different optimization methods. The larger part of the reviewed studies used one of the followed optimization algorithms: Adam, gradient descent (GD), and stochastic gradient descent (SGD). And only a few studies used RMSProp and Adadelta. Figure 3.11 represents the proportion use of the different optimization algorithms across all the reviewed articles.

[4]https://paperswithcode.com/sota/image-classification-on-imagenet/

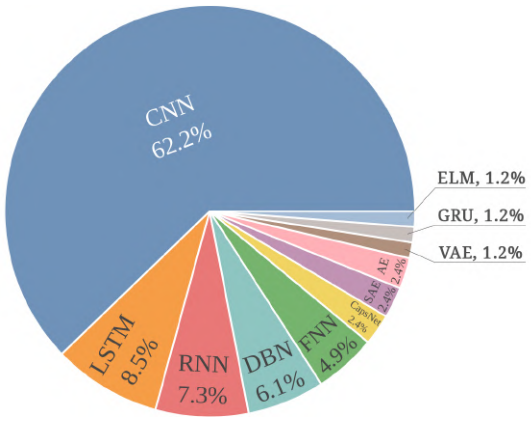


Figure 3.9: Proportion use of the DNN architectures across all the reviewed articles.

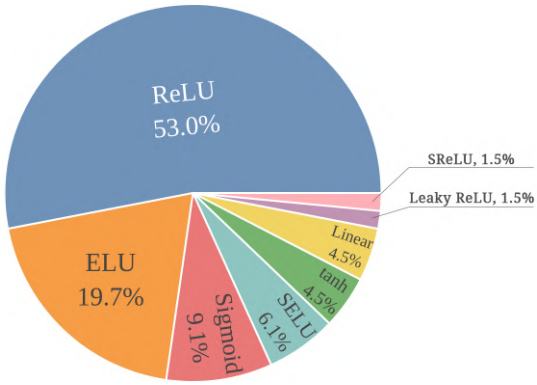


Figure 3.10: Proportion use of different activation functions across all the reviewed articles.

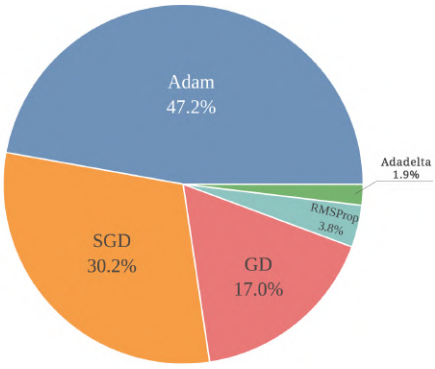


Figure 3.11: Proportion use various optimization algorithms across all the reviewed articles.

Application of deep learning in MI-based BCI can be divided according to the feature extraction part. There are architectures of DNN which work with raw EEG data and others that require certain preprocessing and feature extraction to prepare input data (whether it be images or features) for the subsequent classification. Deep learning architectures which work with raw data perform feature extraction step inside the network.

That's why it was decided to divide all the reviewed studies into three conditional categories according to input formulation types. The number of articles that have used each of the input types is as follows: features (30), time-series (22), and images (19). Figure 3.12 represents the proportion use of input formulation types across all the reviewed articles with the used feature extraction methods for achieving them.

About 27.5% of all the reviewed studies used raw EEG data as input for deep learning models (Figure 3.13a).

About 40% of all the reviewed articles used CSP or FBCSP in the process of features extraction for feature input calculation. While time-frequency analysis methods such as Fast Fourier transform (FFT), Power Spectral Density (PSD), Continuous Wavelet Transform (CWT), Wavelet Packet Decomposition (WPD), Wavelet Packet Transform (WPT), and Continuous Wavelet Decomposition (CWD) take about 24% in total. Figure 3.13b represents detailed proportion use of feature extraction methods for features input formulation across all the reviewed articles.

Short-Time Fourier Transform (STFT), FFT, and Fast Fourier Transform Energy Map (FFTEM) were used by 47.5% of the studies that adopted images in the classification process. While wavelet transforms takes 38%. Figure 3.13c represents detailed proportion use of feature extraction methods for images input formulation across all the reviewed articles.

Different tasks activate different electrical signals on the human scalp. The frequency range of the signals differs. The reviewed studies have suggested different frequency ranges for dealing with the MI tasks during the investigation and classification. However, from the reviewed articles and as shown in Table 3.1, the investigated starting point of the frequency range was within 4–8 Hz; while the observed ending point of the range was within 35–40 Hz for MI related tasks.

According to observed information, CSP and FBCSP are frequently used in case of extraction features as inputs

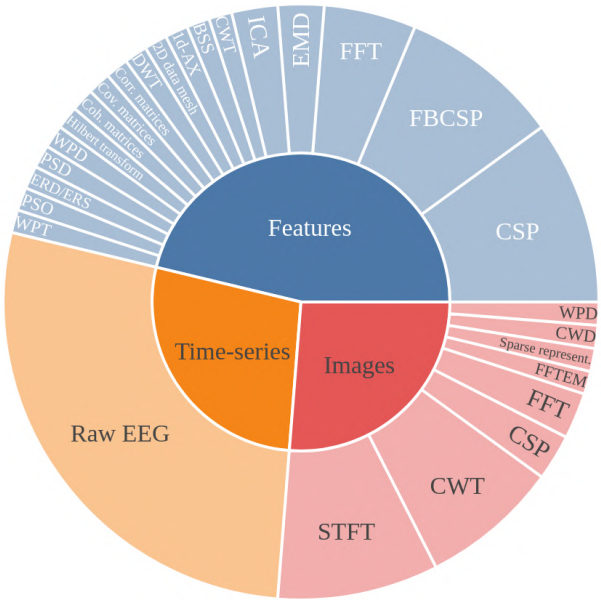


Figure 3.12: Proportion use of input formulation types across all the reviewed articles with the used feature extraction methods for achieving them.

to the neural network and thus could be recommended for this task. While wavelet and Fourier transform-based methods are recommended in case of using images as inputs to the DNN.

Fourier methods analyze signals using a window of fixed length, which is an obvious limitation of these methods due to the non-stationarity of EEG signals. Wavelet transforms could be better suited for capturing images from the raw EEG data than Fourier transform. The wavelet methods analyze the signals with numerous frequencies and time resolutions.

At the same time there are a lot of studies which used raw EEG signal as an input for the models. And the number of such articles grows every year, 12 of the 22 reviewed articles were published after beginning of 2019. Most of these articles used CNN architecture [214, 215, 213, 216, 27, 217, 218, 219, 220, 221, 222, 26, 223, 224, 225, 209, 226].

Application of other architectures such RNN [227, 228], LSTM [225, 228, 229], SAE [207], DBN [230] remains limited for the raw EEG analysis. Some of the studies used stacked variations of different networks, such as RNN + AE + XGBoost [209] or CNN + RNN [213].

Table 3.1: Summary of a deep learning application in MI-based BCIs

Ref	Dataset	Freq. range (Hz)	Feature extraction	Input	DL architecture	Activation function	Optimizer
[214]	BCIC IV 2a [200]	N/M	Raw EEG	time-series	CNN	ELU	Adam
[227]	BCIC IV 2a [200]	0.5-100	Raw EEG	time-series	RNN	ReLU	N/M
[215]	BCIC IV 2a [200]	8-30	Raw EEG	time-series	CNN	Leaky ReLU	N/M
[213]	BCIC IV 2a [200]	1-40	Raw EEG	time-series	CNN, CNN+RNN	ELU	RMSProp
[216]	BCIC IV 2a [200]	1-38	Raw EEG	time-series	CNN	ELU	Adam

Continued on next page

Table 3.1 – *Continued from previous page*

Ref	Dataset	Freq. range (Hz)	Feature extraction	Input	DL architecture	Activation function	Optimizer
[27]	BCIC IV 2a [200]	4-40	Raw EEG	time-series	CNN	Linear, ELU	Adam
[207]	BCIC IV 2a [200]	8-35	Raw EEG	time-series	SAE	N/M	N/M
[217]	BCIC IV 2a [200]	N/M	Raw EEG	time-series	CNN	ELU	SGD
[218]	BCIC IV 2a [200], BCIC IV 2b [200]	0.01-38	Raw EEG	time-series	CNN	Linear, ELU	Adam
[219]	BCIC IV 2a [200], BCIC IV 2b [200], HGD [26]	4-38	Raw EEG	time-series	CNN	Linear	Adam
[220]	BCIC IV 2a [200], BCIC IV 2b [200], Private	4-32	Raw EEG	time-series	CNN	ELU	SGD
[221]	BCIC IV 2a [200], HGD [26]	7-125	Raw EEG	time-series	CNN	ELU	Adam
[222]	BCIC IV 2a [200], HGD [26]	0-40	Raw EEG	time-series	CNN	ELU	Adam
[26]	BCIC IV 2a [200], HGD [26]	0-125	Raw EEG	time-series	CNN	ELU	SGD
[223]	BCIC IV 2b [200]	1-45	Raw EEG	time-series	CNN	ReLU	SGD
[224]	BCIC IV 2b [200]	8-30	Raw EEG	time-series	CNN	ReLU	Adam
[230]	BCIC IV 2b [200]	8-30	Raw EEG	time-series	DBN (RBM)	Sigmoid	Adadelta
[225]	eegmmidb [201]	N/M	Raw EEG	time-series	CNN, LSTM	N/M	Adam
[209]	eegmmidb [201]	N/M	Raw EEG	time-series	CNN, RNN + AE + XGBoost	ReLU	RMSProp
[226]	eegmmidb [201]	N/M	Raw EEG	time-series	CNN	ReLU	Adam
[228]	eegmmidb [201]	N/M	Raw EEG	time-series	LSTM RNN	Sigmoid, tahn	Adam
[229]	eegmmidb [201]	N/M	Raw EEG	time-series	LSTM	N/M	Adam
[231]	BCIC II 3 [197]	8-30	Complex Morlet wavelets	images	CNN	ReLU	SGD
[232]	BCIC II 3 [197]	8-30	CWT	images	CNN	ReLU	GD
[233]	BCIC III [198]	N/M	WPT	features	DBN	N/M	N/M

Continued on next page

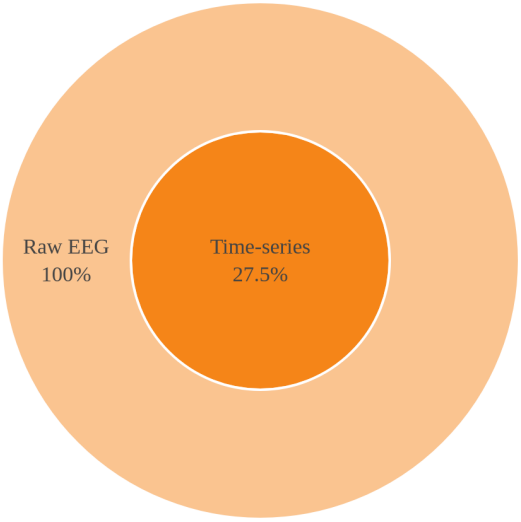
Table 3.1 – *Continued from previous page*

Ref	Dataset	Freq. range (Hz)	Feature extraction	Input	DL architecture	Activation function	Optimizer
[234]	BCIC III [198], BCIC IV [200], Private	8-26	CNN+ DWT	features	LSTM	ReLU	N/M
[235]	BCIC III 4a [198]	N/M	EMD, CSP	features	CNN	ReLU	N/M
[236]	BCIC III 4a [198]	0.5-90	BSS, CWT	features	CNN	ReLU	N/M
[237]	BCIC III 4a [198]	8-15	CSP+ Sparse representation	images	CNN	ReLU	N/M
[238]	BCIC III 4a [198], ULM [203]	N/M	CWT	images	CNN	ReLU	GD
[239]	BCIC IV 2a [200]	4-38	3D matrices	images	CNN	ELU	Adam
[240]	BCIC IV 2a [200]	8-30	CSP	features	CNN	ReLU	Adam
[241]	BCIC IV 2a [200]	4-38	FBCSP	features	CNN, LSTM	ReLU	Adam
[242]	BCIC IV 2a [200]	8-35	1d-AX	features	LSTM	Sigmoid	Adam
[243]	BCIC IV 2a [200]	4-40	FBCSP	features	CNN	ReLU	Adam
[244]	BCIC IV 2a [200]	N/M	FBCSP	features	CNN	ReLU	Adam
[245]	BCIC IV 2a [200]	4-40	FBCSP, Hilbert transform	features	CNN	ReLU	SGD
[246]	BCIC IV 2a [200]	4-30	CSP	features	CNN	N/M	N/M
[247]	BCIC IV 2a [200]	8-30	CSP + FFTEM	images	CNN	ReLU	N/M
[248]	BCIC IV 2a [200], BCIC IV 2b [200]	8-36	FBCSP, PSO	features	CNN	ReLU	Adam
[211]	BCIC IV 2a [200], BCIC IV 2b [200]	8-30	FBCSP	features	GRU	ReLU, tanh	SGD
[249]	BCIC IV 2a [200], Private	9-30	FBCSP	features	CNN	ReLU	SGD
[250]	BCIC IV 2b [200]	4-35	CWT	images	CNN	ReLU	GD
[251]	BCIC IV 2b [200]	4-32	STFT	images	CNN	ReLU	N/M
[212]	BCIC IV 2b [200]	8-30	CSP	features	ELM	Sigmoid	N/M
[252]	BCIC IV 2b [200]	8-35	FFT, WPD	images	DBN	Sigmoid	Conjugate GD
[204]	BCIC IV 2b [200]	8-30	STFT	images	CapsNet	SELU	SGD
[205]	BCIC IV 2b [200]	4-38	STFT	images	CapsNet	SELU	SGD
[253]	BCIC IV 2b [200], BCIC II 3 [197]	0.1-50	CWT	images	CNN	ReLU	GD
[206]	BCIC IV 2b [200], BCIC II 3 [197]	6-30	STFT	images	CNN + SAE	ReLU	GD
[254]	BCIC IV 2b [200], Private	8-30	EMD	features	CNN	ReLU	SGD

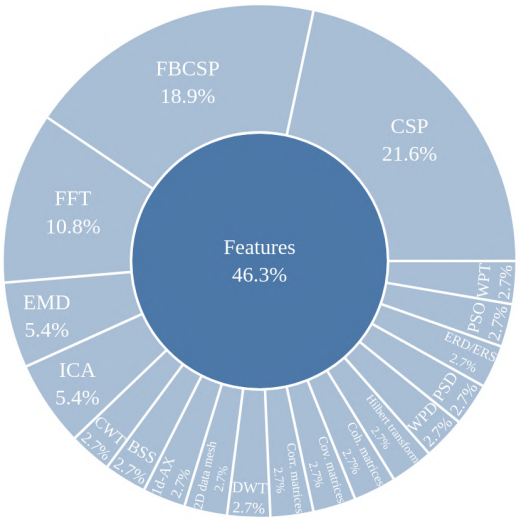
Continued on next page

Table 3.1 – *Continued from previous page*

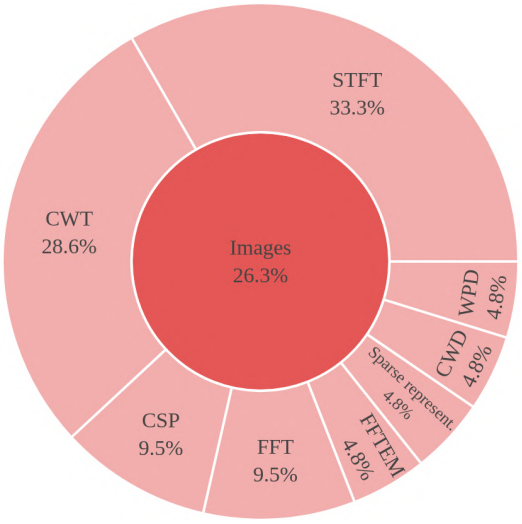
Ref	Dataset	Freq. range (Hz)	Feature extraction	Input	DL architecture	Activation function	Optimizer
[255]	BCIC IV 2b [200], Private	8-30	CSP	features	CNN	ReLU	Adam
[210]	BCIC IV 2b [200], Private	6-30	STFT	images	CNN, VAE	ReLU	GD
[256]	eegmmidb [201]	3-30	FFT	images	CNN	N/M	N/M
[257]	eegmmidb [201]	0.5-90	ERD/ERS + ICA	features	FNN	N/M	N/M
[258]	eegmmidb [201]	8-30	ICA	features	FNN	N/M	N/M
[259]	eegmmidb [201]	N/M	2D EEG data mesh	features	CRNN	N/M	Adam
[260]	eegmmidb [201]	N/M	AE	features	XGBoost	N/M	N/M
[261]	Private	8-30	CSP	features	Recurrent CNN	SReLU	N/M
[208]	Private	N/M	Covariance matrices	features	CNN, RNN + AE+ XGBoost	ReLU	Adam
[262]	Private	5-40	CSP, PSD	features	DNN	N/M	N/M
[263]	Private	5-30	STFT	images	CNN, LSTM	ReLU, ELU, SELU	SGD
[264]	Private	8-35	FFT, WPD	features	DBN	N/M	GD
[265]	Private	8-30	CSP	features	CNN	ReLU	SGD
[266]	Private	0.5-32.5	CWD	images	CNN	ReLU	SGD
[267]	Private	0.1-40	Correlation matrix, coherence matrix	features	CNN	ReLU	Adam
[268]	Private	8-30	FFT	features	DBN	N/M	N/M
[269]	Private	8-30	FFT	features	CNN	tanh, sigmoid	N/M
[270]	Private	0.1-100	FFT	features	DNN	N/M	GD
[271]	Private	8-30	STFT	images	CNN	ReLU, ELU, SELU	SGD
[272]	ULM [203]	0.6-45	CWT	images	CNN	ReLU	Adam



(a) Proportion use of feature extraction methods for time-series input formulation across all the reviewed articles.



(b) Proportion use of feature extraction methods for features input formulation across all the reviewed articles.



(c) Proportion use of feature extraction methods for images input formulation across all the reviewed articles.

Figure 3.13: Detailed proportion use of 3 different input formulation types across all the reviewed articles with the used feature extraction methods for achieving them.

3.6 Proposed architectures for motor intention detection

Summarizing the review in section 3.5, there were observed numerous studies and the most promising architectures were defined. In the case of analyzing small datasets of EEG signals, which are usually noisy and non-stationary, CNN architectures with different regularizing techniques were chosen for the following implementation. A CNN architecture allows working with a small amount of data because of its reduced number of parameters. It was also decided to use raw EEG signals as an input to neural networks to allow CNN to extract features by itself. In this section, different architectures of DNNs, which will be used to obtain results in chapter 5 will be described.

Table 3.2: Parameters of the D-ConvNet Architecture. C is the number of channels and $N = 2$ is number of classes. $T = 3$ s x 128 Hz $= 384$ is the number of samples to consider for classification.

Layer	# filters / (kernel size)	Output
Input		(1, C, $T = 384$)
Conv2D	25 / (1, 5)	(25, C, 380)
Conv2D	25 / (C, 1)	(25, 1, 380)
BatchNorm		(25, 1, 380)
Activation(ELU)		(25, 1, 380)
MaxPool2D	(1, 2)	(25, 1, 190)
Dropout		(25, 1, 190)
Conv2D	50 / (1, 5)	(50, 1, 186)
BatchNorm		(50, 1, 186)
Activation(ELU)		(50, 1, 186)
MaxPool2D	(1, 2)	(50, 1, 93)
Dropout		(50, 1, 93)
Conv2D	100 / (1, 5)	(100, 1, 89)
BatchNorm		(100, 1, 89)
Activation(ELU)		(100, 1, 89)
MaxPool2D	(1, 2)	(100, 1, 44)
Dropout		(100, 1, 44)
Conv2D	200 / (1, 5)	(200, 1, 40)
BatchNorm		(200, 1, 40)
Activation(ELU)		(200, 1, 40)
MaxPool2D	(1, 2)	(200, 1, 20)
Dropout		(200, 1, 20)
Flatten		(4000)
Dense(Softmax)	$N = 2$	(2)

3.6.1 Deep Convolutional Network

Deep Convolutional Network (D-ConvNet) has four convolution-max-pooling blocks, with a special first block designed to handle the EEG input. It is followed by three standard convolution-max-pooling blocks and a dense softmax classification layer. The first convolutional block was split into two convolutional layers to better handle a large number of input channels – one input channel per electrode.

D-ConvNet was proposed in [26] and details of this network architecture can be found in Table 3.2.

3.6.2 Shallow Convolutional Network

The architecture designed for the Shallow Convolutional Network (S-ConvNet) has been inspired by the Filter Bank Common Spatial Pattern (FBCSP) algorithm. The transformations performed by the S-ConvNet are similar to the transformations of FBCSP. Concretely, the first two layers of the S-ConvNet perform temporal and spatial convolutions, as in the D-ConvNet. These steps are analogous to the band-pass filter and the CSP spatial filter steps in FBCSP. In contrast to the D-ConvNet, the temporal convolution of the S-ConvNet had a larger kernel size, allowing a larger range of transformations in this layer. After the two convolutions of the S-ConvNet, a squaring nonlinearity x^2 activation function, a mean pooling layer, and a logarithmic activation function $log(x)$ followed; together these steps are analogous to the trial log-variance computation in FBCSP.

S-ConvNet was proposed in [26] and details of this network architecture can be found in Table 3.3.

Table 3.3: Parameters of the S-ConvNet Architecture. C is the number of channels and $N = 2$ is number of classes. $T = 3$ s x 128 Hz $= 384$ is the number of samples to consider for classification. The $square$ and log activation functions correspond to x^2 and $log(x)$, respectively.

Layer	# filters / (kernel size)	Output
Input		(1, C, $T = 384$)
Conv2D	40 / (1, 13)	(40, C, 372)
Conv2D	40 / (C, 1)	(40, 1, 372)
BatchNorm		(40, 1, 372)
Activation($square$)		(40, 1, 372)
AveragePool2D	(1, 35), stride (1, 7)	(40, 1, 49)
Activation(log)		(40, 1, 49)
Flatten		(1960)
Dropout		(1960)
Dense(Softmax)	$N = 2$	(2)

3.6.3 EEGNet-2.32

EEGNet architecture is presented on Figure 3.14. The first layer of the EEGNet is a temporal convolution to learn frequency filters. Then follows a depthwise convolution layer, connected to each feature map individually to learn frequency-specific spatial filters. The separable convolution is the next convolutional layer; it is a combination of a depthwise convolution, which learns a temporal summary for each feature map individually, followed by a pointwise convolution, which learns how to optimally mix the feature maps together [27].

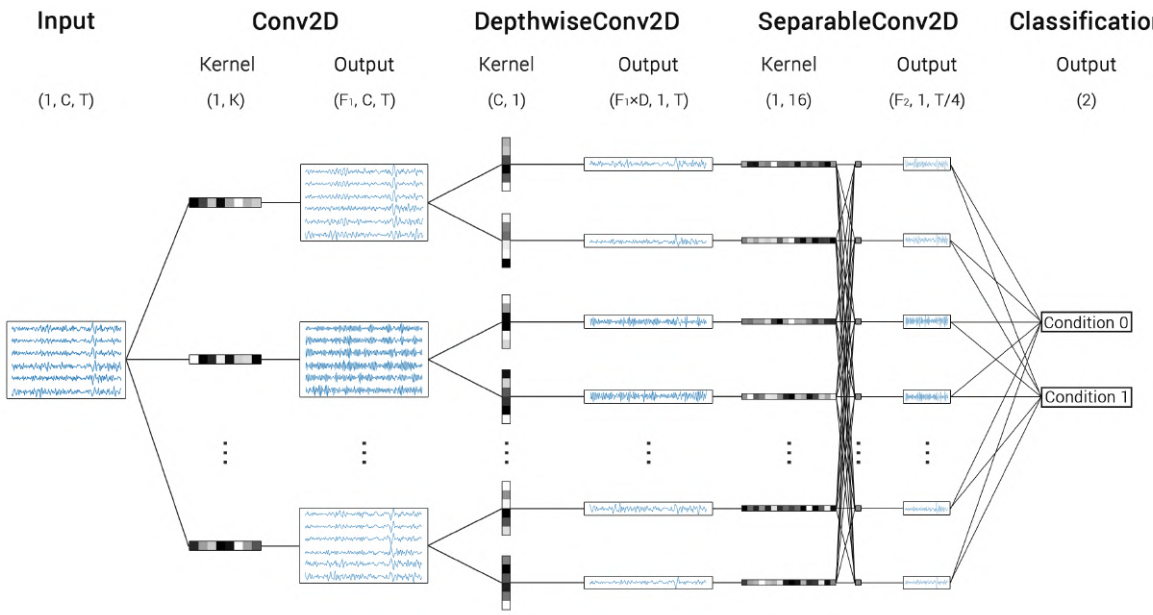


Figure 3.14: EEGNet architecture. Where the number of channels $C = 6$, the number of time samples T, kernel length for the first convolutional layer $K = 8$, multiplier depth $D = 2$, the number of temporal filters F_1 and the number of pointwise filters F_2.

3.6.4 EEGNet-4.8

We tested different sets of parameters (not detailed in the thesis) and select 8 models derived from the original EEGNet which allow us to obtain the best result. We kept 8 temporal filters $F_1 = 8$ and 16 pointwise filters $F_2 = 16$ for all models presented in this thesis. The parameters which have been changing were lengths of temporal kernel size for the first convolutional layer K and multiplier depth D (which determines the number of spatial filters) of the depthwise convolution layer. Thus we set multiplier depth to 2 or 4.

The first layer of filters of the original EEGNet [27] extracts mainly a low frequency information since it has large kernel size $(1, 32)$, and ignores mid- and high frequency information [273]. To remedy this problem, we reduced the first layer kernel size (as in [273]). So for this study the kernel size was set to $(1, 32)$ (as recommended in [27] for sensory motor data), $(1, 16)$, $(1, 8)$ and $(1, 4)$. These new architectures retain much more information in the first layer features concerning the mid and high-frequency bands. More importantly, they also improve the classification performance as shown in section 5.3.2.

We use the notation EEGNet-D.K to denote the number of spatial filters to learn and kernel length for the first convolutional layer; i.e.: EEGNet-4.8 denotes learning 4 spatial filters per temporal filter and kernel length $(1, 8)$. Full details about the network architecture can be found in Table 3.4.

Table 3.4: Parameters of the EEGNet architecture. Where C is the number of channels, T is the number of time samples per channel, F_1 is the number of temporal filters, D is the multiplier depth, F_2 is the number of pointwise filters and N is the number of classes. We use 8 temporal filters $F_1 = 8$, 4 multiplier depth $D = 4$, 16 pointwise filters $F_2 = 16$ and (1, $K = 8$) for the lengths of temporal kernel size for the first convolutional layer.

Layer	# filters / (kernel size)	Output
Input		(1, C, $T = 384$)
Conv2D	8 / (1, K)	(8, C, 384)
BatchNorm		(8, C, 384)
DepthwiseConv2D	4 * 8 / (C, 1)	(32, 1, 384)
BatchNorm		(32, 1, 384)
Activation(ELU)		(32, 1, 384)
AveragePool2D	(1, 4)	(32, 1, 96)
Dropout*		(32, 1, 96)
SeparableConv2D	16 / (1, 16)	(16, 1, 96)
BatchNorm		(16, 1, 96)
Activation(ELU)		(16, 1, 96)
AveragePool2D	(1, 8)	(16, 1, 12)
Dropout*		(16, 1, 12)
Flatten		(192)
Dense(Softmax)	$N = 2$	(2)

3.6.5 Training procedure

All deep neural networks were implemented with the TensorFlow framework [274] and Keras [188] and trained using the Grid'5000 testbed [275] to speed up the computations. The training of the networks was performed with the following configurations.

- ADAM [176] is used as the optimization method. The parameters are set to default values as proposed in [176].
- Categorical cross-entropy is used as the optimization criterion.
- Batch size is selected as 16.
- We run 300 training iterations (epochs) and perform validation stopping, saving the model weights which produced the lowest validation set loss.

- We used a batch normalization [192] layer before the activation layer
- We used a dropout layer after the activation with a probability of 50%. Our investigations showed that 50% dropout probability leads to better (approximately 2%) performance comparing to 25% (as proposed in [27]) and helps to prevent overfitting on datasets with a small number of samples.

The number of training parameters for each neural network is represented in Table 3.5. We could see from the table that EEGNet architectures have much smaller number of parameters comparing to D-ConvNet and S-ConvNet.

Table 3.5: Number of parameters for different neural networks

	Number of channels					
Model	128	47	13	9	6	3
D-ConvNet	214,077	163,452	142,202	139,702	137,827	135,952
S-ConvNet	209,442	79,842	25,442	19,042	14,242	9,442
EEGNet-2.32	3,362	2,066	1,522	1,458	1,410	1,362
EEGNet-4.8	5,794	3,202	2,114	1,986	1,890	1,794

3.7 Conclusion

In this chapter, we have introduced the concept of artificial neural networks. We have represented activation functions and optimization algorithms for artificial neural networks. We have selected Adam for our experiments in section 5.3 as it allows us to use less time and train more efficiently.

Also, we have presented convolutional neural networks, as they could be used for the direct feature extraction and classification from raw EEG signals. We have introduced their architecture and layers, including depthwise convolutional layer and depthwise separable convolutional layer.

Moreover, we have presented the overfitting problem and regularization techniques to prevent it. Overfitting of neural networks and bad generalization due to the small amount of data available for training is one of the reasons why deep learning application is still limited in the MI-based BCI domain. Another challenge is the non-stationarity and low signal-to-noise ratio of the EEG signals recorded in a realistic setting. The above problems make it impossible to use very deep architectures with many layers and require the application of regularization techniques. In this thesis, we will use L1 and L2 regularization for layer's kernels, dropout layer, and batch normalization layer in proposed architectures for MI detection.

In this chapter, we reviewed existing deep learning solutions for the analysis of EEG signals for MI-based BCI in section 3.5. We have selected CNN architectures for raw EEG signal classification such as Deep Convolutional Network, Shallow Convolutional Network, and EEGNet. Also, we proposed the adaptation of EEGNet, namely EEGNet-4.8 for the following investigations for the motor imagery detection in section 3.6).

Also, we have presented the procedure and parameters that will be used for neural network training in further investigations.

In the next chapter, we will present datasets with motor imageries, which we will use in our experiments.

Classification results for deep learning architectures presented in this chapter will be introduced in section 5.3 of chapter 5.

Chapter 4

Materials

Contents

In the last chapters, we have presented machine learning and deep learning methods that could be used to improve feature extraction and classification of motor imageries in BCIs.

In this chapter, we present motor imagery (MI) data sets that will be used for our experiments. Participants, EEG channel layouts paradigms, time schemes, and experimental tasks will be described. Preprocessing and cross-validation details will be discussed as well.

To improve motor imagery detection, we have selected 3 different datasets for experiments.

Firstly in section 4.1, we will present a classic 2-class MI dataset [276] containing right hand and left hand motor imageries.

Secondly, in section 4.2, we will introduce a 8-class MI dataset with combined movement MIs. This is a database that consisted of 7 different MI tasks and resting state. All the data was recorded by members of the NeuroRhythms team in the LORIA laboratory.

Finally, in section 4.3, we will present a dataset consisted of median nerve stimulations during motor imagery and during rest, as well as MI condition separately. All the data was recorded by members of the NeuroRhythms team in the LORIA laboratory.

4.1 2-class motor imagery dataset

We used an open-source 2-class motor imagery dataset described in details in [276]. It is available via PhysioNet [202] web-site. This dataset contains right hand and left hand motor imageries. We used this dataset in the investigation on estimating the quality of motor imageries in BCI using outlier detection. Results are represented in section 5.1.

4.1.1 Participants

The database contains EEG signals for BCIs recorded by NIRx GmbH, Berlin, Germany [276]. The experiment involved 29 right-handed and one left-handed healthy subjects (14 males and 15 females, average age 28.5 ± 3.7 years) who were asked to perform imaginary movements with their right and left hands. All of them were healthy, without reported neurological, psychiatric or other brain-related diseases. All volunteers were informed about the experimental procedure and written consent was obtained from all participants. This study was conducted according to the Declaration of Helsinki [277] and was approved by the Ethics Committee of the Institute of Psychology and Ergonomics, Technical University of Berlin [276].

4.1.2 Experimental tasks

Subjects were supposed to perform kinesthetic MI of their right or left hand at the audible signal for three sessions. More precisely, task consisted of imaginations the opening and closing hands as they were grabbing a ball. An arrow pointing to the left or right side appeared on the screen for two seconds as a visual instruction. After the arrow disappeared, a black fixation cross was displayed during the task period for ten seconds. The subjects were asked to mentally open and close their hands with a 1 Hz pace. MI was accomplished continuously over the task period. The task period was finished with a short beep sound and a 'stop' displayed for 1 s on the screen. Each session contained 20 imaginary movements (10 trials for each left and right hand MI in a single session). So 30 trials for each MI in total [276].

4.1.3 Paradigm and time scheme

The subjects sat on a comfortable armchair in front of a 50-in white screen. The distance between their heads and the screen was 1.6 m. They were asked not to move any part of the body during the data recording. The experiment consisted of three sessions of left and right hand MI. Each session comprised a 1 min pre-experiment resting period, 20 repetitions of the given task and a 1 min post-experiment resting period. The task started with 2 s of a visual introduction of the task, followed by 10 s of a task period and resting period which was given randomly from 15 to 17 s. At the beginning and end of the task period, a short beep (250 ms) was played. Figure 4.1 shows the schematic diagram of the experimental paradigm.

4.1.4 Electroencephalographic data

EEG data was recorded by a multichannel BrainAmp EEG amplifier with thirty active electrodes (Brain Products GmbH, Gilching, Germany) with linked mastoids reference. In the experiment a custom-made stretchy fabric cap (EASYCAP GmbH, Herrsching am Ammersee, Germany) consisted of 30 EEG electrodes placed according to the international 10-5 system (AFp1, AFp2, AFF1h, AFF2h, AFF5h, AFF6h, F3, F4, F7, F8, FCC3h, FCC4h, FCC5h, FCC6h, T7, T8, Cz, CCP3h, CCP, CCP5h, CCP6h, Pz, P3, P4, P7, P8, PPO1h, PPO2h, POO1, POO2, and Fz) was used.

Original dataset includes also electrooculogram (EOG), electrocardiogram (ECG), respiration with a piezo based breathing belt and NIRS data collected from subjects. But we have used only EEG signals in our experiments.

EEG signals were recorded at a sampling rate of 1000 Hz and downsampled to 200 Hz.

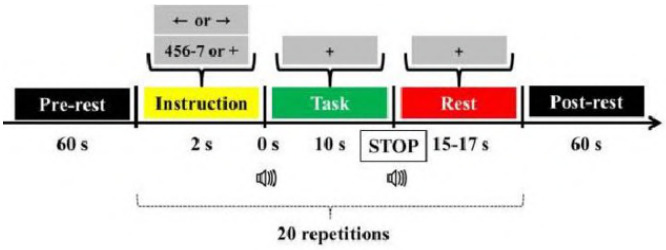


Figure 4.1: Schematic sequence diagram of the experimental paradigm. Each session comprised a 1 min pre-experiment resting period, 20 repetitions of the given task and a 1 min post-experiment resting period. The task started with 2 s of a visual introduction of the task, followed by 10 s of a task period and resting period which was given randomly from 15 to 17 s. At the beginning and end of the task period, a short beep (250 ms) was played [276].

4.1.5 Data preprocessing

Signals from eight electrodes (FCC3h, FCC4h, FCC5h, FCC6h, CCP3h, CCP4h, CCP5h, CCP6h) corresponding to the motor areas of the motor cortex are filtered in the alpha and beta frequency bands. The 5th order Butterworth bandpass filter is used in the alpha (8-12 Hz), lower-alpha (7-10 Hz), upper alpha (11-13 Hz), beta (17-25 Hz), and alpha + beta (7- 30 Hz) ranges.

4.1.6 Feature extraction

For this dataset ERD/ERS feature extraction method was used (see section 2.2.2.1). Based on the obtained filtered signals, a set of features based on event-related desynchronization (ERD) and synchronization (ERS) was calculated [49].

Firstly, we filtered signals from 8 electrodes (described earlier in section 4.1.5) in different bands (four frequency bands were used to obtain ERD features and two frequency bands were used to obtain ERS features).

To obtain ERD features, signals from eight electrodes (described earlier in section 4.1.5) were filtered in (7-30 Hz), (7-10 Hz), (11-13 Hz), and (8-12 Hz) frequency bands for the time window from 0.5 to 5.5 sec. ERS features are obtained from the cropped signals in another time window (7-12 sec) using (7-30 Hz) and (17-25 Hz) frequency ranges.

These features describe processes that occur in localized areas of the motor cortex of the brain [99, 100].

Thereby using 8 electrodes and 6 filters 48 features were obtained.

4.1.7 Cross-validation

We have adopted a subject-specific approach for which data from each subject are investigated separately. In this experiment, we used two loops of cross-validation.

For the outer loop, we used the One-Session-Out approach. And for the inner loop of cross-validation, we used the Leave-One-Out approach. An inner loop was used to select the best parameter of saturation data with outliers.

4.1.7.1 Subject-specific One-Session-Out approach

For the cross-validation, it was chosen subject-specific One-Session-Out approach which is the variation of the K-Fold cross-validation with K equals to the number of sessions in the experiment. This approach is close to real conditions as it does not allow data leakage between different sessions.

Data are splitted into test and training sets as follows: two sessions for training, one session for testing.

Summarizing training set consisted of 40 observations for each subject, while the test set consisted of 20 observations. Three round of cross-validation was performed and then results were averaged.

4.1.7.2 Subject-specific Leave-One-Out approach

Using this cross-validation strategy, one observation remains for the single-item test set, and all the others are used for training. Therefore, the number of folds using Leave-One-Out cross-validation equals the number of

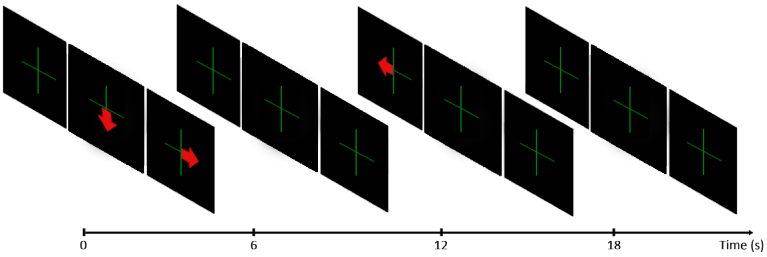


Figure 4.2: Time scheme for motor imagery 8-class database (from [150]).

observations in the data set.

4.2 Motor imagery 8-class database

This is a database that consisted of 7 different MI tasks and resting state. All the data was recorded by members of the team NeuroRhythms (Inria/LORIA) during the PhD thesis of Cecilia Lindig-Léon [278]. All the experiments were carried out with the consent agreement (approved by the ethical committee of INRIA) of each participant and following the statements of the WMA Declaration of Helsinki on ethical principles for medical research involving human subjects [277].

4.2.1 Participants

Ten right-handed healthy volunteer subjects took part in this experiment (7 men and 3 women, 19 to 43 years old). They had no medical history which could have influenced the experiment.

4.2.2 Paradigm and time scheme

Subjects were seated in a comfortable chair with the arms at their sides in front of a screen showing the task cue to be performed, which consisted of one of the seven kinesthetics motor imageries that it is possible to generate with all combinations including the use of the right hand, left hand, both feet together, and rest (i.e., right hand; left hand; feet; both hands together; left hand in combination with feet; the right hand in combination with feet; both hands together with feet, and rest). The whole session consisted of 4 runs, containing each one 10 trials per task, which totals 40 trials per class (320 trials considering the 8 classes).

For the stimulation, three panels were simultaneously displayed on the screen, which was associated from left to right, to the left hand, feet, and right hand. Each trial was randomly presented and lasted for 12 seconds, starting at second 0 with a cross at the center of each panel and an overlaid arrow indicating for the next 6 seconds the task to be performed: an arrow pointing to the left side on the left panel for the left hand, an arrow pointing down on the central panel for feet, an arrow pointing to the right side on the right panel for the right hand, and the simultaneous combination of all of them for the corresponding combined tasks (see Figure 4.2. The rest condition was also considered and it was indicated by the absence of arrows. After the second 6, the task cue disappeared and the crosses were remaining for the next 6 seconds indicating the pause period before the next trial started.

4.2.3 Electroencephalographic data

EEG signals were recorded by the OpenViBE [279] platform from ten right-handed subjects at 256 Hz using a commercial REFA amplifier developed by TMS International^{TM}.

Each subject performed 4 runs of motor imageries. Each run contains 10 trials for each class (Rest, Left Hand, Right Hand, Both Hands, Feet, Left hand & Feet, Right hand & Feet, Both hands & Feet). Thus 40 trials per class are available for each subject.

4.2.4 EEG channels layout

For this experiment EEG cap was fitted with 26 passive electrodes, namely Fp1; Fp_z; Fp2; F_z; FC5; FC3; FC1; FC_z; FC2; FC4; FC6; C5; C3; C1; C_z; C2; C4; C6; CP5; CP3;CP1; CP_z; CP2; CP4; CP6 and P_z.

Such electrodes were placed by using the international 10-10 system positions to cover the primary sensorimotor cortex.

4.2.5 Data preprocessing

Firstly, EEG signals were re-referenced with respect to the common average reference across all channels.

Trials were considered 2 seconds before the task cue appeared on the screen and 6 seconds after it disappeared so that each segment corresponds to a 14s-length signal segment over 26 electrodes.

Raw EEG signals were filtered by 5th order Butterworth band-pass filter between alpha, beta, and alpha+beta bands.

4.2.6 Classification using multi-label approach

4.2.6.1 Feature extraction

To evaluate more precisely the modulation which appeared during the two different time windows, we computed the ERD/ERS using the "band power method" [49] (see Figure 5.4, Figure 5.5 and Figure 5.6).

From each trial, 26 features (one per channel) have been extracted from the filtered signal computing the logarithm of the variance of the amplitude values between 0.5 and 3.5s after the GO signal.

Linear Discriminant Analysis (LDA) [142] (see section 2.5.1) has been used via the Python library scikit-learn [280].

4.2.6.2 Cross-validation

For the cross-validation, it was chosen the One-Session-Out approach as well as described in section 4.1.7. Due to the small number of samples per class, cross-validation has been done using 3 runs for train and 1 run for test. Thus it was used 240 trials for the train and 80 trials for the test in every step of the cross-validation.

4.2.7 Multi-label classification using 3 binary classifiers

4.2.7.1 Feature extraction and classification

As we planned to detect electrical activity in the areas of the primary motor cortex, only 3 channels of EEG (C3, C4, Cz) were chosen for this experiment. Channel C3 corresponds to the right hand movement/imagery, C4 - left hand, Cz - feet.

Six sets of features were investigated:

- 3-feature set (ERD): alpha+beta band (7-30 Hz) for 3 channels of EEG described above;
- 3-feature set (ERS): alpha+beta band (7-30 Hz);
- 6-feature set (ERD & ERS on alpha+beta band): a combination of two sets - alpha+beta band (7-30 Hz);
- 6-feature set (ERD on alpha and ERS on beta band): alpha band (8-12 Hz) and beta band (17-25 Hz) each for 3 channels of EEG described above;
- 18-feature set: low-alpha band (7-10 Hz), upper-alpha band (11-13 Hz), low-beta band (15-17 Hz), mid-beta band (18-20 Hz), mid-beta band (21-23 Hz) and upper-beta band (24-26 Hz) each for 3 channels of EEG described above;
- 30-feature set: ERD filtered in following bands (7-10 Hz, 11-13 Hz, 8-12 Hz, 7-30 Hz) and ERS filtered in following bands (15-17 Hz, 18-20 Hz, 21-23 Hz, 24-26 Hz, 17-25 Hz, 7-30 Hz) each for 3 channels of EEG described above.

For all filtered signals ERD/ERS and logarithmic ERD/ERS were calculated. ERD is observed during and immediately after the motor imagery task and better represented in the alpha band, while ERS is observed after the task and better expressed in the beta band. That's why features filtered in alpha bands (ERD) were calculated during 0.5-6 seconds of each trial, and features filtered in beta bands (ERS) were calculated from 7 to 12 sec.

As a classifiers were used following algorithms:

- Shrinkage Linear Discriminant Analysis (see section 2.5.1);
- Support Vector Machine with linear, rbf, sigmoid and polynomial kernels (see section 2.5.2) ;
- Decision Tree Classifier (see section 2.5.3).

4.2.7.2 Cross-validation

For the cross-validation, it was chosen the One-Session-Out approach as well as described in section 4.1.7. Cross-validation was implemented using 3 runs for train and 1 run for test. To create the training set we used the 'One vs. All' approach, which means that we labeled one class as '1' and the other three classes as '0' for each of three classifiers:

- classifier to detect right hand activity (right hand vs. rest & left hand & feet).
- classifier to detect left hand activity (left hand vs. rest & right hand & feet).
- classifier to detect feet activity (feet vs. rest & left hand & right hand).

Each training set had the same number of observations of each of the classes in every step of cross-validation (for example 30 samples of right hand imageries vs. 10 rest + 10 left hand + 10 feet).

The test set consists of 1 run of each of the 8 classes, it equals 80 trials (10 per class).

4.3 Median nerve stimulation database

A median nerve stimulation database is used in our investigation of motor intention detection with deep learning to prevent intraoperative awareness.

This database consisted of median nerve stimulations during motor imagery and during rest, as well as MI condition separately. All the data was recorded by members of the team NeuroRhythms (Inria/LORIA) Pierre Riff and Sébastien Rimbert during the PhD thesis of Sébastien Rimbert [20].

This experiment followed the statements of the WMA Declaration of Helsinki on ethical principles for medical research involving human subjects [277]. Besides, participants signed an informed consent which was approved by the ethical committee of Inria (COERLE, approval number: 2016-011/01) as it satisfied the ethical rules and principles of the institute.

The aim of the experiments with this database is to detect the occurrence of motor patterns under two different conditions: median nerve stimulation during motor imagery (MI + MNS) and median nerve stimulation during rest (MNS) (see Figure 4.3). These two conditions aim at showing that a MNS can be used and is more helpful as a trigger to improve the detection of intraoperative awareness. The reason for using MNS is its ability to modulate the motor cortex activity by first generating an ERD during stimulation and then an ERS post-stimulation, which resulted in a higher accuracy of classification [20] (see section 1.3.4).

4.3.1 Participants

Twenty-two right-handed healthy volunteers (12 females; 19 to 57 years old; 28.56 $\pm$ 13.3 years old) were recruited for this study. All subjects satisfied the inclusion criteria (right-handed, between 18-60 years old, without medical history which could have influenced the task), such as diabetes, antidepressant treatment or neurological disorders [20].

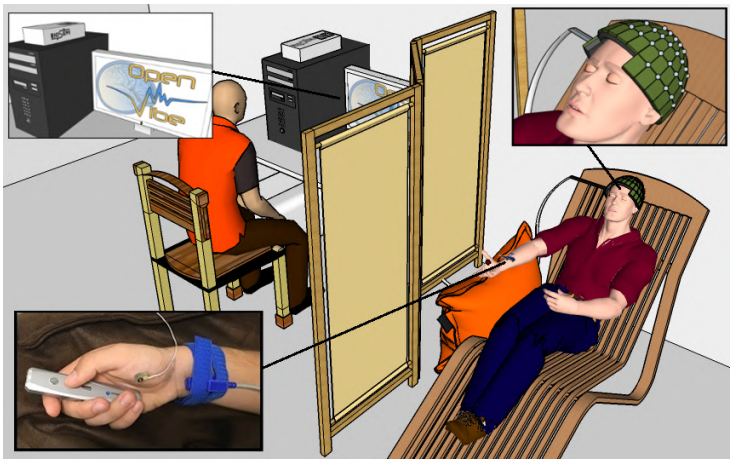


Figure 4.3: Schematics of the experimental setting (from [20]).

4.3.2 Paradigm and time scheme

During EEG recordings (Figure 4.3), the subject was lying on a comfortable chair with eyes closed. Their legs rest on a footrest and the right forearm rests on a cushion to prevent movement. The subject mentally presses and releases a remote button. The operator monitors the EEG signals during the experiment.

Participants performed each condition randomly during two runs of 26 trials per condition. So in total, the database consists of 52 recordings of each class for every subject.

As an additional inclusion criterion for participants, we considered the fact that the stimulation was not felt as painful and that it caused a slight movement between the thumb and the index finger. The stimulation intensity was adapted for each subject and varied between 8 and 15 mA.

Condition 1: motor imagery

For the MI condition (C1), subjects had to imagine an isometric grasp between the thumb and the index finger on a pointer button, i.e., they had to try to feel a maximum of sensations caused by the real movement, but without any movement. A low-frequency beep indicated when the subject had to start the motor imagery, the grasping MI was maintained during 2 s, then a second beep indicated the end of the imagined task [20] (Figure 4.4).

Condition 2: motor imagery with a median nerve stimulation

During the MI + MNS condition (C2), subjects had to perform a motor imagination while their median nerve was stimulated at 750 milliseconds after the start of the motor imaginary task (Figure 4.4). A low-frequency beep was indicated when the subject had to start the motor imagery and a high-frequency beep indicated when to stop it. We chose the 750 milliseconds of delay according to the reaction time of the average person, to stimulate during the ERD corresponding to the imagination started. The duration of stimulation was 100 ms [20].

Condition 3: median nerve stimulation only

The MNS only condition (C3) consisted of a series of stimulation of the median nerve during rest (Figure 4.4). We placed two stimulating electrodes on the wrist according to the standards [57, 281] and recorded the EEG data from subjects [20].

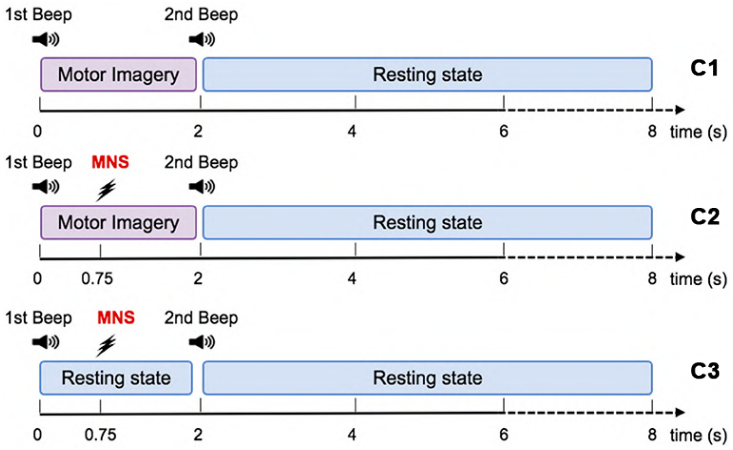


Figure 4.4: Representation scheme for one trial. Timing schemes of a trial for C1, C2, and C3. For all motor tasks, one low-frequency beep indicates when to start the task. For the MNS+MI condition, the MNS occurs at 750 milliseconds after the first beep. The end of the MI is announced by a high-frequency beep and followed by a rest period of 6 seconds (adopted from [20]).

4.3.3 Electroencephalographic data

EEG signals were acquired using the *OpenViBE* platform [279] with a *BioSemi Active Two* 128-channel EEG system, arranged in the *BioSemi*'s ABC system covering the entire scalp (Figure 4.5) at 2048Hz. Among all registered sites, some of the electrodes were localized around primary motor cortex, motor cortex, somatosensory cortex, and occipital cortex (Figure 4.5), which allowed to observe the physiological changes due to kinesthetic motor imagery and median nerve stimulation [282, 283, 57, 47].

In BioSemi^{TM} system the ground electrodes used were two separate electrodes: Common Mode Sense (CMS) active electrode and Driven Right Leg (DRL) passive electrode located over the parietal lobe.

Impedance was kept below 10kΩ for all electrodes to ensure that the background noise in the acquired signal was low.

An external electromyogram (EMG) electrode was added to verify that there was no movement during the MI task.

4.3.4 EEG channel layouts

Initially 128 EEG channels were registered with *BioSemi*'s ABC layout (Figure 4.5). The recordings with a reduced number of channels were constructed to investigate the possible influence of using a smaller number of electrodes on the motor imagery detection performance.

We also investigated additional channels from the frontal lobe as an interesting area for our experiment because it is considered as the emotional controller center and also responsible for motor functions, motor speech functions and impulse control [284].

The following electrode setups covering mostly brain areas responsible for movements were tested (see Figure 4.6):

- 128 channels (all channels as present on Figure 4.5);
- 47 channels (Figure 4.6a) covered mainly motor cortex area (*A1, A2, A3, B1, B2, B15, B16, B17, B18, B19, B20, B21, B22, B23, B24, B25, B28, B29, B30, B31, B32, C1, C2, C23, C24, C22, C11, D1, D2, D9, D10, D11, D12, D13, D14, D15, D16, D17, D18, D19, D20, D21, D22, D25, D26, D27, D28*);
- 13 channels (Figure 4.6b) over motor cortex (*D12, C23, B31, D21, D19, D14, A1, B20, B22, B24, D28, A3, B18*);
- 13 channels (Figure 4.6c) over motor cortex and frontal lobe (*C29, C17, C16, D11, D13, D19, D27, D17, B32, B30, B22, B19, B17*);

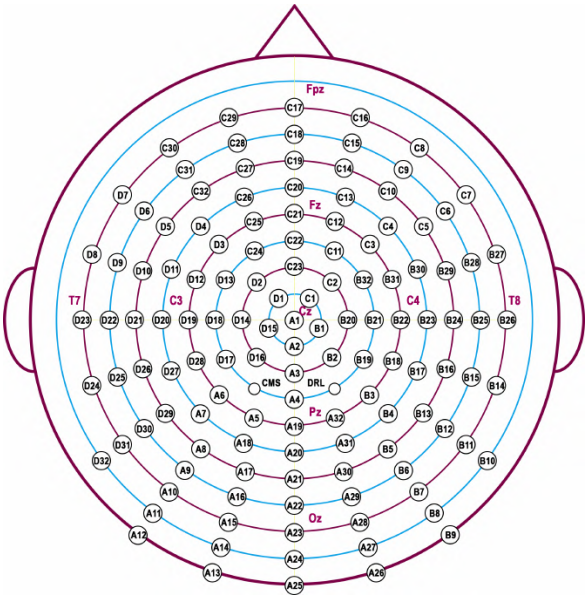


Figure 4.5: BioSemi 128 channels headcap ABC layout

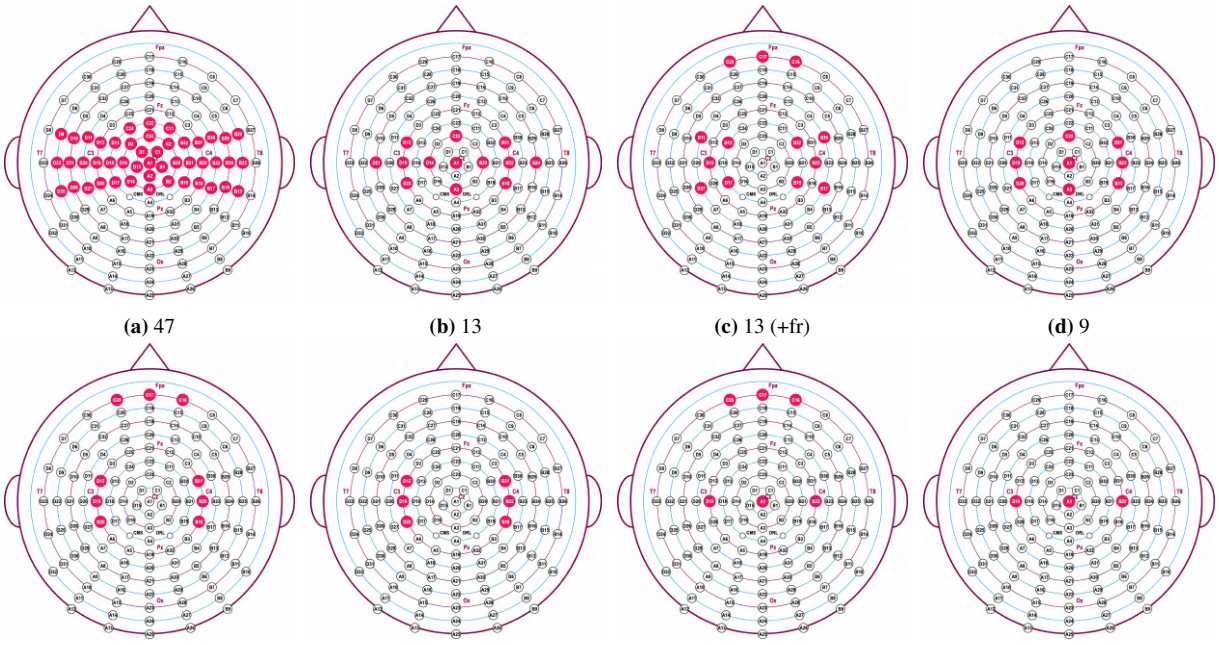
- 9 channels (Figure 4.6d) over motor cortex (*D12, D19, D28, C23, A1, A3, B31, B22, B18*);
- 9 channels (Figure 4.6e) over motor cortex and frontal lobe (*D12, D19, D28, C29, C17, C16, B31, B22, B18*);
- 6 channels (Figure 4.6f) over motor cortex (*D12, D19, D28, B31, B22, B18*);
- 6 channels (Figure 4.6g) over motor cortex and frontal lobe (*C29, C17, C16, D19, A1, B22*);
- 6 channels only over left side of motor cortex and frontal lobe (*C30, C29, C17, D12, D19, D28*);
- 3 channels (Figure 4.6h) over motor cortex (*D19, A1 and B22* which corresponds to the electrodes *C3, Cz* and *C4* in International 10–20 System respectively).
- 3 channels over frontal lobe (*C29, C17, C16*)
- 3 channels over only left side of frontal lobe (*C30, C29, C17*)

4.3.5 Data preprocessing

After selecting a certain channel layout, EEG signals were re-referenced to a common average reference, then EEG signals were resampled at 128 Hz. While power modulations during motor activity are observed mainly in mu, alpha and beta bands of an electroencephalographic signal measured over the sensorimotor area of the brain [125, 285, 286] sampling rate of 128 Hz is enough to represent mu, beta, and low gamma frequency range according to the Nyquist-Shannon-Kotelnikov sampling theorem.

Then each signal was filtered in a specific frequency band using a 4th-order Butterworth band-pass filter [114] and divided into trials of 3 s duration starting at 250 ms after the first beep (Figure 4.4). This corresponds to 384 samples (3 s at 128 Hz) per each channel.

We investigated filtering in few frequency bands to see is it efficient enough to add theta (4-8 Hz) or low gamma (30-38 Hz) range to standard mu+beta (8-30 Hz) frequency band especially for MNS vs. MI+MNS discriminating task because it has not been studied before. Thereby we used the following bands: theta-beta (4-30 Hz), standard mu-beta (8-30 Hz), mu-gamma (8-38 Hz), and theta-gamma (4-38 Hz) as was observed in the section 3.5. We did not apply any feature extraction technique to these data. So filtered EEG signals were directly fed to the deep learning architectures as an input.



(a) 47 **(b)** 13 **(c)** 13 (+fr) **(d)** 9

(e) 9 (+fr) **(f)** 6 **(g)** 6 (+fr) **(h)** 3

Figure 4.6: Layouts with electrodes placement

4.3.6 Cross-validation

In section 5.3 we report subject-specific classification results corresponding to ten-fold blockwise cross-validation, where eight of ten blocks are chosen as the training set, one block as the validation set, and the remaining block as the test set.

4.4 Conclusion

In this chapter three different datasets were introduced that were used in our experiments: motor imagery 2-class dataset, motor imagery 8-class dataset containing combined movements, and median nerve stimulation dataset.

Participants, experimental tasks, time schemes, EEG data and channel layouts, preprocessing, feature extraction, classification approaches (if exists), and cross-validation strategies were described for 3 datasets.

Results of investigations on these datasets will be introduced in the following chapter 5. Firstly, we will represent results of outlier detection as quality index of motor imagery estimation in section 5.1 on 2-class motor imagery dataset introduced in section 4.1. Secondly, we will use data from the 8-class motor imagery dataset described in section 4.2 to improve combined movement motor imagery classification. We will show results for the classification of combined movement motor imageries using binary classifiers in section 5.2.2. Also, we will introduce results of feature selection for combined movement MI classification on the 8-class motor imagery dataset in section 5.2.3. Finally, we will represent results for motor imagery detection using deep learning to prevent intraoperative awareness in section 5.3 on the dataset introduced in section 4.3.

Chapter 5

Experimental results

Contents

In previous chapters, we have presented materials and methods for motor imagery detection from raw EEG signals.

More precisely, in chapter 1, we have introduced the background to BCIs, EEG signals, brain rhythms, and motor imageries as the most popular paradigm in the MI-based BCI domain. Also, we have presented the concept of BCI for accidental awareness during general anesthesia based on median nerve stimulation as one of the possible applications of BCI.

In chapter 2, we have introduced machine learning techniques for motor imagery detection. We have presented different stages of EEG signal processing, such as filtering, feature extraction, feature selection, outlier detection, and classification. Also, we have proposed new methods to improve motor imagery classification in this chapter (see section 2.4.1 and section 2.5.6).

In chapter 3, we have represented background to deep neural networks as a combined solution for feature extraction and classification from raw EEG signals. We have reviewed the application of deep learning for motor imagery detection and proposed convolutional neural network architectures for implementation.

In chapter 4, we have presented three different datasets as well as details of preprocessing, feature extraction, classification, and validation techniques used for these data.

In this chapter, we will present a series of results on different parts of the BCI processing. Each section presents results on the improvement for motor imagery detection.

Results of machine learning and deep learning application will be presented through this chapter for three different tasks:

- Outlier detection method presented in section 2.4.1 will be applied for motor imagery quality estimation prior training a BCI to improve accuracy (section 5.1);

- Application of three parallel-connected binary classifiers presented in section 2.5.6 for combined movement MI classification improvement (section 5.2);
- Deep learning methods presented in chapter 3 will be applied to median nerve stimulation-based BCI for the task of intraoperative awareness prevention (section 5.3).

5.1 Improving a BCI using outlier detection as quality index of motor imagery estimation

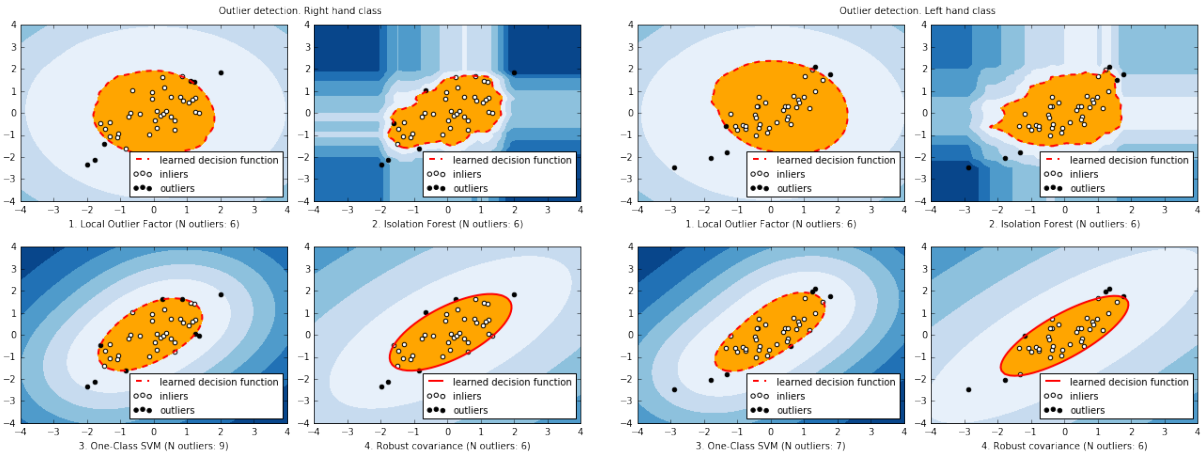
In this section, we propose to detect outliers in the data and exclude them before training. We use the method presented in section 2.4.1 for outlier detection to analyze the quality of data [287].

We use 2-class motor imagery dataset described previously in section 4.1. Dataset consists of EEG signals from 29 subject performing the right hand and the left hand motor imagery.

We use 48 ERD/ERS features from different channels filtered in different bands (described in section 4.1.6). Then, we apply outlier detection techniques on the feature set to find outliers in the training set and remove them using different saturation parameter separately for each class. We calculate classification accuracy using logistic regression for each parameter of saturation to choose outlier detectors with the best parameter of saturation and use them to clean test set. After removing the outliers, classifiers predict classes on the cleared test set.

As cross-validation, we have applied a subject-specific strategy with two loops: the One-Session-Out approach for the outer loop and the Leave-One-Out approach for the inner loop. An inner loop was used to select the best parameter of saturation data with outliers.

Classification accuracies obtained using the model described in section 2.4.1 are represented in Tables 5.1 and 5.2).



(a) Right hand MI

(b) Left hand MI

Figure 5.1: Results of outlier detection for three classes: (a) right hand MI, (b) left hand MIusing different methods: local outlier factor, Isolation forest, one-class SVM, robust covariance.

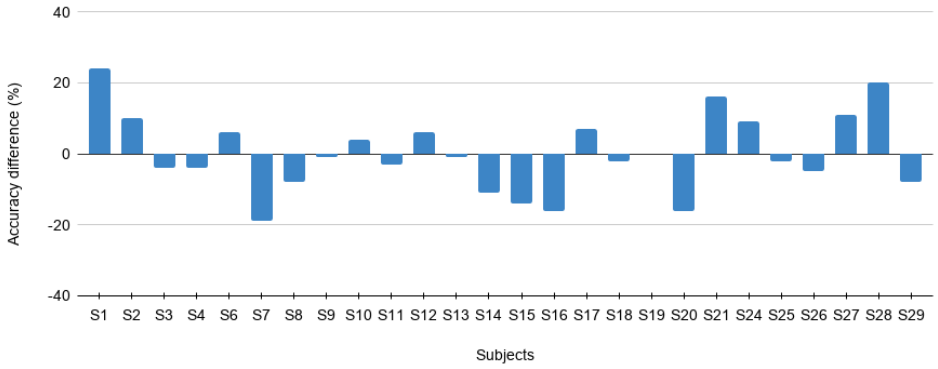


Figure 5.2: Classification accuracy difference after outlier detection using Isolation forest method for different subjects

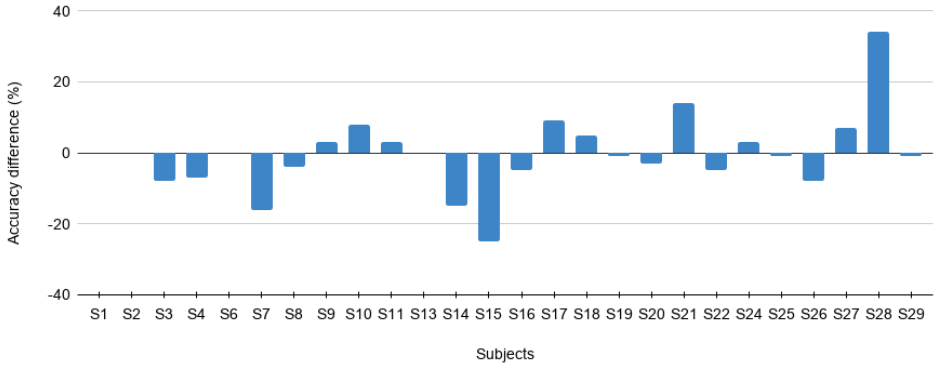


Figure 5.3: Classification accuracy difference after outlier detection using one-class SVM method for different subjects

Tables 5.1 and 5.2 show the accuracy of imaginary movements classification before and after outlier detection using isolation forest and one-class SVM methods. Subjects 5, 22, 23 were deleted due to poor data quality by isolation forest and 5, 12, 23 were deleted by one-class SVM.

As we can see from the represented data the average accuracy across all subjects remained at the same level for both methods. However, accuracy increased for 10 subjects (+11.3% in average) using isolation forest method (Figure 5.2) and for 9 subjects (+9.6% in average) using one-class SVM method (Figure 5.3). This suggests that removing outliers from the training data helps to build more accurate classification models. But this approach is subject-specific and can be successfully applied to a specific group of subjects (see section 6.1). At the same time, the method of outlier detection in the initial data has to be improved.

Table 5.1: Classification accuracy before and after outlier detection using Isolation forest method

Subject	Accuracy before outlier detection	Accuracy after outlier detection	The best parameter of data saturation by outliers (%)
S1	0.73 ± 0.06	0.97 ± 0.05	40.0 ± 22.86
S2	0.75 ± 0.0	0.85 ± 0.11	48.0 ± 5.66
S3	0.5 ± 0.04	0.46 ± 0.1	28.0 ± 14.24
S4	0.58 ± 0.02	0.54 ± 0.1	33.33 ± 20.74
S6	0.48 ± 0.08	0.54 ± 0.14	34.67 ± 16.11
S7	0.52 ± 0.05	0.33 ± 0.27	50.67 ± 6.8
S8	0.6 ± 0.08	0.52 ± 0.1	42.67 ± 14.73
S9	0.67 ± 0.09	0.66 ± 0.01	24.0 ± 17.28
S10	0.62 ± 0.09	0.66 ± 0.09	32.0 ± 9.8
S11	0.58 ± 0.06	0.55 ± 0.17	45.33 ± 10.5
S12	0.47 ± 0.05	0.53 ± 0.1	40.0 ± 14.97
S13	0.55 ± 0.11	0.54 ± 0.14	41.33 ± 23.63
S14	0.48 ± 0.02	0.37 ± 0.09	38.67 ± 27.39
S15	0.7 ± 0.11	0.56 ± 0.09	44.0 ± 5.66
S16	0.75 ± 0.07	0.59 ± 0.07	26.67 ± 18.57
S17	0.47 ± 0.05	0.54 ± 0.05	14.67 ± 15.08
S18	0.52 ± 0.09	0.5 ± 0.13	32.0 ± 11.78
S19	0.78 ± 0.08	0.78 ± 0.05	17.33 ± 21.75
S20	0.6 ± 0.04	0.44 ± 0.14	25.33 ± 19.14
S21	0.42 ± 0.06	0.58 ± 0.12	26.67 ± 16.44
S24	0.47 ± 0.05	0.56 ± 0.04	34.67 ± 6.8
S25	0.93 ± 0.09	0.91 ± 0.08	5.33 ± 7.54
S26	0.58 ± 0.06	0.53 ± 0.1	45.33 ± 8.22
S27	0.75 ± 0.04	0.86 ± 0.1	20.0 ± 11.78
S28	0.38 ± 0.05	0.58 ± 0.11	42.67 ± 12.36
S29	0.72 ± 0.02	0.64 ± 0.06	38.67 ± 27.39
AVG	0.6 ± 0.06	0.6 ± 0.1	33.54 ± 14.89

Table 5.2: Classification accuracy before and after outlier detection using One-class SVM method

Subject	Accuracy before outlier detection	Accuracy after outlier detection	The best parameter of data saturation by outliers (%)
S1	0.73 ± 0.06	0.73 ± 0.12	37.33 ± 3.77
S2	0.75 ± 0.0	0.75 ± 0.04	12.0 ± 11.31
S3	0.5 ± 0.04	0.42 ± 0.07	18.67 ± 13.2
S4	0.58 ± 0.02	0.51 ± 0.04	24.0 ± 16.97
S6	0.48 ± 0.08	0.48 ± 0.11	29.33 ± 21.25
S7	0.52 ± 0.05	0.36 ± 0.25	32.0 ± 6.53
S8	0.6 ± 0.08	0.56 ± 0.21	40.0 ± 17.28
S9	0.67 ± 0.09	0.7 ± 0.03	12.0 ± 6.53
S10	0.62 ± 0.09	0.7 ± 0.12	28.0 ± 9.8
S11	0.58 ± 0.06	0.61 ± 0.1	42.67 ± 12.36
S13	0.55 ± 0.11	0.55 ± 0.05	8.0 ± 5.66
S14	0.48 ± 0.02	0.33 ± 0.09	42.67 ± 24.51
S15	0.7 ± 0.11	0.45 ± 0.32	44.0 ± 16.97
S16	0.75 ± 0.07	0.7 ± 0.02	22.67 ± 15.43
S17	0.47 ± 0.05	0.56 ± 0.09	50.67 ± 7.54
S18	0.52 ± 0.09	0.57 ± 0.14	24.0 ± 8.64
S19	0.78 ± 0.08	0.77 ± 0.09	12.0 ± 8.64
S20	0.6 ± 0.04	0.57 ± 0.27	40.0 ± 22.63
S21	0.42 ± 0.06	0.56 ± 0.11	21.33 ± 7.54
S22	0.5 ± 0.07	0.45 ± 0.15	22.67 ± 20.74
S24	0.47 ± 0.05	0.5 ± 0.12	32.0 ± 14.97
S25	0.93 ± 0.09	0.92 ± 0.08	20.0 ± 25.51
S26	0.58 ± 0.06	0.5 ± 0.14	25.33 ± 16.44
S27	0.75 ± 0.04	0.82 ± 0.15	21.33 ± 1.89
S28	0.38 ± 0.05	0.72 ± 0.21	30.67 ± 17.99
S29	0.72 ± 0.02	0.71 ± 0.05	32.0 ± 22.63
AVG	0.6 ± 0.06	0.6 ± 0.12	27.9 ± 13.72

5.2 Improving combined movement motor imagery classification using machine learning

In this section, data from the database described in section 4.2 is used. Data consists of 7 different MI tasks (right hand; left hand; feet; both hands together; left hand in combination with feet; the right hand in combination with feet; both hands together with feet) and resting-state recorded from 10 right-handed healthy volunteer subjects.

Each subject performed 4 sessions of motor imageries contains 10 trials for each class. Thus, 40 trials per class are available for each subject.

Figure 5.4 shows the time course of ERD/ERS over electrode C3 (i.e, the activity source associated with the use of the right hand). We can observe that all motor imageries involving the use of the right hand (i.e., right hand, right hand in combination with feet, both hands, and both hands in combination with feet) present considerably lower ERD/ERS values than those that do not include it (i.e., rest, left hand, feet, and left hand in combination with feet).

Similarly, Figure 5.5 presents the time course of ERD/ERS over electrode C_z (i.e, the activity source associated with the use of the feet). Over this region, it can also be observed that the motor imageries involving the use of the feet (i.e., feet, left hand in combination with feet, right hand in combination with feet, and both hands in combination with feet) present lower ERD/ERS values than those that do no it include it (i.e., rest, left hand, right hand, and both hands). However, in comparison to the contrast observed over electrode C3 between the tasks involving the use of the right hand and those that do not involve it, differences are not as visible. In fact, we can observe that the use of both left and right hands induce some desynchronization over the central region, possibly

because the associated sources are quite close to electrode C_z, which might cause it to incorporate some of this activity.

Finally, the same comparison is illustrated in Figure 5.6 for the time course of ERD/ERS generated over electrode C4 (i.e, the activity source linked to the use of the left hand). In this case, similarly to the analysis over electrode C3, it is possible to observe a significant contrast between the low ERD/ERS values obtained from the activity induced by the motor imageries involving the use of the left hand (i.e., left hand, left hand in combination with feet, both hands, and both hands in combination with feet) from the high ERD/ERS values associated to the tasks that do not involve the use of the left hand (i.e., rest, feet, right hand, and right hand in combination with feet).

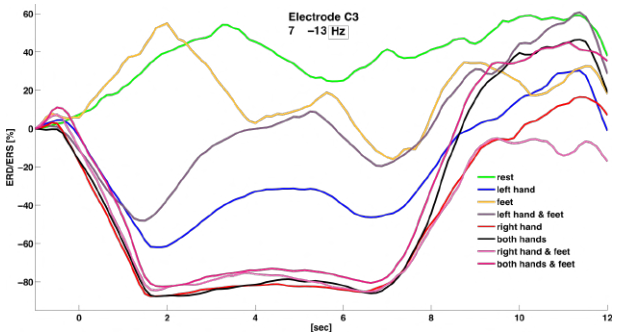


Figure 5.4: ERD/ERS% time series within the mu range for all motor tasks over electrode C3 (i.e, the activity source associated with the use of the right hand). ERD/ERS values were estimated over the EEG trials from subject 2 (from [150]).

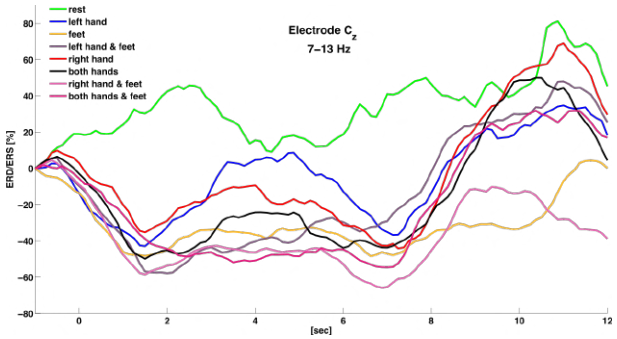


Figure 5.5: ERD/ERS% time series within the mu range for all motor tasks over electrode C_z. ERD/ERS values were estimated over the EEG trials from subject 2 (from [150]).

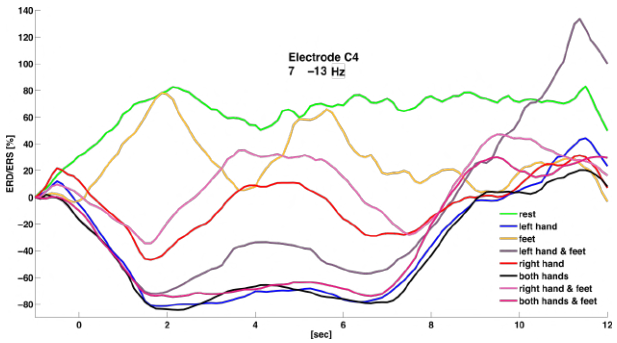


Figure 5.6: ERD/ERS% time series within the mu range for all motor tasks over electrode C4. ERD/ERS values were estimated over the EEG trials from subject 2 (from [150]).

5.2.1 Combined movement MI classification using classic methods

Table 5.3 reports classification accuracy for 8 classes obtained by a standard shrinkage LDA for 10 subjects. These results were obtained using all available data from 8 classes for train according to One-Session-Out cross-validation approach. From each trial, 26 ERD/ERS features (one per channel) have been extracted from the filtered (8-30 Hz) EEG signal.

The mean accuracy of 32.7% is quite low. However, for two subjects (S2 and S10) there is higher accuracy compared with other subjects. It is interesting to see that the best subject in term of accuracy had also the powerful ERD during the MI-task (Figure 5.4, Figure 5.5 and Figure 5.6). All of the above once again confirms the fact that this is a subject-specific task. In the next section, we will discuss possible solutions to improve the classification accuracy.

Table 5.3: Accuracy obtained by a shrinkage LDA for 10 subjects for 8-class MI classification task.

Subject	**Accuracy & STD**
1	0.219 ± 0.34
2	0.581 ± 0.50
3	0.347 ± 0.27
4	0.284 ± 0.66
5	0.269 ± 0.40
6	0.237 ± 0.18
7	0.278 ± 0.66
8	0.266 ± 0.35
9	0.353 ± 0.30
10	0.434 ± 0.52
AVG	$\mathbf{0.327 \pm 0.42}$

Table 5.4 shows the main misclassifications. In particular, feet motor imageries have been confused with a resting state (Rest → Feet, Feet → Rest, Both hands → Feet, Left hand & Feet → Left hand, Right hand & Feet → Right hand, Both hands & Feet → Both hand). One explanation could be that the activity of the right hand (respectively the left hand and both hands) goes over the Cz location.

Table 5.4: Confusion Matrix (%) over all subjects

		Predicted labels							
		Rest	LH	RH	BH	F	LH & F	RH & F	BH & F
	Rest	**44**	5	7	5	18	9	6	6
	LH	6	**22**	11	15	7	16	9	14
	RH	6	10	**25**	17	9	6	16	11
	BH	5	11	12	**28**	7	10	8	19
	F	17	6	8	5	**42**	8	9	5
	LH & F	6	13	6	9	11	**35**	9	11
	RH & F	6	8	14	9	7	9	**36**	11
True labels	BH & F	6	13	9	20	6	12	11	**23**

The accuracy obtained by a shrinkage linear discriminant analysis is not good enough to allow devises control with combined movements. Further work will be focused on building the ensemble of binary classifiers (section 5.2.2) to save time on BCI training and feature selection for better condition separation (section 5.2.3). Then one-versus-one approach, one-versus-the-rest approach, or a hierarchical approach should improve the detection of combined motor imageries.

5.2.2 Classification of combined movement motor imageries using binary classifiers

The aim of this experiment was to investigate the performance of the classification using 3 parallel connected binary classifiers instead of using a multi-label classifier. The idea was to train 3 binary classifiers on data with

single motor imageries and then test it on the set which contains all possible combinations of these movements and rest [288]. It is useful because it can help to save a lot of time on recording the EEG while collecting data for MI classification. Instead of recording all combinations of combined movements, it can permit to collect of the sets only with single movements.

In this section different methods of supervised learning were used to estimate the performance of our approach with 3 binary classifiers in the task of the multi-label combined movement classification.

The following algorithms of supervised learning were selected: SVM with different kernels (section 2.5.2), LDA with shrinkage regularization (section 2.5.1) and Decision Tree Classifier (section 2.5.3).

Investigation was conducted with six sets of features: 3-feature set (ERD) (table 5.5), 3-feature set (ERS) (table 5.6), 6-feature set (ERD & ERS on alpha+beta band) (table 5.7), 6-feature set (ERD on alpha and ERS on beta band) (table 5.8), 18-feature set (table 5.9) and 30-feature set (table 5.10) (described in section 4.2.7).

Average accuracies and standard deviation were calculated for all subjects with leave-one-session-out cross-validation (see section 4.2.6.2). SVM with sigmoid and polynomial kernels are not presented due to low-level performance.

Table 5.5: Accuracies of combined MI movements classification on 3-ERD-feature set

State	Classifiers			
	SVM (linear)	SVM (RBF)	LDA	Decision Tree
Average	**0.264 ± 0.08**	**0.255 ± 0.08**	**0.261 ± 0.09**	**0.194 ± 0.08**
Rest	0.152 ± 0.11	0.163 ± 0.15	0.16 ± 0.12	0.198 ± 0.11
Left hand	0.348 ± 0.2	0.287 ± 0.2	0.342 ± 0.19	0.258 ± 0.15
Right hand	0.367 ± 0.29	0.35 ± 0.3	0.37 ± 0.28	0.297 ± 0.25
Both hands	0.282 ± 0.24	0.282 ± 0.26	0.29 ± 0.24	0.16 ± 0.08
Feet	0.427 ± 0.25	0.443 ± 0.23	0.412 ± 0.24	0.24 ± 0.17
Left hand and feet	0.265 ± 0.24	0.24 ± 0.21	0.255 ± 0.24	0.165 ± 0.12
Right hand and feet	0.185 ± 0.13	0.22 ± 0.13	0.173 ± 0.14	0.133 ± 0.08
Both hands and feet	0.085 ± 0.1	0.053 ± 0.05	0.085 ± 0.09	0.1 ± 0.06

Table 5.6: Accuracies of combined MI movements classification on 3-ERS-feature set

State	Classifiers			
	SVM (linear)	SVM (RBF)	LDA	Decision Tree
Average	**0.168 ± 0.03**	**0.168 ± 0.03**	**0.166 ± 0.03**	**0.146 ± 0.03**
Rest	0.043 ± 0.06	0.045 ± 0.06	0.048 ± 0.05	0.138 ± 0.06
Left hand	0.338 ± 0.17	0.282 ± 0.15	0.335 ± 0.16	0.17 ± 0.05
Right hand	0.233 ± 0.13	0.223 ± 0.12	0.248 ± 0.12	0.16 ± 0.09
Both hands	0.115 ± 0.08	0.118 ± 0.08	0.11 ± 0.08	0.133 ± 0.04
Feet	0.283 ± 0.16	0.31 ± 0.17	0.263 ± 0.13	0.17 ± 0.07
Left hand and feet	0.11 ± 0.07	0.098 ± 0.09	0.09 ± 0.07	0.128 ± 0.07
Right hand and feet	0.22 ± 0.13	0.26 ± 0.16	0.227 ± 0.15	0.155 ± 0.07
Both hands and feet	0.005 ± 0.01	0.008 ± 0.01	0.008 ± 0.01	0.113 ± 0.06

Table 5.7: Accuracies of combined MI movements classification on 6 ERD & ERS-feature set

State	Classifiers			
	SVM (linear)	SVM (RBF)	LDA	Decision Tree
Average	**0.285 ± 0.09**	**0.277 ± 0.09**	**0.284 ± 0.09**	**0.203 ± 0.07**
Rest	0.205 ± 0.09	0.195 ± 0.14	0.23 ± 0.1	0.232 ± 0.13
Left hand	0.467 ± 0.22	0.43 ± 0.22	0.44 ± 0.19	0.288 ± 0.15
Right hand	0.455 ± 0.28	0.46 ± 0.28	0.46 ± 0.27	0.34 ± 0.26
Both hands	0.265 ± 0.19	0.183 ± 0.14	0.268 ± 0.19	0.133 ± 0.05
Feet	0.455 ± 0.19	0.487 ± 0.22	0.463 ± 0.19	0.318 ± 0.14
Left hand and feet	0.17 ± 0.16	0.188 ± 0.17	0.18 ± 0.15	0.143 ± 0.07
Right hand and feet	0.227 ± 0.12	0.247 ± 0.16	0.202 ± 0.12	0.113 ± 0.06
Both hands and feet	0.033 ± 0.04	0.028 ± 0.04	0.03 ± 0.04	0.063 ± 0.04

Table 5.8: Accuracies of combined MI movements classification on 6-feature set: alpha band (8-12 Hz) and beta band (17-25 Hz)

State	**Classifiers**			
	SVM (linear)	**SVM (RBF)**	**LDA**	**Decision Tree**
Average	**0.275 ± 0.09**	**0.272 ± 0.09**	**0.281 ± 0.1**	**0.222 ± 0.08**
Rest	0.205 ± 0.12	0.225 ± 0.14	0.228 ± 0.14	0.35 ± 0.14
Left hand	0.467 ± 0.2	0.477 ± 0.2	0.458 ± 0.23	0.338 ± 0.18
Right hand	0.412 ± 0.27	0.425 ± 0.28	0.423 ± 0.25	0.372 ± 0.25
Both hands	0.25 ± 0.12	0.205 ± 0.14	0.268 ± 0.15	0.12 ± 0.08
Feet	0.497 ± 0.14	0.507 ± 0.18	0.488 ± 0.14	0.36 ± 0.12
Left hand and feet	0.168 ± 0.1	0.15 ± 0.09	0.2 ± 0.12	0.103 ± 0.06
Right hand and feet	0.165 ± 0.09	0.168 ± 0.11	0.158 ± 0.1	0.11 ± 0.05
Both hands and feet	0.035 ± 0.05	0.02 ± 0.02	0.03 ± 0.04	0.025 ± 0.03

Table 5.9: Accuracies of combined MI movements classification on 18-feature set

State	Classifiers			
	SVM (linear)	SVM (RBF)	LDA	Decision Tree
Average	**0.286 ± 0.08**	**0.294 ± 0.09**	**0.288 ± 0.09**	**0.205 ± 0.06**
Rest	0.253 ± 0.12	0.255 ± 0.13	0.253 ± 0.11	0.487 ± 0.13
Left hand	0.492 ± 0.12	0.487 ± 0.16	0.505 ± 0.13	0.27 ± 0.1
Right hand	0.45 ± 0.21	0.435 ± 0.31	0.463 ± 0.23	0.263 ± 0.14
Both hands	0.232 ± 0.12	0.253 ± 0.15	0.247 ± 0.14	0.103 ± 0.07
Feet	0.45 ± 0.18	0.507 ± 0.18	0.443 ± 0.19	0.328 ± 0.12
Left hand and feet	0.19 ± 0.07	0.188 ± 0.11	0.192 ± 0.07	0.098 ± 0.07
Right hand and feet	0.185 ± 0.08	0.19 ± 0.12	0.172 ± 0.09	0.073 ± 0.05
Both hands and feet	0.033 ± 0.04	0.035 ± 0.06	0.028 ± 0.03	0.017 ± 0.02

Table 5.10: Accuracies of combined MI movements classification on 30-feature set

State	Classifiers			
	SVM (linear)	SVM (RBF)	LDA	Decision Tree
Average	**0.303 ± 0.11**	**0.305 ± 0.1**	**0.321 ± 0.09**	**0.213 ± 0.07**
Rest	0.358 ± 0.14	0.375 ± 0.15	0.357 ± 0.14	0.422 ± 0.09
Left hand	0.535 ± 0.21	0.522 ± 0.24	0.533 ± 0.22	0.315 ± 0.16
Right hand	0.583 ± 0.23	0.555 ± 0.25	0.595 ± 0.23	0.398 ± 0.25
Both hands	0.22 ± 0.16	0.21 ± 0.14	0.253 ± 0.19	0.072 ± 0.04
Feet	0.497 ± 0.19	0.588 ± 0.21	0.508 ± 0.17	0.325 ± 0.12
Left hand and feet	0.193 ± 0.11	0.158 ± 0.07	0.202 ± 0.1	0.105 ± 0.07
Right hand and feet	0.187 ± 0.1	0.175 ± 0.1	0.19 ± 0.1	0.073 ± 0.06
Both hands and feet	0.015 ± 0.02	0.025 ± 0.05	0.015 ± 0.02	0.008 ± 0.02

Table 5.11: Average accuracies of combined MI movements classification for different feature sets

Feature set	Classifiers			
	SVM (linear)	SVM (RBF)	LDA	Decision Tree
3-ERD (7-30)	0.264 ± 0.08	0.255 ± 0.08	0.261 ± 0.09	0.194 ± 0.08
3-ERS (7-30)	0.168 ± 0.03	0.168 ± 0.03	0.166 ± 0.03	0.146 ± 0.03
6: ERD (7-30) & ERS (7-30)	0.285 ± 0.09	0.277 ± 0.09	0.284 ± 0.09	0.203 ± 0.07
6: ERD (8-12) & ERS (17-25)	0.275 ± 0.09	0.272 ± 0.09	0.281 ± 0.1	0.222 ± 0.08
18-feature set	0.286 ± 0.08	0.294 ± 0.09	0.288 ± 0.09	0.205 ± 0.06
30-feature set	0.303 ± 0.11	0.305 ± 0.1	0.321 ± 0.09	0.213 ± 0.07

As we can see from the tables 5.5 - 5.10 the best performance is demonstrated by linear classifiers (Linear SVM and LDA). Average accuracies of combined MI movements classification for different feature sets are represented in Table 5.11 to simplify comparison. But this level of accuracy is lower than the standard multi-label approach. An important moment is that level of classification of the "rest" class is very low, although this class was known to the classifier. It can suggest that the "rest" class was confused with the "feet" class, as was presented in Table 5.4. Also, the accuracy of the "both hands" classification is higher than the accuracy of other unknown classes. But in general, we have a good result on known classes and bad performance on the unknown classes (accuracy of the "both hands and feet" classification even equals to 0.8-2.5% in the 30-feature set).

It is also showed that with an increasing number of different bands the accuracy also increases (26.1% for LDA classifiers on 3-ERD feature set (see Table 5.5) increased to 32.1% for the 30-feature set (see Table 5.10)). But it is doubtful that all features are equally useful for the classification. It is also known that with a small number of data examples, a large number of features may lead to model overfitting, which results in low accuracy on the validation data. That's why using feature selection techniques (see section 2.3) could be promising to improve model performance. Results obtained using feature selection techniques are represented in section 5.2.3.

5.2.3 Feature selection for combined movement MI classification

Observing modulation in a wide frequency band such as the alpha+beta bands together (8-30 Hz) ensures having desynchronization and synchronization. In this section, we investigated features to know the variability between bands. It can help to understand which bands and features are more useful.

For this task meta-transformer, SelectFromModel was used. This algorithm was realized in the Python library Scikit-learn. More information about this algorithm and its parameters see in section 2.3.

We test this algorithm on 30-feature set (see section 4.2.7) with linear SVM ($C = 0.01$, $L1$ penalty, $threshold = 10^{-5}$) for each of 3 binary classifiers for all subjects. Results shows how frequently features were marked as important by classifier (maximum mark is 40 - it equals 10 subjects with 4-fold cross-validation). Results are represented in the tables 5.12-5.14.

Table 5.12: Result of feature selection for 'Right vs. rest & Left & Feet' classifier. Frequently used features are indicated in green, occasionally used features indicated in yellow, features that are rarely used are indicated in red.

Feature	Channels		
	C3	Cz	C4
ERD 8-10	2	2	3
ERD 11-13	30	4	7
ERD 8-12	22	6	14
ERD 8-30	25	12	18
ERS 15-17	2	0	2
ERS 18-20	9	0	2
ERS 21-23	4	1	3
ERS 24-26	0	2	0
ERS 17-25	16	4	10
ERS 8-30	24	8	16

Table 5.13: Result of feature selection for 'Left vs. rest & Right & Feet' classifier. Frequently used features are indicated in green, occasionally used features indicated in yellow, features that are rarely used are indicated in red.

Feature	Channel		
	C3	Cz	C4
ERD 8-10	11	7	9
ERD 11-13	16	3	32
ERD 8-12	14	10	20
ERD 8-30	16	11	21
ERS 15-17	4	0	6
ERS 18-20	3	0	7
ERS 21-23	9	0	0
ERS 24-26	2	3	0
ERS 17-25	12	6	18
ERS 8-30	12	6	22

Table 5.14: Result of feature selection for 'Feet vs. rest & Right & Left' classifier. Frequently used features are indicated in green, occasionally used features indicated in yellow, features that are rarely used are indicated in red.

Feature	Channels		
	C3	Cz	C4
ERD 8-10	10	9	4
ERD 11-13	14	14	12
ERD 8-12	19	11	21
ERD 8-30	12	16	12
ERS 15-17	3	0	1
ERS 18-20	5	2	3
ERS 21-23	6	9	1
ERS 24-26	2	9	0
ERS 17-25	5	24	5
ERS 8-30	18	10	18

From the results given in the Tables 5.12 and 5.13 we can see that for the separation of left and right hand activity from other activities alpha ERD and wide beta ERS are useful, and other ERS features can be removed. As for the feet activity detection we have to use ERS beta (17-25 Hz, 7-30 Hz) features on Cz and ERD alpha band feature (8-12 Hz) on C3 and C4 to separate feet activity from others.

To see variability between these features histograms of distribution logarithmic ERD/ERS depending on the state (class) also were created. Histograms are represented in Figures 5.15, 5.16.

Figures 5.15, 5.16 shows that separation is not clear even for the features, which was chosen by feature selector. The best separation is in the low-alpha ERD. But even here overlapping between "feet" and "rest" classes exists. The same could be seen on the 3D plot of the sample distribution (Figure 5.7a).

Table 5.15: Histograms of sample distribution for ERD features (calculated from 0.5 sec till 6 sec)

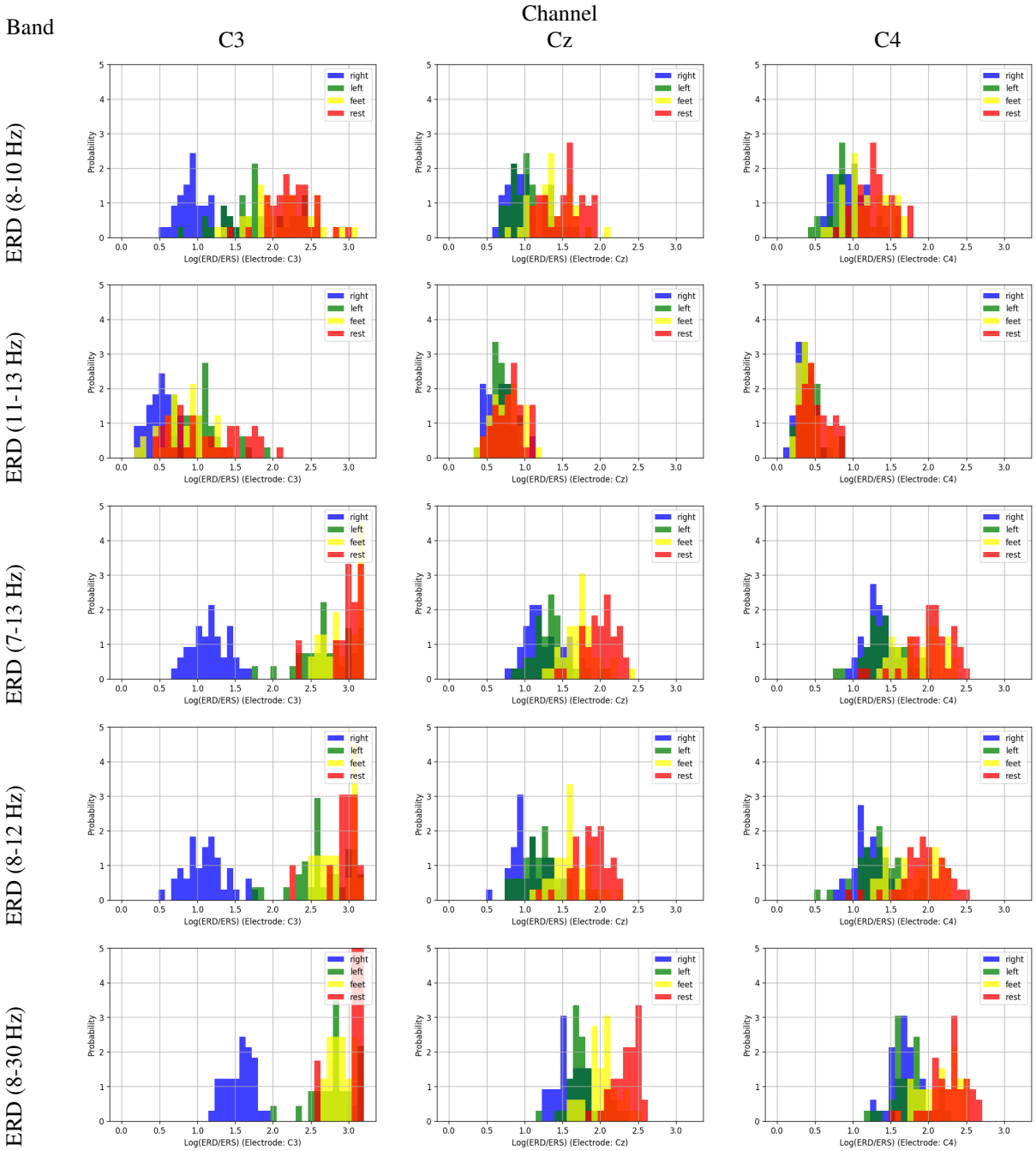
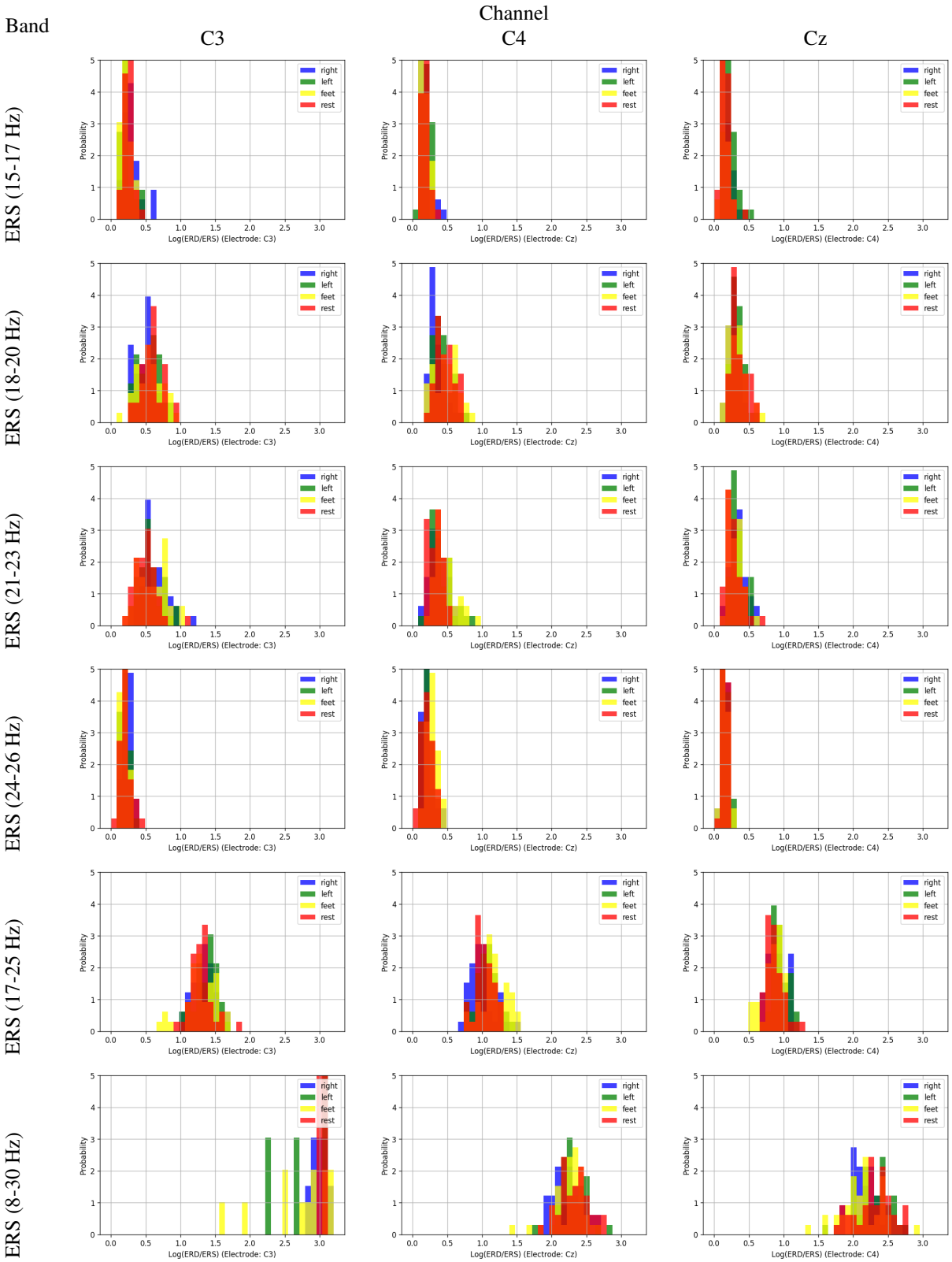
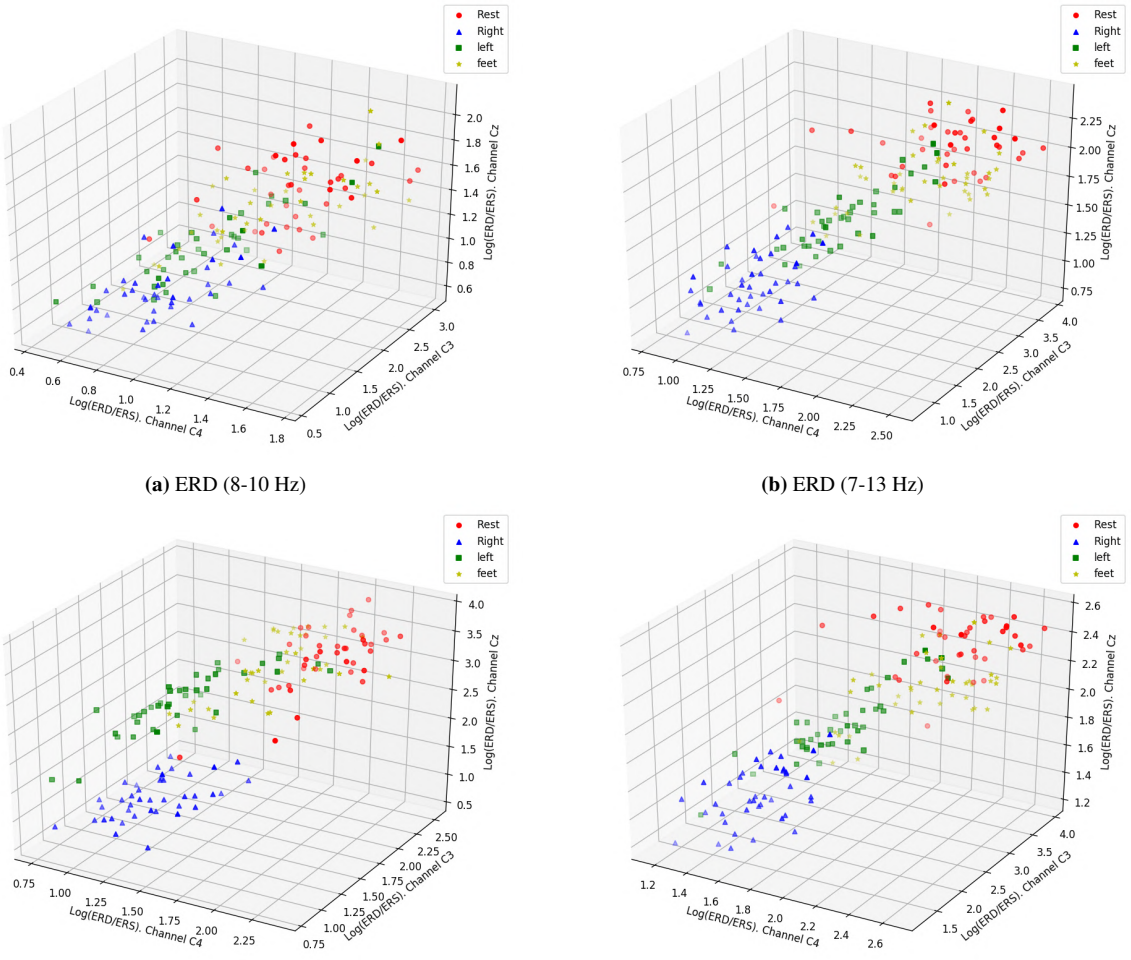


Table 5.16: Histograms of sample distribution for ERS features (calculated from 7 sec till 12 sec)





(a) ERD (8-10 Hz)

(b) ERD (7-13 Hz)

(c) ERD (8-12 Hz)

(d) ERD (8-30 Hz)

Figure 5.7: 3D plot of the sample distribution for different feature sets: (a) ERD (8-10 Hz); (b) ERD (7-13 Hz); (c) ERD (8-12 Hz); (d) ERD (8-30 Hz).

Interesting thing that ERD feature on 8-12 Hz band (Figure 5.7c) have much better separation of classes than both ERD on 8-10 (Figure 5.7a) and 11-13 Hz and even than 7-13 Hz (Figure 5.7b). But the separation of 'rest' and 'feet' classes is still very bad. It also could be seen on histograms (table 5.15). Another useful feature could be ERD/ERS in the alpha+beta band (8-30 Hz). As we can see, using features from this band separation of all classes is not perfect, but the separation of the 'rest' and 'feet' classes are much better than in other features.

In this investigation, we also computed accuracies after feature selection for the 30-feature set (Table 5.17).

Table 5.17: Accuracies of combined movement classification on 30-feature set using binary classifiers and feature selection

State	Classifiers		
	SVM (linear)	sLDA	Decision Tree
Average before feature selection	0.303 ± 0.11	0.321 ± 0.09	0.213 ± 0.07
Average after feature selection	0.323 ± 0.11	0.332 ± 0.10	0.215 ± 0.07

We can see that after feature selection (Table 5.17) classification accuracy is higher. But at the same time, we have not a significant difference between results.

Furthermore, the best accuracy obtained using binary classifies with feature selection ($33.2\% \pm 1\%$) is even higher than for standard multilabel approach ($32.7\% \pm 4.2\%$) (see Table 5.3).

5.3 Optimizing motor imagery detection with deep learning to prevent intraoperative awareness

In this section, we have presented results obtained on the dataset introduced in section 4.3. This database consisted of median nerve stimulations during motor imagery and during rest, as well as MI condition separately. EEG signals were recorded from 22 right-handed healthy volunteers.

The goal of this experiment is to improve and optimize motor intention detection using different deep learning architectures, frequency bands, electrode layouts, and classification techniques. Also, the aim of the experiments with this database is to detect the occurrence of motor patterns under two different conditions: median nerve stimulation during motor imagery (MI + MNS) and median nerve stimulation during rest (MNS). These two conditions aim at showing that a MNS can be used and is more helpful as a trigger to improve motor imagery detection (see section 1.3.4).

The standard machine learning classification methods (such as CSP+LDA, MDRM and TS+LR) were used for MI-based BCI described in chapter 2, as well as CNN architectures described in chapter 3.

We selected three state-of-the-art deep neural network architectures (such as D-ConvNet, S-ConvNet, EEGNet) to compare their results to the best-known approaches for classification in the MI-based BCI domain.

For feature extraction we used two efficient methods: CSP filters [17, 19, 18] (see section 2.2.3.2) and Riemannian geometry [21, 23, 22] (see section 2.2.4).

For classification we used first a linear discriminant analysis (LDA) [289] (see section 2.5.1). We also selected approaches based on Riemannian geometry: MDRM algorithm [21, 22, 23] (see section 2.5.5), tangent space projection of covariance matrices with following logistic regression classifier application on projected data (TS+LR) as described in [20, 21] (see section 2.2.4.3).

Our motivation to use convolutional neural networks is to not predefine features and expect better accuracy. We used i) a deep convolutional network (D-ConvNet) as reference for deep networks (section 3.6.1), ii) a shallow convolutional network (S-ConvNet) because they are designed specifically for oscillatory signal classification (section 3.6.2), and iii) a generic architecture (EEGNet) proposed in [27] for EEG data that already demonstrated good results on several BCI paradigms (section 3.6.3); iv) adaptation of EEGNet, namely EEGNet-4.8, proposed in the section 3.6.4, that better handle mid- and high frequency information.

We present how deep neural nets architectures can improve classification accuracy for two MI discriminating paradigms: the standard (MI vs. rest) BCI task and the same task when a stimulation of the median nerve is added (MI+MNS vs. MNS). We addressed the following aspects in the experiments:

- using extended frequency ranges with theta and low gamma bands (section 5.3.1);
- using various deep learning architectures and standard classifiers (section 5.3.4);
- EEGNet adaptation to better handle medium frequency components (section 5.3.2);
- dropout parameter optimization (section 5.3.3);
- comparing the two discriminating tasks: MNS vs. MNS+MI and MI vs. rest (section 5.3.5);
- frontal electrodes influence investigation (section 5.3.6);
- comparing different EEG montages with reduced electrodes number.

5.3.1 Frequency band selection

In the first set of experiments, we investigate the influence of the EEG frequency band on the classification accuracy. The recommended frequency band for detecting motor imagery is 8-30 Hz [290, 291], but sometimes it could be useful to extend the analysis to wider frequency bands such as 4-38 Hz [205, 131, 256, 246] (see Table 3.1). We investigated separately and together the impact of adding the theta 4-8 Hz and the low gamma 30-38 Hz frequency ranges on the classification performance. We did this study for all models, i.e. CSP+LDA, MDRM, TS+LR, EEGNet, D-ConvNet, and S-ConvNet. Similar behavior was obtained for all models, so we only present the most representative part of it (see section 5.3.4) in Figure 5.11 showing classification accuracies for TS+LR and EEGNet-4.8 models trained on data filtered with different bands for MNS vs. MNS+MI and MI vs. rest tasks.

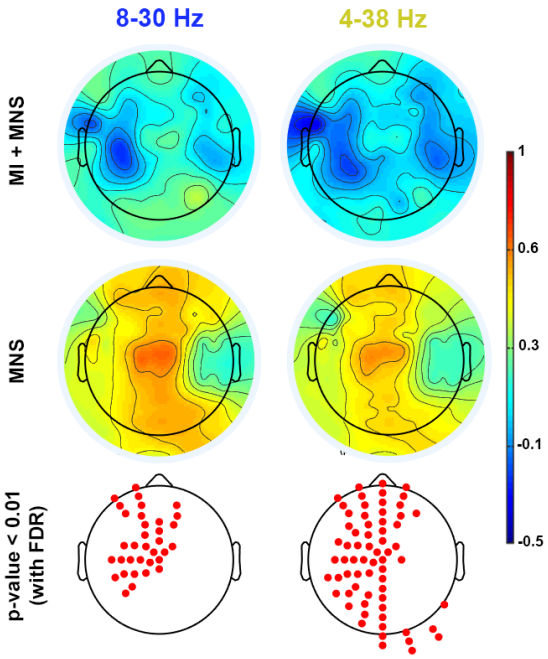


Figure 5.8: Topographic map of ERD/ERS% (grand average, n = 22) in the 8-30 Hz and 4-38 Hz bands during two conditions: MI + MNS and MNS only. A red color corresponds to a strong ERS and a blue one to a strong ERD. Red electrodes indicate a significant difference between the two conditions ($p < 0.01$).

5.3.1.1 MNS+MI vs. MNS

Figure 5.11c shows that including separately and both together the theta and the low gamma bands significantly ($p-value < 0.01$) improves the accuracy. Indeed, discrimination power using 4-30 Hz and 8-38 Hz bands exceeds the one of the 8-30 Hz band. Moreover, both together they improve even more accuracy. The positive effect on the accuracy with the low gamma band is stronger ($p - value < 0.001$) for 128 electrodes. That may indicate the presence of useful information in theta and low gamma oscillations outside the motor cortex and the frontal lobe since for the other montages with 47 electrodes or less recorded brain information is reduced to these regions.

To analyze the differences between MI + MNS and MNS conditions, we performed an event-related spectral perturbation (ERSP) analysis for two different frequency bands (8-30 Hz vs. 4-38 Hz) and separately (4-8 Hz and 30-38 Hz) using EEGLAB. This analysis allows us to understand how MI + MNS and MNS conditions can be discriminated and which time parameters we can choose to guide the classification. We have decided to compute ERSPs in a merged band (mu+beta, 8–30 Hz) and a wider band (4-38 Hz) for MI + MNS and MNS conditions (Figure 5.8). A surrogate permutation test ($p < 0.01$; 4,000 permutations) from the EEGLAB toolbox was used to validate differences in terms of localization of these ERSPs. In addition to this analysis, we applied a false discovery rate (FDR) correction test (see section 2.8) in order to clarify how the FDR was controlled for multiple comparisons to show the difference between MI + MNS and MNS conditions.

Figures 5.9 and 5.10 shows grand average topographic maps of ERD/ERS% for all subjects in the 4-8 Hz and 30-38 Hz bands during two conditions: MI + MNS and MNS only. The red color corresponds to a strong ERS and a blue one to a strong ERD. Red electrodes indicate a significant difference between the two conditions ($p < 0.01$).

Figure 5.8 shows difference between two bands (8-30 Hz and 4-38 Hz) for MNS vs. MI+MNS task. It could be seen that (4-38 Hz) band better handle ERD during MI+MNS condition. Moreover, the FDR test shows a larger number of electrodes that have a significant difference for two conditions for (4-38 Hz) frequency band.

As we can see from Figure 5.9 (4-8 Hz) frequency band could be very useful in case of discriminating MNS condition from the MI+MNS. In addition, FDR shows a huge number of electrodes all over the cortex which indicates a significant difference between the two conditions. At the same time Figure 5.10 indicates that (30-38

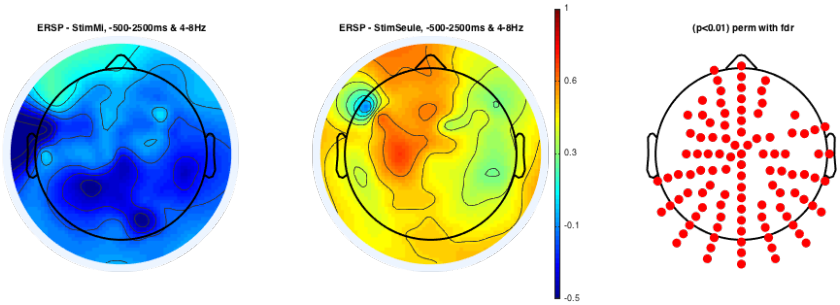


Figure 5.9: Topographic map of ERD/ERS% (grand average, n = 22) in the 4-8 Hz band during two conditions: MI + MNS and MNS only. A red color corresponds to a strong ERS and a blue one to a strong ERD. Red electrodes indicate a significant difference between the two conditions ($p < 0.01$).

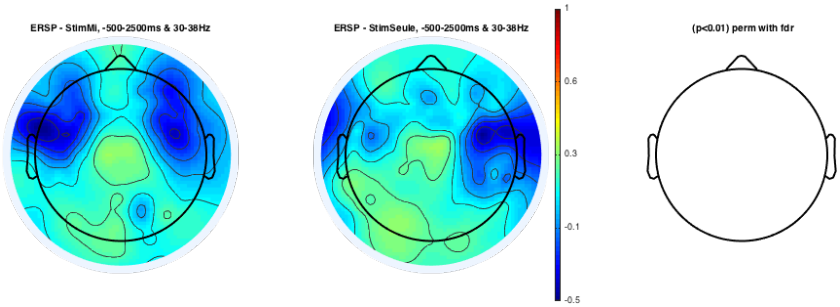


Figure 5.10: Topographic map of ERD/ERS% (grand average, n = 22) in the 30-38 Hz band during two conditions: MI + MNS and MNS only. A red color corresponds to a strong ERS and a blue one to a strong ERD. Red electrodes indicate a significant difference between the two conditions ($p < 0.01$).

Hz) frequency band is not so useful for discriminating MNS vs. MI+MNS task.

Classification accuracy obtained by different classifiers for 22 subjects using different electrode layouts and different frequency bands for MNS vs. MI+MNS discriminating task represented in Tables A.1 - A.48.

Wilcoxon signed-rank test shows significant ($p - value < 0.01$ or higher) difference between EEGNet-4.8 performance on data filtered in 8-30 Hz and 4-38 Hz frequency bands for MNS vs. MI+MNS discriminating task (Figure 5.12). This is the case for 128, 47, 13, 9 and 6-electrode setups. At the same time, for 47, 13, and 6-electrode setups the difference was even more pronounced ($p - value < 0.001$). The advantage of using a 4-38 Hz band can achieve accuracy improvement up to 4% for MNS vs. MI+MNS task. Unlike EEGNet-4.8, the TS+LR model showed a highly significant ($p - value < 0.01$) difference in accuracy obtained on data filtered in 8-30 Hz and 4-38 Hz frequency bands only for 128 electrode setup, although the average growth in accuracy is around 2%.

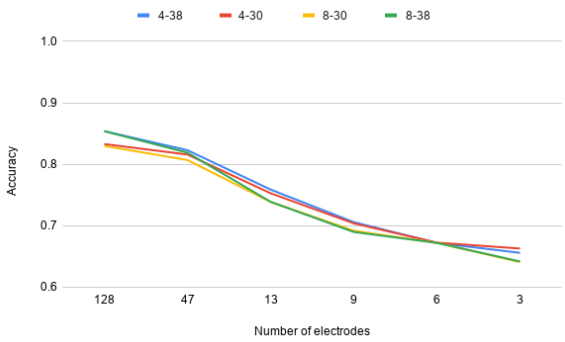
5.3.1.2 Motor imagery vs. rest

Although for the MI vs. rest task the addition of the theta band effects classification accuracy, but no significant difference has been noticed based on Wilcoxon signed-rank test (Figure 5.11d).

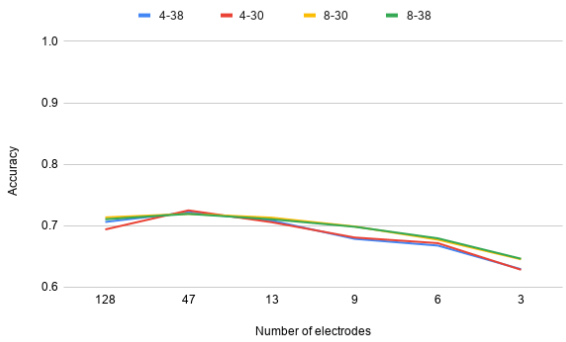
Classification accuracy obtained by different classifiers for 22 subjects using different electrode layouts and different frequency bands for MI vs. rest discriminating task represented in Tables B.1 - B.48.

Recommendations

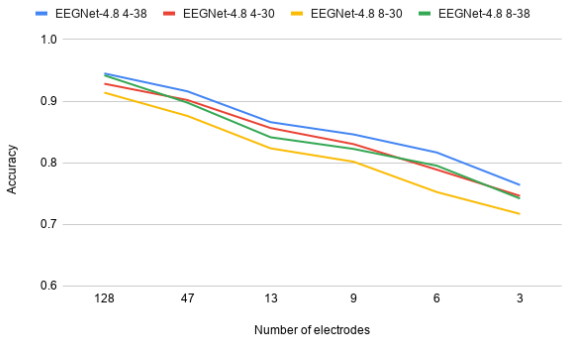
When using MNS it is better to use the wider frequency band 4-38 Hz. At the same time, there are no reasons to change the standard 8-30 Hz band for the Rest vs. MI task.



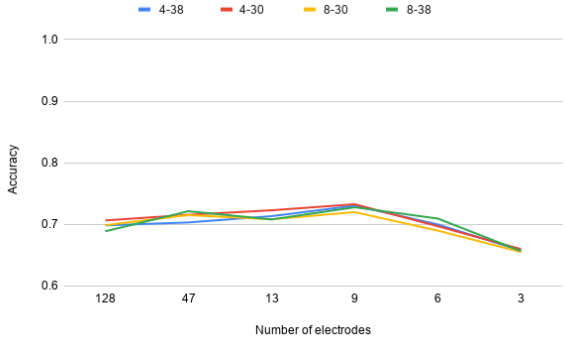
(a) TS+LR for MNS vs. MI+MNS



(b) TS+LR for MI vs. rest task



(c) EEGNet-4.8 for MNS vs. MI+MNS task



(d) EEGNet-4.8 for MI vs. rest task

Figure 5.11: Classification accuracy for TS+LR (5.11a, 5.11b) and EEGNet-4.8 (5.11c, 5.11d), averaged over all folds ($k = 10$) and all subjects ($n = 22$) for different electrodes setups of MNS vs. MI+MNS (5.11a, 5.11c) and MI vs. rest (5.11b, 5.11d) discriminating task. Input data filtered in 4-38 Hz, 4-30 Hz, 8-30 Hz and 8-38 Hz frequency ranges.

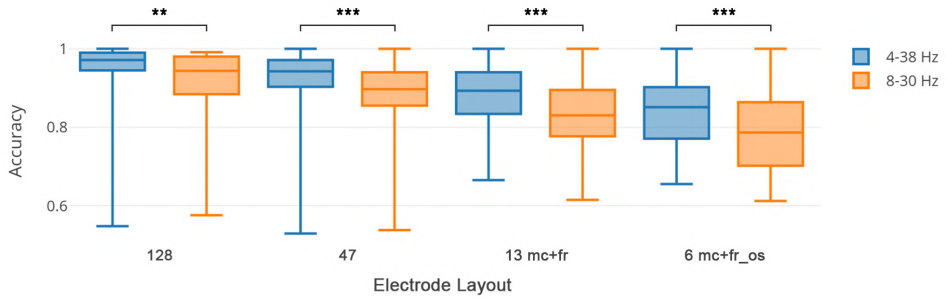


Figure 5.12: Boxplots showing the distribution of EEGNet-4.8 classification accuracy using two frequency bands (4-38 Hz) and standard (8-30 Hz) for different electrode layouts, averaged over all folds ($k = 10$) and all subjects ($n = 22$) for MNS vs. MI+MNS discriminating task. Significance levels of Wilcoxon signed-rank test comparisons: $^{**}p < 0.01$, $^{***}p < 0.001$.

5.3.2 EEGNet architectures

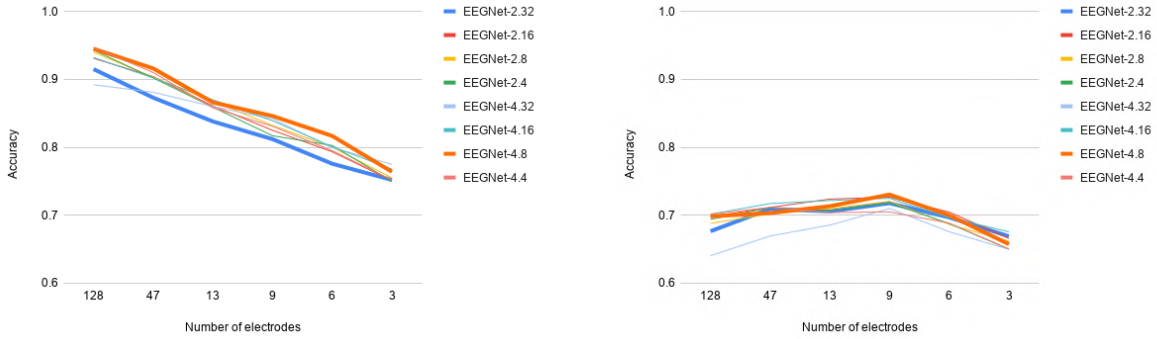
EEGNet is a very promising architecture [27] for spatio-temporal analysis of EEG signals. To better investigate this architecture, many parameters of EEGNet may be adjusted. In the present investigation, we focus on the temporal kernel size K for the first convolutional layer and the multiplier depth D of the depthwise convolution layer which to our best knowledge were not investigated deeply earlier. The accuracy obtained using different architectures of EEGNet for MNS vs. MI+MNS task represented in Figure 5.13a. We can see that decreasing the kernel size of the first convolutional layer and increasing the multiplier depth outperforms the standard set of

parameters of EEGNet-2.32 [27] for a kernel size of 2 and a multiplier depth of 32. We selected EEGNet-4.8 for future investigations based on its higher classification performance for almost all electrode layouts.

5.3.2.1 MNS vs. MI+MNS

More precisely comparing EEGNet-4.8 proposed in this work with standard EEGNet-2.32 architecture for sensory-motor data [27] we can see that it works especially better for larger setups of electrodes for MNS vs. MI+MNS task (Figure 5.13a). Wilcoxon signed-rank test showed a highly significant ($p-value < 0.01$) difference between performances of EEGNet-4.8 and EEGNet-2.32 for all setups of electrodes except setup with 3 electrodes over the motor cortex.

Classification accuracy obtained by different variations of EEGNet for 22 subjects using different electrode layouts for MNS vs. MI+MNS discriminating task represented in Tables C.1 - C.12.



(a) MNS vs. MI+MNS task

(b) MI vs. rest task

Figure 5.13: Classification accuracy for different variants of EEGNet, averaged over all folds ($k = 10$) and all subjects ($n = 22$) for different electrodes setups of MNS vs. MI+MNS (5.13a) and MI vs. rest (5.13b) discriminating task. Input data filtered in 4-38 Hz frequency band for 5.13a and in 8-30 Hz frequency band for 5.13b.

5.3.2.2 Motor imagery vs. rest

Classification accuracy obtained by different variations of EEGNet for 22 subjects using different electrode layouts for MI vs. rest discriminating task represented in Tables D.1 - D.12.

For MI vs. rest task (Figure 5.13b) the difference between EEGNet-2.32 and EEGNet-4.8 is not statistically significant, but the classification accuracy of EEGNet-4.8 is better for most electrode layouts.

5.3.3 Dropout investigation

Also we investigated the influence of dropout (section 3.4.2) on classification performance of EEGNet variants for MNS vs. MNS+MI discriminating task. Results for electrode setup with 128 and 47 electrodes represented in Table 5.18 and 5.19 respectively. It is showed that classification accuracy obtained with dropout rate $D = 0.5$ is higher for all models in contrast to $D = 0.25$ proposed in [27].

Table 5.18: Classification accuracy obtained by different variations of EEGNet with different dropout rate (D) for 22 subjects using 128 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Subject	EEGNet-2.32		EEGNet-4.8	
	D = 0.25	D = 0.5	D = 0.25	D = 0.5
S1	0.873636	0.923	0.941818	0.931
S2	0.972727	0.954	0.990909	0.991
S3	0.95	0.961	0.98	0.970
S4	0.93	0.961	0.96	0.980
S5	0.779091	0.825	0.847273	0.843
S6	0.595455	0.587	0.538182	0.548
S7	0.99	1.000	1	0.990
S8	0.962727	0.971	0.981818	0.972
S9	0.99	1.000	0.990909	1.000
S10	0.829091	0.804	0.858182	0.945
S11	0.871818	0.923	0.950909	0.960
S12	0.940909	0.951	0.96	0.950
S13	0.990909	0.991	0.980909	0.990
S14	0.943636	0.962	0.941818	0.972
S15	0.930909	0.865	0.950909	1.000
S16	0.961818	0.990	0.99	1.000
S17	0.858182	0.915	0.943636	0.954
S18	0.970909	0.990	0.99	0.990
S19	0.934545	0.916	0.962727	0.925
S20	0.95	0.901	0.940909	0.970
S21	0.868182	0.857	0.933636	0.926
S22	0.916364	0.886	0.971818	0.972
AVG	**0.910**	**0.915**	**0.937**	**0.945**

Table 5.19: Classification accuracy obtained by different variations of EEGNet with different dropout rate (D) for 22 subjects using 47 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Subject	EEGNet-2.32		EEGNet-4.8	
	D = 0.25	D = 0.5	D = 0.25	D = 0.5
S1	0.766364	0.835	0.932727	0.922
S2	0.906364	0.896	0.926364	0.954
S3	0.942727	0.915	0.951818	0.915
S4	0.99	0.910	1	0.970
S5	0.713636	0.739	0.916364	0.895
S6	0.576364	0.547	0.536364	0.529
S7	0.774545	0.845	0.941818	0.990
S8	0.885455	0.933	0.931818	0.952
S9	1	1.000	1	1.000
S10	0.763636	0.827	0.820909	0.790
S11	0.770909	0.865	0.924545	0.971
S12	0.930909	0.940	0.99	0.960
S13	0.981818	0.990	0.971818	0.972
S14	0.912727	0.923	0.904545	0.933
S15	0.847273	0.895	0.895455	0.855
S16	0.873636	0.905	0.932727	0.942
S17	0.886364	0.865	0.940909	0.943
S18	0.98	0.990	0.99	0.990
S19	0.943636	0.883	0.951818	0.980
S20	0.855455	0.875	0.882727	0.903
S21	0.787273	0.762	0.82	0.867
S22	0.78	0.868	0.942727	0.932
AVG	**0.858**	**0.873**	**0.914**	**0.916**

5.3.4 Classification results

In this section classification results obtained by all models will be compared using graphical representations and statistical tests both for MNS vs. MI+MNS task and MI vs. rest task.

5.3.4.1 MNS vs. MI+MNS task

Classification accuracies obtained by EEGNet and other standard and deep learning models for all electrode layouts for MNS vs. MI+MNS task are presented in Figure 5.14 and Table 5.20.

Classification accuracies obtained by different models for 22 subjects using layouts with 128, 47, 13 and 6 electrodes, averaged over all folds (k= 10) for MNS vs. MI+MNS condition presented in Figures 5.15 - 5.18.

Boxplots showing the distribution of different models classification accuracy using layouts with 128, 47, 13 and 6 electrodes, averaged over all folds ($k = 10$) and all subjects ($n = 22$) for MNS vs. MI+MNS discriminating task represented in Figures 5.19-5.22. Significance levels of Wilcoxon signed-rank test comparisons showed in these figures means: NS – not significant, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

From the results presented in this section we can see that EEGNet-4.8 significantly ($p - value < 0.01$) outperforms all standard and deep learning models for all electrode setups (Figures 5.19-5.22), except 3-electrode setups where difference to second best model (EEGNet-2.32) is not significant (Figure 5.14) for MNS vs. MI+MNS task. The second best model for all electrodes setups is also an EEGNet architecture (EEGNet-2.32).

Also from Figure 5.14 we can see that performance of D-ConvNet and S-ConvNet become better when the electrode number decreases from 128 to 47, while it becomes worse for all other models. All deep learning models outperform standard models starting from setups with 47 electrodes and less.

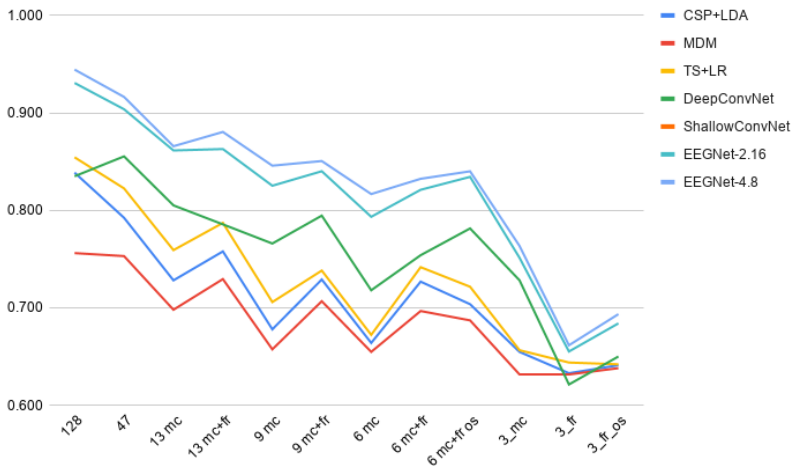


Figure 5.14: Grand average classification accuracy for different electrode layouts using different models, averaged over all folds ($k = 10$) and all subjects ($n = 22$) for MNS vs. MI+MNS condition. Input data filtered in 4-38 Hz frequency band.

Table 5.20: Grand average classification accuracy obtained by different classifiers using different layouts for MNS vs. MI+MNS discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation and across 22 subjects

Electrode layout	CSP+LDA	MDRM	TS+LR	D-ConvNet	S-ConvNet	EEGNet-2.32	EEGNet-4.8
128	0.839	0.756	0.854	0.835	0.834	0.915	0.945
47	0.792	0.753	0.823	0.855	0.867	0.873	0.916
13_mc	0.728	0.698	0.759	0.805	0.813	0.838	0.866
13_mc+fr	0.758	0.729	0.787	0.786	0.832	0.851	0.880
9_mc	0.678	0.657	0.706	0.766	0.793	0.812	0.846
9_mc+fr	0.729	0.707	0.738	0.795	0.804	0.835	0.851
6_mc	0.664	0.655	0.672	0.718	0.769	0.776	0.817
6_mc+fr	0.727	0.697	0.742	0.754	0.789	0.812	0.832
6_mc+fr_os	0.704	0.687	0.722	0.781	0.789	0.828	0.840
3_mc	0.655	0.632	0.656	0.728	0.738	0.752	0.764
3_fr	0.633	0.632	0.644	0.621	0.635	0.665	0.662
3_fr_os	0.641	0.638	0.642	0.650	0.660	0.669	0.693

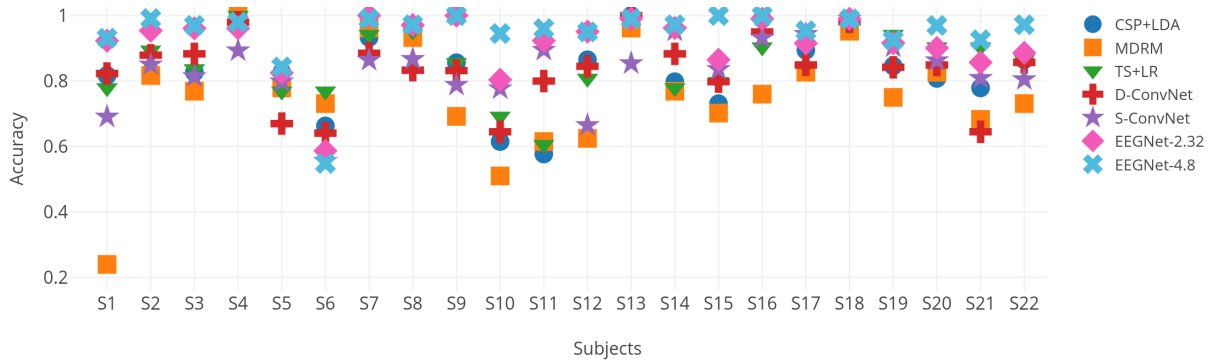


Figure 5.15: Classification accuracies obtained by different models for 22 subjects using layout with 128 electrodes, averaged over all folds ($k = 10$) for MNS vs. MI+MNS condition. Input data filtered in 4-38 Hz frequency band.

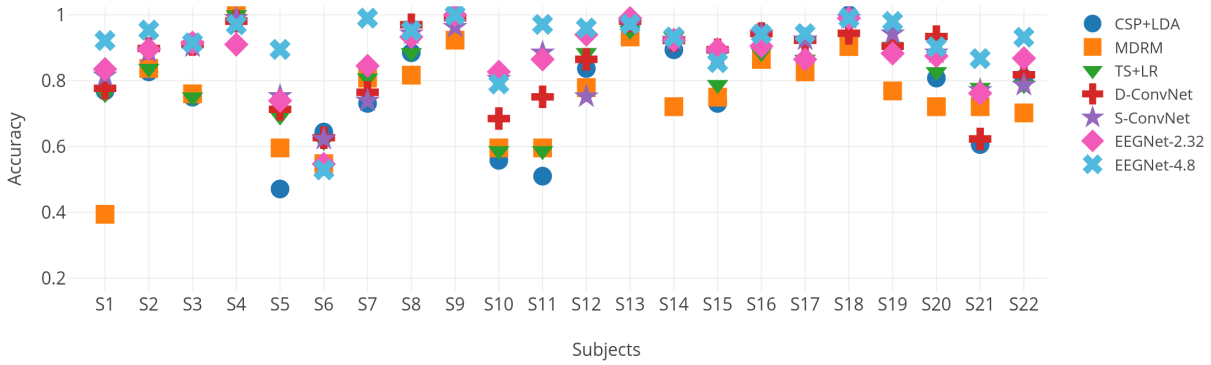


Figure 5.16: Classification accuracies obtained by different models for 22 subjects using layout with 47 electrodes, averaged over all folds ($k = 10$) for MNS vs. MI+MNS condition. Input data filtered in 4-38 Hz frequency band.

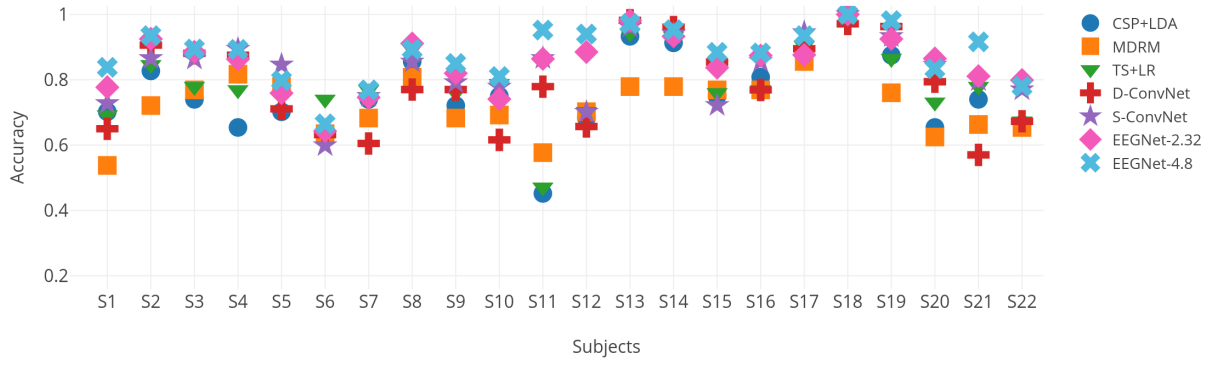


Figure 5.17: Classification accuracies obtained by different models for 22 subjects using layout with 13 electrodes over motor cortex and frontal lobe, averaged over all folds ($k = 10$) for MNS vs. MI+MNS condition. Input data filtered in 4-38 Hz frequency band.

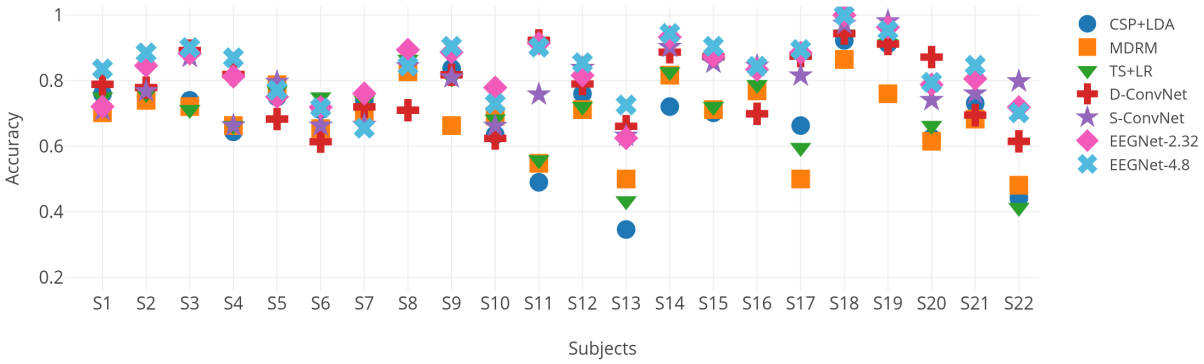


Figure 5.18: Classification accuracies obtained by different models for 22 subjects using layout with 6 electrodes over one side of motor cortex and frontal lobe, averaged over all folds ($k = 10$) for MNS vs. MI+MNS condition. Input data filtered in 4-38 Hz frequency band.

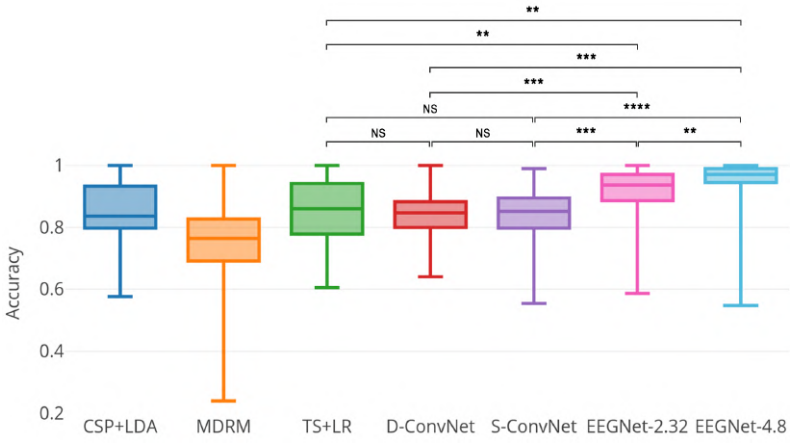


Figure 5.19: Boxplots showing the distribution of different models classification accuracy using layout with 128 electrodes, averaged over all folds ($k = 10$) and all subjects ($n = 22$) for MNS vs. MI+MNS discriminating task. Significance levels of Wilcoxon signed-rank test comparisons: NS – not significant, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

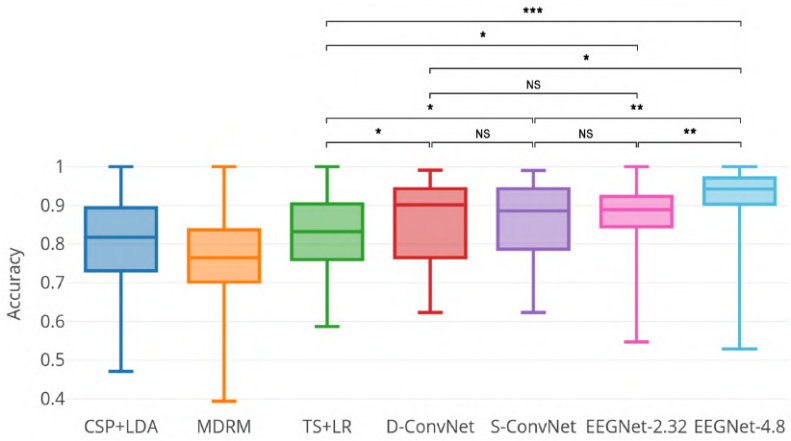


Figure 5.20: Boxplots showing the distribution of different models classification accuracy using layout with 47 electrodes, averaged over all folds ($k = 10$) and all subjects ($n = 22$) for MNS vs. MI+MNS discriminating task. Significance levels of Wilcoxon signed-rank test comparisons: NS – not significant, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

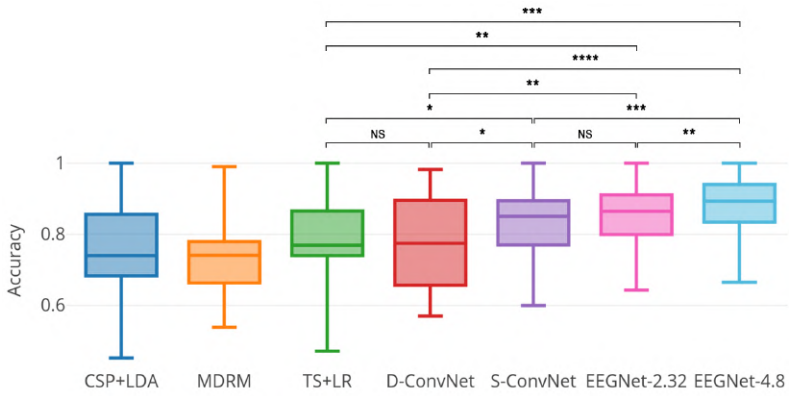


Figure 5.21: Boxplots showing the distribution of different models classification accuracy using layout with 13 electrodes over motor cortex and frontal lobe, averaged over all folds ($k = 10$) and all subjects ($n = 22$) for MNS vs. MI+MNS discriminating task. Significance levels of Wilcoxon signed-rank test comparisons: NS – not significant, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

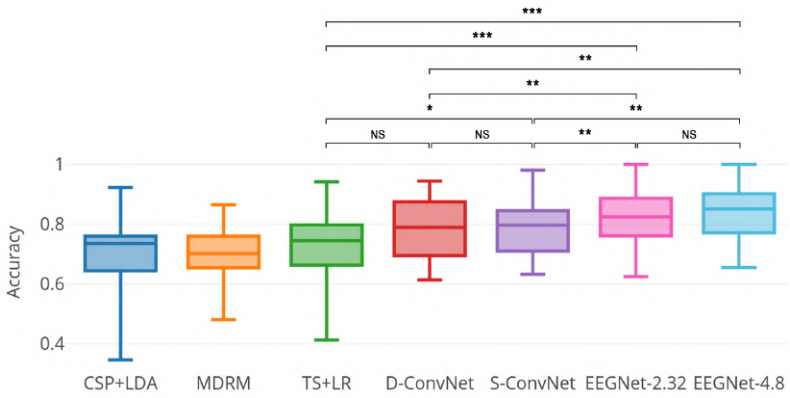


Figure 5.22: Boxplots showing the distribution of different models classification accuracy using layout with 6 electrodes over one side of motor cortex and frontal lobe, averaged over all folds ($k = 10$) and all subjects ($n = 22$) for MNS vs. MI+MNS discriminating task. Significance levels of Wilcoxon signed-rank test comparisons: NS – not significant, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

5.3.4.2 Motor imagery vs. rest

Classification accuracies obtained by EEGNet and other standard and deep learning models for all electrode layouts for MI vs. rest task are presented in Figure 5.23 and Table 5.21.

We can see that TS+LR outperforms other models for 128 and 47 electrodes for MI vs. rest task (Figure 5.23) but then performance decreases a lot when decreasing number of electrodes. At the same time, deep learning models are more stable to decreasing of electrodes number. And we even observe that a decreasing number of electrodes leads to increasing classification performance for all deep learning models with a peak at 9 electrode setup.

Statistical tests show that EEGNet-4.8 significantly ($p-value < 0.05$) outperforms TS+LR (and other standard methods) for 9 and 6 electrode setups.

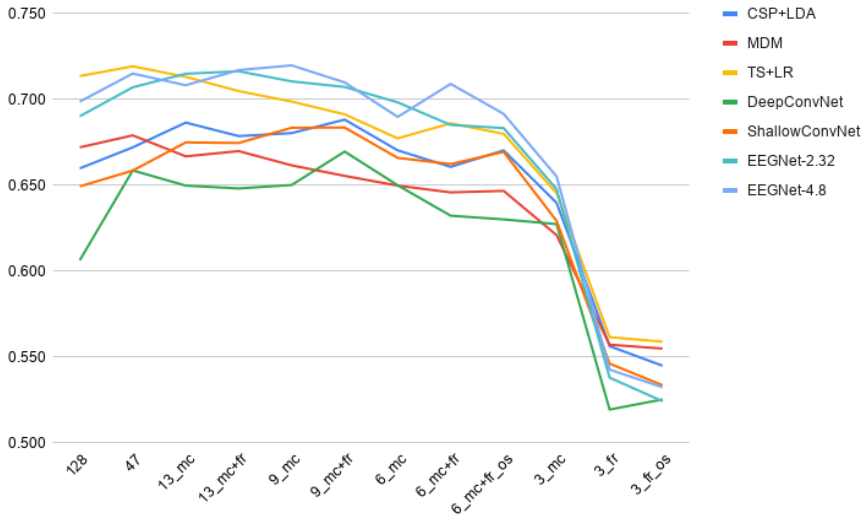


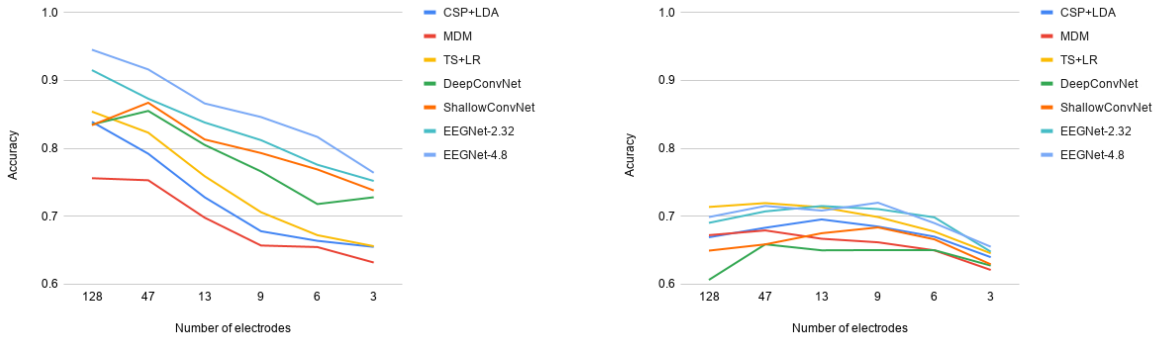
Figure 5.23: Grand average classification accuracy for different electrode layouts using different models, averaged over all folds ($k = 10$) and all subjects ($n = 22$) for MI vs. rest condition. Input data filtered in 8-30 Hz frequency band.

Table 5.21: Grand average classification accuracy obtained by different classifiers using different layouts for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation and across 22 subjects

Electrode layout	CSP+LDA	MDRM	TS+LR	D-ConvNet	S-ConvNet	EEGNet-2.32	EEGNet-4.8
128	0.639	0.654	0.706	0.615	0.639	0.677	0.699
47	0.644	0.670	0.722	0.642	0.681	0.709	0.703
13_mc	0.662	0.658	0.708	0.640	0.669	0.706	0.713
13_mc+fr	0.674	0.661	0.692	0.652	0.684	0.706	0.738
9_mc	0.664	0.653	0.679	0.644	0.685	0.718	0.730
9_mc+fr	0.672	0.653	0.695	0.642	0.670	0.690	0.711
6_mc	0.672	0.641	0.668	0.640	0.662	0.697	0.700
6_mc+fr	0.656	0.635	0.660	0.605	0.661	0.696	0.689
6_mc+fr_os	0.653	0.628	0.666	0.619	0.677	0.676	0.686
3_mc	0.629	0.612	0.629	0.608	0.631	0.669	0.657
3_fr	0.544	0.542	0.557	0.515	0.562	0.537	0.541
3_fr_os	0.525	0.534	0.529	0.514	0.555	0.529	0.539

5.3.5 Comparing paradigms

Classification performance obtained by EEGNet and other standard and deep learning models for several electrode layouts for MNS vs. MI+MNS and MI vs. rest tasks could be compared in Figure 5.24.



(a) MNS vs. MI+MNS task

(b) MI vs. rest task

Figure 5.24: Comparison of classification accuracy for different electrode layouts (128, 47, 13, 9, 6, 3), averaged over all folds ($k = 10$) and all subjects ($n = 22$) for MNS vs. MI+MNS (5.24a) and MI vs. rest (5.24b) conditions. Input data filtered in 4-38 Hz frequency band for 5.24a and in 8-30 Hz frequency band for 5.24b.

The difference between the results of MNS vs. MI+MNS and MI vs. rest paradigms was demonstrated earlier in [20]. But previous investigations were done only for 128 electrode setup. In this work, we also present it for a smaller number of electrodes. Boxplots with classification accuracy distribution for EEGNet-4.8 model for setups with 13, 9, 6 electrodes located over motor cortex are shown in Figure 5.25. Statistical tests indicate highly significant ($p < 0.001$) difference between EEGNet-4.8 accuracies of both tasks for 13, 9 and 6 electrodes.

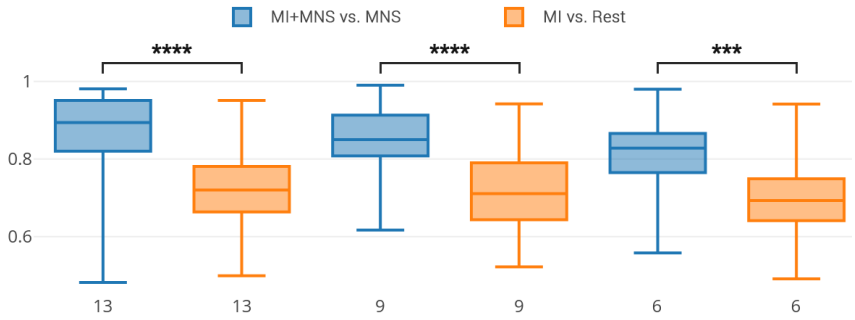


Figure 5.25: Boxplots showing the distribution of EEGNet-4.8 classification accuracy for different setups of n electrodes over motor cortex and frontal lobe, averaged over all folds ($k = 10$) and all subjects ($n = 22$) for both discriminating tasks. Significance levels of Wilcoxon signed-rank test comparisons ***$p < 0.001$, ****$p < 0.0001$.

5.3.6 Frontal electrodes investigation

To explore the opportunities of including the EEG activity from different functional areas of the brain, we investigated setups with electrodes not only from motor cortex but also adding electrodes from frontal lobe (Figures 4.6c, 4.6e, 4.6g) because frontal electrodes are used in many monitoring systems to determine the depth of anesthesia [38].

Figure 5.26 shows classification performance of models for MI + MNS vs MNS discriminating task on the data from 13, 9, and 6 electrodes over the motor cortex and frontal lobe (distinguished by prefix $+fr$ in the chart), and motor cortex only.

Results show that the addition of frontal lobe electrodes increases accuracy almost for all models and all numbers of electrodes. Wilcoxon signed-rank tests confirm significant differences between setups with and without frontal lobe electrodes for many models. For example, the difference is significant ($p - value < 0.05$) for TS+LR and EEGNet-2.32 for setups with 9 and 6 electrodes. Moreover, it is highly significant for the TS+LR model using 6 electrodes. At the same time, statistical tests show that there is no significant difference between setups with and without frontal lobe electrodes for EEGNet-4.8.

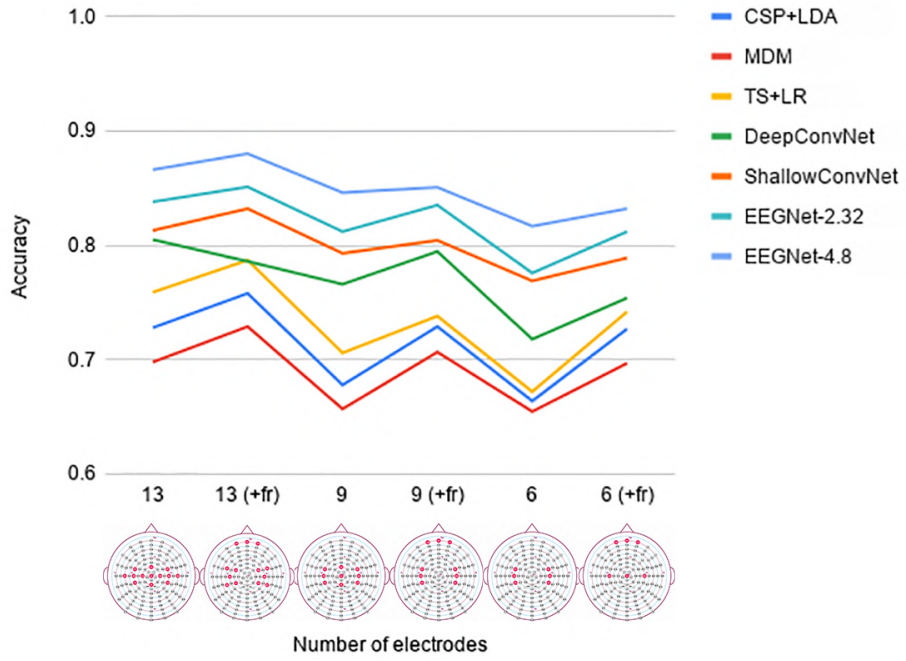


Figure 5.26: Models classification accuracy for electrode setups over motor cortex and over motor cortex and frontal lobe, averaged over all folds ($k = 10$) and all subjects ($n = 22$) for MNS vs. MI+MNS condition.

5.4 Conclusion

In this chapter results for 3 different machine and deep learning applications to improve BCI performance are represented.

First, it was presented an approach of using outlier detection to estimate the quality of motor imageries prior to BCI training. Results showed that this approach is subject-specific and increases classification accuracy in BCI for a group of subjects without significant average accuracy improvement.

Secondly, the enhancement of combined movement motor imagery classification was studied. Method using 3 binary classifiers was proposed for combined MI movement classification mainly to reduce the time for signal acquisition, training, and calibration of BCI. Using this approach together with the feature extraction method it is possible to achieve the same results as using the standard multi-class approach. Add here why this study and result is useful for mankind and science, anyway.

In this chapter, we have investigated motor imagery detection using deep learning methods. During this study, two different discriminating tasks were considered: median nerve stimulation (MNS) vs. MI+MNS and MI vs. rest as a standard task for the BCI domain. It is a complex study that includes several components:

- frequency band selection, where using wider (4-38 Hz) frequency band was proposed instead of standard (8-30 Hz) band. It was showed that using a wider (4-38 Hz) band allows to significantly increase classification performance;
- EEGNet architectures investigation, where lengths of temporal kernel size for the first convolutional layer and multiplier depth, which determines the number of spatial filters of the depthwise convolution layer were studied. The parameter of the dropout rate for the model also was investigated. EEGNet adaptation, namely EEGNet-4.8 showed the best performance for different electrode layouts for both tasks. Results of EEGNet-4.8 classification are significantly higher than any other classical and deep learning models;
- classification performance for two paradigms MNS vs. MI+MNS and MI vs. rest were compared. It was showed that using MNS vs. MI+MNS discriminating task allows achieving significantly higher classification accuracy than standard MI vs. rest for all investigated electrode layouts.

- frontal electrode investigation, where it was showed that electrode layouts with electrodes over the frontal lobe together with electrodes over the motor cortex allows achieving significant classification accuracy improvement comparing with electrode layouts without frontal electrodes;
- electrode layout investigation where 12 setups with different channel numbers and localization were studied. Results showed that it is possible to achieve high classification performance using a reduced number of electrodes (84.0% using just 6 electrodes over one side of the motor cortex and frontal lobe).

All of the above allows us to discuss the application of deep learning to MNS-based BCI for intraoperative awareness prevention in the next chapter.

Chapter 6

Discussion

Contents

In previous chapters, we have proposed and considered three different ways of improving motor imagery detection in the BCI domain on tree different datasets described in chapter 4.

Firstly, we have proposed to assess the quality of motor imageries using outlier detection in section 2.4.1. This was followed by the cleaning of the original data from outliers to improve model training and accuracy of constructing decision boundaries for classifiers. Results for improving a BCI using outlier detection as a quality index of motor imagery estimation are presented in section 5.1.

Secondly, to improve combined movement motor imagery classification and reduce efforts for EEG signal recording and BCI calibration, we have proposed the solution based on three parallel-connected binary classifiers in section 2.5.6. We have represented results for classification of combined movement motor imageries using binary classifiers in section 5.2.2. Also, we have introduced results for feature selection applied to parallel-connected binary classifiers in section 5.2.3.

Finally, we have presented four convolutional neural networks (D-ConvNet, S-ConvNet, EEGNet-2.32, and an adaptation of the latest EEGNet-4.8) as advanced feature extraction and classification techniques for motor imagery detection from raw EEG signals in section 3.6. Results for optimizing MI detection with deep learning using MNS-based BCI are presented in section 5.3.

In this chapter, proposed methods and results presented in the previous chapter will be analyzed and discussed. We will discuss the following points:

- Outlier detection for automatic assessing the quality of imaginary movements in MI-based BCI in section 6.1;
- Combined movement motor imagery classification in section 6.2, where we will review the classification of combined movement motor imageries using binary classifiers and feature selection;
- Imaginary movements detection using deep learning to prevent intraoperative awareness in section 6.4. Here we will touch on questions about different deep learning architectures, frequency bands, and layout selections. We will discuss the importance of frontal electrodes for classification accuracy. Also will discuss perspectives of the proposed method for prevention of intraoperative awareness during general anesthesia.

6.1 Outlier detection for automatic assessing the quality of imaginary movements in MI-based BCI

Assessing the quality of imaginary movements before training a BCI using outlier detection is an important problem that could increase the accuracy of the classification and improve BCI training. The clustering algorithms used in the work (one-class SVM method and isolation forest) allow to detect and remove around 30% of outliers from the data (see Tables 5.1 and 5.2) and thus perform classification on the cleaned dataset.

However, the obtained classification accuracy of imaginary movements before and after the outlier detection allows us to conclude that this approach is subject-specific and requires additional adjustment of parameters separately for each subject. Imaginary movement classification accuracy increased for ten over twenty-nine subjects (+11.3% in average) using isolation forest method (Figure 5.2 page 83) and for nine subjects (+9.6% in average) using one-class SVM method. But for other subjects classification accuracy was not changed or decreased. This suggests that this approach is subject-specific and can be successfully applied to a specific group of subjects. At the same time, average accuracy across all subjects was not changed so this means that the method of outlier detection in the initial data has to be improved.

In our opinion, this can be achieved by one or more of the following options: changing the architecture of the method; using more discriminative features to better separate classes before outlier detection; using deep learning methods for both feature extraction and outlier detection parts.

6.2 Classification of combined movement motor imageries using binary classifiers

Our study shows that the electrical activity of combined motor imageries measured on the scalp is very similar to the electrical activity measured for each simple motor imagery taken individually. Thus an efficient classification should be able to distinguish them.

The low classification rate obtained with classical methods in section 5.2.1 could be linked with the difficulty to perform combined movement motor imagery in a proper way as well as with bad class separation using current feature extraction techniques.

We tried to achieve mainly three goals in our study about combined movement classification:

- to reduce the time for signal acquisition, BCI training, and calibration using an architecture with 3 binary classifiers, which does not require combined movement motor imageries for training, just three classes of MIs: right hand, left hand, and feet.
- to investigate the importance of features obtained from the signals filtered in different bands;
- to improve the performance of combined movement MIs classification overall.

Results for 3 parallel stacked binary classifiers represented in section 5.2.2.

It is obvious that the proposed solution could be useful to save time for recording, training, and calibration stages as we will not collect data for combined movements before training. But at the same time the results (Tables 5.5 - 5.10) of such approach are worth than using multi-label classification (Table 5.3).

The best accuracies are obtained using a 30-feature set, which confirms the hypothesis that usage of features filtered in different frequency bands could be useful. We show that an increasing number of features leads to increased accuracy of classification even with our small number of training examples. But the final results are still too small to implement this approach to combined movements classification.

It can be seen that the accuracy of the resting state classification is low (Table 5.10), even considering the fact that the resting state was in the training data. But in general, we have a predictable tendency to obtain a good result for known classes and bad performance on the unknown classes.

To start with, detection of the rest state should be improved because accuracy for these conditions is very low. It was shown on histograms - 'rest' and 'feet' classes are not well separated (Figures 5.15-5.16). The solution could be in addition one more serial connected classifier before 3 parallel to detect any movement activity. If there is no movement activity - it is a resting state.

6.3 Feature selection for combined movement motor imagery detection

The proposed method for feature selection showed that not all of the features in our feature set are equally useful for the classification. Most prominent features are alpha ERD and wide beta ERS features (Tables 5.12 and 5.13) for the right and left hand MI detection. And ERS beta (17-25 Hz) together with ERS alpha+beta (7-30 Hz) features on Cz were chosen by feature selection for the feet MI detection. It corresponds to described in the literature earlier [49, 23]. But it is interesting that the feature selection method also chooses ERD alpha band (8-12 Hz) features from C3 and C4 channels to separate feet activity from others (Table 5.14)). It could be evident that this feature was used not for feet detection, but for the detection of right and left hand activity in the data which was trained versus the feet class.

Histograms of logarithmic ERD/ERS features distribution (Figures 5.15, 5.16) demonstrate that interclass separation is low even for the best features chosen by the feature selection method. Feet and rest classes overlap (Figure 5.7a) which can be a reason for low classification accuracy for the Rest class.

Another useful feature could be ERD/ERS in the alpha+beta band (8-30 Hz). As we can see the separation of the 'rest' and 'feet' classes are much better than in other features for this band (table 5.15).

It was showed that classification accuracy is higher after feature selection (Table 5.17). But at the same time, we have not a significant difference between results.

As already showed and described in the section 5.2.3 the best accuracy obtained using binary classifies with feature selection (33.2% $\pm$ 1%) is higher than using standard multi-label approach (32.7% $\pm$ 4.2%) (see Table 5.3). But the difference is not significant and quite low.

Results of feature selection investigation and classification represented only for 10 subjects and require further investigations. But even with this amount of data, we can make a generalized view of the vulnerabilities of our approach. The main achievement of this approach is reduced time for model training and calibration (using 4 classes for training to predict 8 classes). But the result remains low for implementing this approach to control a BCI.

To improve the current method more features could be considered as well. As was presented in section 5.2.3, usage of alpha, beta, and alpha+beta bands together can be a solution to better separate classes. And it can have a positive influence on the classification accuracy. Furthermore, as known from [285, 286] it is better to use different bands for detection of different movements (for example for finger movement the largest beta ERS was found in the 16$\pm$21 Hz band and for foot movement in the slightly higher 19$\pm$26 Hz band). But for the current experimental design, it seems impossible due to combined several different movements.

Also, classification in the BCI domain is subject-specific and there are a lot of parameters that could increase accuracy for each individual subject (for example we use only one timer for ERD calculation and one timer for ERS calculation). Thus creating user-profiles and an automated recognition system is required.

Moreover, better feature extraction techniques should be found. Results obtained in section 5.2.3 showed that classes are not well separated.

An incremental motor imagery learning using first real movements and then imaginary movements could be useful to train the user to produce good motor imagery.

Also, each class is represented by a few numbers of samples (30 for training and 10 for testing). Increasing the number of samples for each class should improve performance.

6.4 Imaginary movements detection using deep learning to prevent intra-operative awareness

Results in section 5.3 show the ability of deep learning techniques to extract features from a few EEG channels to detect motor imagery when the subject is being stimulated over the median nerve, which could represent in particular a regain of consciousness during general anesthesia.

6.4.1 Different deep learning architectures

We tested eight different modifications of EEGNet, where we i) reduced the first layer kernel size from 32 to 16, 8 and 4 to better handle mid- and high-frequency features; and ii) changed multiplier depth from 2 to 4, which

determines the number of spatial filters of the depthwise convolution layer. We choose EEGNet-4.8 (see section 5.3.2) as most prominent architecture based on the results.

Also, we tested another deep learning models such as D-ConvNet and S-ConvNet. S-ConvNet and EEGNets showed high performances. Statistical tests show that EEGNet-4.8, proposed in this paper, significantly outperformed all standard and deep learning approaches especially for MNS + MI vs. MNS task in terms of accuracy.

The higher performance of EEGNet-4.8 compare to EEGNet-2.32 for MNS + MI vs. MNS task can be explained by the presence of useful high-frequency components in the EEG signals for this task unlike MI vs. rest task. These components are better kept using a shorter kernel size. It is also shown in Figure 5.11, where we can see better performances for 8-38 Hz than for standard 8-30 Hz frequency band. Usually, there is intuition to use smaller kernels for detecting high-frequency features and larger kernels for low-frequency features [273]. That's why reducing the first layer kernel size from 32 to 8 is useful for MNS + MI vs. MNS task.

Also as we use the same kernel for a different set of samples in a signal, the same weights are shared across these sample sets. This means that a smaller kernel size leads to using fewer weights to back-propagate on. So the benefit of using a smaller kernel is a reduction in computational costs.

Changing multiplier depth from 2 to 4 in the second layer to increase the number of spatial filters allowed to store more information in this layer. Moreover, EEGNet models have a significantly lower number of training parameters comparing to other deep learning models D-ConvNet and S-ConvNet (Table 3.5), which results in better overfitting resistance.

6.4.2 Frequency band selection

We found that using a 4-38 Hz frequency band is helpful for better distinguishing motor patterns resulting in MNS + MI and MNS. Previously, the 8-30 Hz frequency range was mainly used in MI-based BCI (see Table 3.5) , but adding the theta and low gamma frequency ranges resulted in an increase of accuracy for each electrode layout studied in our experiments.

Though, the effect of hand motor function in the theta 4.8–7.9 Hz band of somatosensory electrical stimulation has already been shown [292]. They observed a significant decrease in mean motor ipsilesional resting-state power.

On the other hand, there is evidence of motor activities in the gamma band [293, 294, 295, 296].

6.4.3 Addition of the frontal electrodes to improves motor imaginary detection

When investigating the effect of adding channels from the frontal cortex as additional input data to the models, we noticed that for many models the accuracy increases comparing to the case when only the channels from the motor cortex were used. This was the case for the less accurate models (e.g. TS+LR and EEGNet-2.32), while there were no significant improvements for the best-obtained model (EEGNet-4.8). The evident increase in classification accuracy between MI + MNS vs. MNS after adding frontal electrodes might have different causes, leading to models perceiving the presence of MI conditioned to nerve stimulation.

Possible reasons may be rooted in participation of the prefrontal cortex in response implementation, providing a functional substrate for the transformation of perception into action [297], or regions of the frontal cortex being a part of the cortical networks involved in the complex experience of pain [298]. In [299] is reported that individuals with the poor ability of motor imagery demonstrated the recruitment of orbitofrontal cortex, among others, while performing MI, and results of [300] demonstrated that the sports experts had activation patterns in prefrontal areas while performing motor imagery tasks. These results suggest that characteristic activity related to our MI + MNS vs. MNS might be encountered in the motor cortex area.

Although current results show that the proposed EEGNet-4.8 model with the highest accuracy demonstrates that the inclusion of the frontal electrodes does not lead to even more statistically significant improvements in terms of accuracy, we suggest to use frontal electrodes for discriminating MI + MNS and MNS since adding them statistically significantly improves the performance for most of the classifiers (Figure 5.26). Therefore we believe that the potential of looking at the frontal area is high and this direction will be addressed in our future research. Information from the frontal cortex is useful and the ways of integrating it into classification models in order to obtain a more significant increase in accuracy should be explored further.

6.4.4 Using reduced electrode layouts for motor imagery detection

In parallel with testing the performance of many models, we addressed the problem of minimization of the electrode number in the recording setup, to reduce the preparation and calibration time in clinical settings. It is especially important for the task of detecting intraoperative awareness, where preparation time before surgery is a precious resource. Starting from 128 electrodes layout which is standard for the Biosemi system (Figure 4.5), we tested the dependency of the performance on the electrode number for 48, 13, 9, 6, and 3 electrodes covering mainly the motor cortex area.

We found that it is possible to reduce the number of electrodes to 13 or even 6 electrodes without substantial loss of classification performance. The best model (EEGNet-4.8) achieved 88% accuracy for 13 electrodes setup (84% for 6 electrodes setup) for MNS + MI vs. MNS task. Since such a task with a few electrodes was not addressed in the literature previously, this is the highest accuracy to the best of the authors' knowledge. These values are higher than the available values for a model that distinguishes MI vs. rest, which is 73.8% for 13 electrodes setup (68.9% for 6 electrodes setup). Deep learning models showed higher stability when decreasing the number of electrodes for MNS + MI vs. MNS task. At the same time, the best classification performance for MI vs. rest task has been achieved on the setup with 9 electrodes.

Conclusions

This manuscript presents three main contributions to improve the detection of motor imageries which are used by numerous brain-computer interfaces as a means of interaction. Firstly, we proposed to estimate the quality of motor images by detecting outliers and removing them before learning. Secondly, we studied the selection of features (timing, frequency bands, and electrode) for motor imagery classification and created a method based on three binary parallel-connected classifiers to distinguish 7 motor imageries using just 3 of them for training the BCI. Thirdly, we presented a large review on deep learning and proposed a deep learning architecture based on EEGNet applicable directly to filtered electroencephalographic signals adapted to the number of electrodes. We showed in particular its benefits for improving the detection of intraoperative awareness and other MI-based BCI applications.

Application of the outlier detection method for preliminary preprocessing a dataset and remove bad trials allowed to significantly increase classification accuracy of motor imagery detection for 10 over 29 subjects (+11.3% on average) using isolation forest method and for 9 over 29 subjects (+9.6% on average) using one-class SVM method. But for other subjects classification accuracy was not changed or decreased. Thus, removing outliers from the training data helps to build more accurate classification models for a group of subjects.

The proposed approach for classifying combined motor imageries consists of feature selection and classification using an architecture with three binary parallel-connected classifiers for detecting three basic movements (right hand, left hand, and feet) and further detection of any combinations of these movements. We have studied the selection of different band-power features according to timing windows, frequency bands, and electrodes. We have found that the best features were obtained from the alpha (8-12 Hz) and wider alpha+beta (8-30 Hz) band just after the beginning of the MI task (0.5-6 seconds), corresponding to the event-related desynchronization over C3, C4, and Cz electrodes. We have shown that using this approach it is possible to achieve the same classification accuracy level (33.2% $\pm$ 1%) as using ERD/ERS+LDA standard multi-class approach (32.7% $\pm$ 4.2%) but with reduced efforts for signal acquisition, training, and calibration of a BCI because we use only 3 electrodes instead 26 for EEG acquisition and train BCI using only 3 MI classes (right hand, left hand, and feet) to detect 7 classes (combinations of these basic MI movements are added to the previously mentioned ones).

In this work, it was proposed to use median nerve stimulation together with deep learning on the filtered EEG data for motor imagery detection. We demonstrated that for motor imagery detection while the subject is being stimulated via the median nerve, it is advisable to use EEG signals recorded from as few as 6 electrodes in an extended 4-38 Hz frequency range processed by an adapted architecture of EEGNet deep neural network, which can give 84% of classification accuracy, comparing to 74.2% obtained with the best performed standard method (TS+LR). The combination of the novel approach of MNS with a larger number of electrodes results in even higher accuracy (94.5% for 128 electrodes and 88.0% for 13 electrodes, comparing to 85.4% and 78.7% respectively, obtained with TS+LR). These results set the background for clinical application with anesthetics of the proposed setup for accessing the intraoperative awareness.

Contribution

The main contribution of this thesis is a demonstration that the deep learning approach to motor imagery detection in raw EEG signals is also more advanced than standard existing machine learning methods.

We have done a complex inclusive review of deep learning for EEG signal analysis in the motor imagery-based BCI domain. Data from 71 reviewed articles were summarized. The main trends in the application of deep learning architectures, optimization algorithms, activation functions, and feature extraction techniques for MI-based BCIs

were considered.

We have improved EEGNet architecture to better handle mid and high-frequency components from filtered EEG data, identified wider frequency band, determining the importance of frontal electrodes, and showed on a dataset with 22 subjects that these improvements all together allows us to obtain a classification accuracy of 94.5% for 128 electrode layout and 84% for 6 electrode layout for the proposed MNS vs. MI+MNS discriminative task. This result is the state of the art in the task of MI detection while the median nerve is stimulated. This result is evidence that deep learning architectures not only can work well on raw EEG data with a small number of samples but also beat and replace existing standard machine learning techniques in this domain.

Thus, the current thesis contributes to the MI-based BCI domain. Moreover, the findings of this work could be applied to the task of accidental awareness detection during general anesthesia. Also, the proposed deep learning architecture with a set of parameters for feature extraction and classification from raw EEG signals might be beneficial to the EEG signal analysis community in general.

Future work

We introduce time horizons for future work highlighting short-term, intermediate-term, and long-term planning.

- Short-term plans

  The main short-term purpose is to integrate the three contributions of this thesis in one pipeline. We already applied our knowledge obtained in the study on the selection of timing, frequency band, and electrode characteristics to the application of accidental awareness detection during general anesthesia. The integration of unsupervised outlier detection for cleaning the dataset requires further investigations to improve classification performance.

- Intermediate-term plans

  Although the use of median nerve stimulation is promising, it still requires more detailed investigations. Another challenge for this kind of BCI remains to be used in clinical routine.

  Although most anesthetics reduce the global metabolic activity of the central nervous system by decreasing neuronal excitability [301, 302], the effect of anesthetics on the sensorimotor cortex appears to be relatively well preserved [95, 45]. In fact, regions of the cortex would still be receptive to information, but their ability to communicate with other regions seems to be damaged during general anesthesia [303], suggesting that our method can be used in patients with clinical conditions [45].

  Further experiments are planned to be done in clinical conditions under general anesthesia to investigate the influence of anesthetics on the ability to detect motor activity in real-life conditions.

  Also, it is important to continue studying user profiles with a set of parameters in case of subject-specificity due to high variance between subjects. Further investigations should be aimed at application different cross-validation strategies, such as leave-one-subject-out and leave-one-group-of-subjects-out. Such an experimental design is much closer to real-life experience and requires subject-to-subject transfer learning, so this might lead to a decrease in classification performance. The current cross-validation approach presented in this thesis can be well-suited if we consider taking some training examples from an observed subject for training or calibration which is not always possible.

  In the context of detection of intraoperative awareness from EEG signals, no training patterns of the intention of movement under propofol are available in advance to train a specific model for the current subject. So the findings of this thesis concerning the benefits of an extension of the frequency band to 4-38 Hz and of using frontal electrodes are useful but the calibration phase has to be adapted. A transfer learning method from other subjects performing a motor intention under propofol should be done.

  In order to tackle this challenge, two different machine learning approaches could be explored: i) transfer learning approaches, with the idea to learn a motion intention detection model on a subject excluding anesthesia (where we can have a lot of data) and to transfer this model to a context (where we have little data), in

particular by transforming data under anesthesia into data outside anesthesia. The transformation necessary to pass from the covariance matrix of the signals recorded during rest under anesthesia (for example at the very beginning of the operation) to that of the rest signals recorded without anesthesia will provide us with an estimate of the matrix of covariance of motion intent signals during anesthesia from motion intent signals outside anesthesia; ii) adaptive learning approaches, where the idea is to build adaptively, starting from zero, a model describing the properties of EEG signals [74]. An intention to move will then be identified by rupture detection when the property of the EEG signals changes.

One of the possible approaches is to adapt Riemannian geometry to transfer learning. We already have a database with 15 subjects for this experiment and a method such proposed by [304] can be applied in the intermediate-term perspective.

- Long-term plans

Another approach based on transfer learning from other subjects performing a motor intention under propofol could be done in the long-term perspective. This method consists of training a network using a large EEG database such as ones available for automatic sleep scoring [305, 306, 307] to extract principal EEG features. Then, pre-trained layers of this network could be used to train the last layers [308] on our specific EEG data under general anesthesia.

Thus, transfer learning can allow avoiding a long calibration of the model on patient-specific EEG of motor imageries using data and models of other anesthetized subjects.

In particular, it is not possible to ask the patient to follow a protocol for several minutes to calibrate a model before a surgery. Deep learning can also be used for collecting larger synthetic amounts of data [309, 310]. A generative adversarial network could be applied in this case to generate and collect missing training patterns of the intention of movement under propofol using only EEG signals acquired while subjects stimulated via the median nerve.

Also, in the long-term, we plan to study the application of median nerve stimulation to detect several various motor imageries (for example, right arm, left arm, head, and full body) to increase the number of possible interactions of BCI.

Bibliography

[1] L. E. H. van Dokkum, T. Ward, and I. Laffont, "Brain computer interfaces for neurorehabilitation–its current status as a rehabilitation strategy post-stroke," *Annals of Physical and Rehabilitation Medicine*, vol. 58, pp. 3–8, Feb. 2015.

[2] R. G. Lupu, D. C. Irimia, F. Ungureanu, M. S. Poboroniuc, and A. Moldoveanu, "BCI and FES Based Therapy for Stroke Rehabilitation Using VR Facilities," Apr. 2018. ISSN: 1530-8669 Pages: e4798359 Publisher: Hindawi Volume: 2018.

[3] A. A. Frolov, O. Mokienko, R. Lyukmanov, E. Biryukova, S. Kotov, L. Turbina, G. Nadareyshvily, and Y. Bushkova, "Post-stroke Rehabilitation Training with a Motor-Imagery-Based Brain-Computer Interface (BCI)-Controlled Hand Exoskeleton: A Randomized Controlled Multicenter Trial," *Frontiers in Neuroscience*, vol. 11, 2017. Publisher: Frontiers.

[4] M. A. Cervera, S. R. Soekadar, J. Ushiba, J. d. R. Millán, M. Liu, N. Birbaumer, and G. Garipelli, "Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis," *Annals of Clinical and Translational Neurology*, vol. 5, pp. 651–663, Mar. 2018.

[5] D. J. McFarland, J. Daly, C. Boulay, and M. Parvaz, "Therapeutic Applications of BCI Technologies," *Brain computer interfaces (Abingdon, England)*, vol. 47, no. 1-2, pp. 37–52, 2017.

[6] L. Huang and G. v. Luijtelaar, "Brain Computer Interface for Epilepsy Treatment," *Brain-Computer Interface Systems - Recent Progress and Future Prospects*, June 2013. Publisher: IntechOpen.

[7] D. Zhang, D. Cao, and H. Chen, "Deep Learning Decoding of Mental State in Non-invasive Brain Computer Interface," *arXiv:1911.05661 [cs, eess]*, Nov. 2019. arXiv: 1911.05661.

[8] A. Myrden and T. Chau, "A Passive EEG-BCI for Single-Trial Detection of Changes in Mental State," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 345–356, Apr. 2017. Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.

[9] T. S. Rached and A. Perkusich, "Emotion Recognition Based on Brain-Computer Interface Systems," *Brain-Computer Interface Systems - Recent Progress and Future Prospects*, June 2013. Publisher: IntechOpen.

[10] M. Mohammadpour, S. M. R. Hashemi, and N. Houshmand, "Classification of EEG-based emotion for BCI applications," in *2017 Artificial Intelligence and Robotics (IRANOPEN)*, pp. 127–131, Apr. 2017.

[11] E. P. Torres P., E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo, "EEG-Based BCI Emotion Recognition: A Survey," *Sensors*, vol. 20, p. 5083, Jan. 2020. Number: 18 Publisher: Multidisciplinary Digital Publishing Institute.

[12] E. Maby, M. Perrin, O. Bertrand, G. Sanchez, and J. Mattout, "BCI Could Make Old Two-Player Games Even More Fun: A Proof of Concept with "Connect Four"," *Advances in Human-Computer Interaction*, December 2012.

[13] I. Martišius and R. Damaševičius, "A Prototype SSVEP Based Real Time BCI Gaming System," *Computational Intelligence and Neuroscience*, March 2016.

[14] L.-D. Liao, C.-Y. Chen, I.-J. Wang, S.-F. Chen, S.-Y. Li, B.-W. Chen, J.-Y. Chang, and C.-T. Lin, "Gaming control using a wearable and wireless EEG-based brain-computer interface device with novel dry foam-based sensors," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, p. 5, Jan. 2012.

[15] K. Oum, H. Ayaz, P. A. Shewokis, and P. Diefenbach, "MindTactics: A Brain Computer Interface gaming platform," in *2010 2nd International IEEE Consumer Electronics Society's Games Innovations Conference*, pp. 1–5, Dec. 2010. ISSN: 2166-675X.

[16] M. Ahn, M. Lee, J. Choi, and S. C. Jun, "A Review of Brain-Computer Interface Games and an Opinion Survey from Researchers, Developers and Users," *Sensors (Basel, Switzerland)*, vol. 14, pp. 14601–14633, Aug. 2014.

[17] Y. Fang, M. Chen, X. Zheng, and R. F. Harrison, "Extending CSP to Detect Motor Imagery in a Four-class BCI," *Journal of Information & Computational Science*, vol. 9, no. 1, 2012.

[18] Y. Wang, S. Gao, and X. Gao, "Common Spatial Pattern Method for Channel Selelction in Motor Imagery Based Brain-computer Interface," *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 5, pp. 5392–5395, 2005.

[19] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanaba, and K. Müller, "Optimizing spatial filters for robust EEG single-trial analysis [revealing tricks of the trade]," *IEEE Signal processing magazine*, 2008.

[20] S. Rimbert, P. Riff, N. Gayraud, D. Schmartz, and L. Bougrain, "Median Nerve Stimulation Based BCI: A New Approach to Detect Intraoperative Awareness During General Anesthesia," *Frontiers in Neuroscience*, vol. 13, p. 622, June 2019.

[21] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Riemannian geometry applied to BCI classification," in *International Conference on Latent Variable Analysis and Signal Separation*, pp. 629–636, Springer, 2010.

[22] M. Congedo, A. Barachant, and A. Andreev, "A New Generation of Brain-Computer Interface Based on Riemannian Geometry," *arXiv:1310.8115 [cs, math]*, Oct. 2013. arXiv: 1310.8115.

[23] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass Brain-Computer Interface Classification by Riemannian Geometry," *IEEE Transactions on Biomedical Engineering*, vol. 59, pp. 920–928, Mar. 2012.

[24] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of Neural Engineering*, vol. 16, p. 051001, Aug. 2019.

[25] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of Neural Engineering*, vol. 16, p. 031001, June 2019.

[26] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, pp. 5391–5420, Nov. 2017. arXiv: 1703.05051.

[27] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, p. 056013, July 2018.

[28] O. Avilov, S. Rimbert, A. Popov, and L. Bougrain, "Deep learning techniques to improve intraoperative awareness detection from electroencephalographic signals," *Conference proceedings: 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society. July 20-24, 2020, Montréal, Canada (forthcoming)*, 2020.

[29] T. Weiser, A. Haynes, G. Molina, S. Lipsitz, M. Esquivel, T. uribe Leitz, R. Fu, T. Azad, T. Chao, W. Berry, and A. Gawande, "Size and distribution of the global volume of surgery in 2012," *Bull World Health Organ*, vol. 94, no. 3, pp. 201–209, 2016.

[30] G. Mashour and M. Avidan, "Intraoperative awareness: controversies and non-controversies," *British Journal of Anaesthesia*, vol. 115, pp. 20–26, 2015.

[31] K. MacGregor, "A waking nightmare: how can we avoid accidental awareness during general anaesthesia?," *J Perioper Pract*, vol. 23, no. 9, pp. 185–90, 2013.

[32] D. Almeida, "Awake and unable to move: what can perioperative practitioners do to avoid accidental awareness under general anesthesia?," *J Perioper Pract*, vol. 25, no. 12, pp. 257–261, 2015.

[33] J. J. Pandit, J. Andrade, D. G. Bogod, J. M. Hitchman, W. R. Jonker, N. Lucas, J. H. Mackay, A. F. Nimmo, K. O'Connor, E. P. O'Sullivan, R. G. Paul, J. H. M. G. Palmer, F. Plaat, J. J. Radcliffe, M. R. J. Sury, H. E. Torevell, M. Wang, J. Hainsworth, T. M. Cook, Royal College of Anaesthetists, and Association of Anaesthetists of Great Britain and Ireland, "5th national audit project (NAP5) on accidental awareness during general anaesthesia: summary of main findings and risk factors," *British Journal of Anaesthesia*, vol. 113, pp. 549–559, 2014.

[34] R. Blacher, "On awakening paralyzed during surgery. a syndrome of traumatic neurosis," *Br J Anaesth*, no. 234, pp. 67–8, 1975.

[35] S. Tasbighou, M. Vogels, and A. Absalom, "Accidental awareness during general anaesthesia - a narrative review," *Anaesthesia*, vol. 73, no. 1, pp. 112–122, 2018.

[36] P. Myles, K. Leslie, J. McNeil, A. Forbes, and M. Chan, "Bispectral index monitoring to prevent awareness during anaesthesia: the b-aware randomised controlled trial," *The Lancet*, vol. 363, no. 9423, pp. 1757–1763, 2004.

[37] C. Kent and K. Domino, "Depth of anesthesia," *Curr opin Anaesthesiol*, vol. 22, no. 6, pp. 782–787, 2009.

[38] P. L. Purdon, A. Sampson, K. J. Pavone, and E. N. Brown, "Clinical Electroencephalography for Anesthesiologists Part I: Background and Basic Signatures," *Anesthesiology*, vol. 123, pp. 937–960, Oct. 2015.

[39] Y. Punjasawadwong, A. Phongchiewboon, and N. Bunchungmongkol, "Bispectral index for improving anaesthetic delivery and postoperative recovery," *Cochrane Database of Systematic Reviews*, no. 6, 2014.

[40] M. S. Avidan, L. Zhang, B. A. Burnside, K. J. Finkel, A. C. Searleman, J. A. Selvidge, L. Saager, M. S. Turner, S. Rao, M. Bottros, C. Hantler, E. Jacobsohn, and A. S. Evers, "Anesthesia awareness and the bispectral index," *New England Journal of Medicine*, vol. 358, no. 11, pp. 1097–1108, 2008. PMID: 18337600.

[41] M. S. Avidan, E. Jacobsohn, D. Glick, B. A. Burnside, L. Zhang, A. Villafranca, L. Karl, S. Kamal, B. Torres, M. O'Connor, A. S. Evers, S. Gradwohl, N. Lin, B. J. Palanca, and G. A. Mashour, "Prevention of intraoperative awareness in a high-risk surgical population.," *N Engl J Med*, vol. 365, pp. 591–600, Aug. 2011.

[42] M. S. Avidan and G. A. Mashour, "The incidence of intraoperative awareness in the UK: under the rate or under the radar?," *British Journal of Anaesthesia*, vol. 110, no. 4, pp. 494–497, 2013.

[43] P. Schuller, S. Newell, P. Strickland, and J. Barry, "Response of bispectral index to neuromuscular block in awake volunteers," *British Journal of Anaesthesia*, vol. 115, pp. 95–103, 2015.

[44] S. Rimbert, D. Schmartz, L. Bougrain, C. Meistelman, C. Baumann, and P. Guerci, "Motana: study protocol to investigate motor cerebral activity during a propofol sedation," *Trial*, vol. 20, p. 9, Sept. 2019.

[45] S. Rimbert, P. Guerci, N. Gayraud, C. Meistelman, and L. Bougrain, "Innovative brain-computer interface based on motor cortex activity to detect accidental awareness during general anesthesia," in *IEEE SMC 2019 - IEEE International Conference on Systems, Man, and Cybernetics*, (Bari, Italy), Oct. 2019.

[46] M. M. Ghoneim, R. I. Block, M. Haffarnan, and M. J. Mathews, "Awareness during anesthesia: risk factors, causes and sequelae: a review of reported cases in the literature," *Anesthesia and Analgesia*, vol. 108, no. 2, pp. 527–535, 2009.

[47] S. Salenius, A. Schnitzler, R. Salmelin, V. Jousmäki, and R. Hari, "Modulation of human cortical rolandic rhythms during natural sensorimotor tasks," *NeuroImage*, vol. 5, no. 3, pp. 221–228, 1997.

[48] C. Neuper and G. Pfurtscheller, "Event-related dynamics of cortical thythms: frequency-specific features and functional correlates," *Int J Psychophysiology*, vol. 43, pp. 41–58, 2001.

[49] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin Neurophysiol*, vol. 110, pp. 1842–57, Nov. 1999.

[50] B. E. Kilavik, M. Zaepffel, A. Brovelli, W. A. MacKay, and A. Riehle, "The ups and downs of beta oscillations in sensorimotor cortex.," *Exp Neurol*, vol. 245, pp. 15–26, July 2013.

[51] G. Pfurtscheller and A. Aranibar, "Event-related cortical desynchronization detected by power measurement of the scalp EEG," *Clinical Neurophysiology*, vol. 42, pp. 817–826, 1977.

[52] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, pp. 1123–1134, July 2001.

[53] O. Bai, Z. Mari, S. Vorbach, and M. Hallett, "Asymmetric spatiotemporal patterns of event-related desynchronization preceding voluntary sequential finger movements: a high-resolution EEG study," *Clinical Neurophysiology*, vol. 116, pp. 1213–1221, 2005.

[54] G. Pfurtscheller, A. Stancák, Jr, and C. Neuper, "Post-movement beta synchronization. a correlate of an idling motor area?," *Electroencephalogr Clin Neurophysiol*, vol. 98, pp. 281–93, Apr. 1996.

[55] S. Fok, R. Schwartz, M. Wronkiewicz, C. Holmes, J. Zhang, T. Somers, D. Bundy, and E. Leuthardt, "An eeg-based brain computer interface for rehabilitation and restoration of hand control following stroke using ipsilateral cortical physiology.," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2011, pp. 6277–6280, 2011.

[56] E. Formaggio, S. Storti, I. Galazzo, M. Gandolfi, C. Geroin, N. Smania, L. Speria, A. Waldner, A. Fiaschi, and P. Manganotti, "Modulation of event-related desynchronization in robot-assisted hand performance: brain oscillatory changes in active, passive and imagined movements," *J Neuroeng Rehabil*, pp. 10–24, 2013.

[57] A. Schnitzler, S. Salenius, R. Salmelin, V. Jousmaki, and R. Hari, "Involvement of primary motor cortex in motor imagery: a neuromagnetic study.," *Neuroimage*, vol. 6, pp. 201–208, Oct. 1997.

[58] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.

[59] F. Boi, T. Moraitis, V. De Feo, F. Diotalevi, C. Bartolozzi, G. Indiveri, and A. Vato, "A Bidirectional Brain-Machine Interface Featuring a Neuromorphic Hardware Decoder," *Frontiers in Neuroscience*, vol. 10, Dec. 2016.

[60] D. C. Irimia, R. Ortner, G. Krausz, C. Guger, and M. S. Poboroniuc, "BCI Application in Robotics Control," *IFAC Proceedings Volumes*, vol. 45, pp. 1869–1874, May 2012.

[61] M. Tariq, P. M. Trivailo, and M. Simic, "EEG-Based BCI Control Schemes for Lower-Limb Assistive-Robots," *Frontiers in Human Neuroscience*, vol. 12, 2018. Publisher: Frontiers.

[62] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Noninvasive Electroencephalogram Based Control of a Robotic Arm for Reach and Grasp Tasks," *Scientific Reports*, vol. 6, p. 38565, Dec. 2016. Number: 1 Publisher: Nature Publishing Group.

[63] R. H. Abiyev, N. Akkaya, E. Aytac, I. Günsel, and A. Çağman, "Brain-Computer Interface for Control of Wheelchair Using Fuzzy Neural Networks," Sept. 2016.

[64] J. Tang, Y. Liu, D. Hu, and Z. Zhou, "Towards BCI-actuated smart wheelchair system," *BioMedical Engineering OnLine*, vol. 17, Aug. 2018.

[65] R. Li, X. Zhang, Z. Lu, C. Liu, H. Li, W. Sheng, and R. Odekhe, "An Approach for Brain-Controlled Prostheses Based on a Facial Expression Paradigm," *Frontiers in Neuroscience*, vol. 12, 2018. Publisher: Frontiers.

[66] S. Yahud and N. A. Abu Osman, "Prosthetic Hand for the Brain-computer Interface System," in *3rd Kuala Lumpur International Conference on Biomedical Engineering 2006* (F. Ibrahim, N. A. A. Osman, J. Usman, and N. A. Kadri, eds.), IFMBE Proceedings, (Berlin, Heidelberg), pp. 643–646, Springer, 2007.

[67] J. V. Erp, F. Lotte, and M. Tangermann, "Brain-computer interfaces: Beyond medical applications," *Computer*, vol. 45, no. 4, p. 26–34, 2012.

[68] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain Computer Interfaces, a Review," *Sensors (Basel, Switzerland)*, vol. 12, pp. 1211–1279, Jan. 2012.

[69] W. J. Freeman, L. J. Rogers, M. D. Holmes, and D. L. Silbergeld, "Spatial spectral analysis of human electrocorticograms including the alpha and gamma bands," *Journal of Neuroscience Methods*, vol. 95, pp. 111–121, Feb. 2000.

[70] N. M. Gage and B. J. Baars, "Chapter 3 - Observing the Brain," in *Fundamentals of Cognitive Neuroscience (Second Edition)* (N. M. Gage and B. J. Baars, eds.), pp. 53–97, San Diego: Academic Press, Jan. 2018.

[71] A. Torricelli, D. Contini, A. Pifferi, M. Caffini, R. Re, L. Zucchelli, and L. Spinelli, "Time domain functional NIRS imaging for human brain mapping," *NeuroImage*, vol. 85, pp. 28–50, Jan. 2014.

[72] G. Buzsáki, *Rhythms of the brain*. Rhythms of the brain, New York, NY, US: Oxford University Press, 2006.

[73] J. Malmivuo, P. o. B. a. H. o. t. R. G. I. J. Malmivuo, R. Plonsey, and P. o. B. E. R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 1995.

[74] J. S. Kumar and P. Bhuvaneswari, "Analysis of Electroencephalography (EEG) Signals and Its Categorization–A Study," *Procedia Engineering*, vol. 38, pp. 2525–2536, Jan. 2012.

[75] S. Matsuoka, "Theta rhythms: State of consciousness," *Brain Topography*, vol. 3, pp. 203–208, Sept. 1990.

[76] E. Niedermeyer and F. L. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Oxford University Press, 2005.

[77] G. Pfurtscheller, "Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest," *Electroencephalography and Clinical Neurophysiology*, vol. 83, pp. 62–69, July 1992.

[78] E. M. Whitham, K. J. Pope, S. P. Fitzgibbon, T. Lewis, C. R. Clark, S. Loveless, M. Broberg, A. Wallace, D. DeLosAngeles, P. Lillie, A. Hardy, R. Fronsko, A. Pulbrook, and J. O. Willoughby, "Scalp electrical recording during paralysis: quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG," *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, vol. 118, pp. 1877–1888, Aug. 2007.

[79] G. Buzsaki, *Rhythms of the Brain*. Oxford ; New York: Oxford University Press, illustrated edition ed., June 2011.

[80] O. College, "Illustration from Anatomy & Physiology, Connexions Web site. http://cnx.org/content/col11496/1.6/, Jun 19, 2013.."

[81] Y. Yang and A. Raine, "Prefrontal Structural and Functional Brain Imaging findings in Antisocial, Violent, and Psychopathic Individuals: A Meta-Analysis," *Psychiatry research*, vol. 174, pp. 81–88, Nov. 2009.

[82] E. K. Miller, D. J. Freedman, and J. D. Wallis, "The prefrontal cortex: categories, concepts and cognition.," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 357, pp. 1123–1136, Aug. 2002.

[83] A. Riehle and E. Vaadia, *Motor Cortex in Voluntary Movements: A Distributed System for Distributed Functions*. CRC Press, Dec. 2004.

[84] j. rothwell, *Control of Human Voluntary Movement*. Springer Science & Business Media, Dec. 2012.

[85] D. A. Rosenbaum, *Human Motor Control*. Elsevier, June 2014.

[86] B. S. Matthew Levy, Bruce Koeppen, *Berne and Levy Principles of Physiology*. Mosby, 2006.

[87] G. H. Klem, H. O. Lüders, H. H. Jasper, and C. Elger, "The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology," *Electroencephalography and Clinical Neurophysiology. Supplement*, vol. 52, pp. 3–6, 1999.

[88] L. Avanzino, A. Giannini, A. Tacchino, E. Pelosin, P. Ruggeri, and M. Bove, "Motor imagery influences the execution of repetitive finger opposition movements," *Neuroscience Letters*, vol. 466, no. 1, pp. 11–15, 2009.

[89] J. Decety and J. A. Stevens, "Action representation and its role in social interaction," in *Handbook of imagination and mental simulation*, pp. 3–20, New York, NY, US: Psychology Press, 2009.

[90] M. Jeannerod, "Mental imagery in the motor context," *Neuropsychologia*, vol. 33, pp. 1419–32, Nov. 1995.

[91] M. Lotze and U. Halsband, "Motor imagery," *J Physiol Paris*, vol. 99, pp. 386–95, June 2006.

[92] C. Neuper, R. Scherer, M. Reiner, and G. Pfurtscheller, "Imagery of motor actions: Differential effects of kinesthetic and visual–motor mode of imagery in single-trial EEG," *Cognitive Brain Research*, vol. 25, no. 3, pp. 668–677, 2005.

[93] A. Guillot, C. Collet, V. A. Nguyen, F. Malouin, C. Richards, and J. Doyon, "Brain activity during visual versus kinesthetic imagery: an fMRI study," *Human Brain Mapping*, vol. 30, pp. 2157–2172, July 2009.

[94] A. J. Butler and S. J. Page, "Mental practice with motor imagery: evidence for motor recovery and cortical reorganization after stroke.," *Arch Phys Med Rehabil.*, vol. 87, pp. S2–11, Dec. 2006.

[95] Y. Blokland, J. Farquhar, J. Lerou, J. Mourisse, G. J. Scheffer, G.-J. van Geffen, L. Spyrou, and J. Bruhn, "Decoding motor responses from the eeg during altered states of consciousness induced by propofol," *Journal of Neural Engineering*, vol. 13, no. 2, p. 026014, 2016.

[96] M. Liu, T. Fujiwara, K. Shindo, Y. Kasashima_Shindo, Y. Otaka, T. Tsuji, and J. Ushiba, "Newer challenges to restore hemiparetic upper extremity after stroke: HANDS therapy and BMI neurorehabilitation," *Hong Kong Physiotherapy Journal*, vol. 30, pp. 83–92, Dec. 2012.

[97] C. Vidaurre and B. Blankertz, "Towards a Cure for BCI Illiteracy," *Brain Topography*, vol. 23, no. 2, pp. 194–198, 2010.

[98] M. C. Thompson, "Critiquing the Concept of BCI Illiteracy," *Science and Engineering Ethics*, vol. 25, pp. 1217–1233, Aug. 2019.

[99] G. Pfurtscheller and A. Berghold, "Patterns of cortical activation during planning of voluntary movement," *Electroencephalography and Clinical Neurophysiology*, vol. 72, no. 3, p. 250–258, 1989.

[100] W. Klimesch, G. Pfurtscheller, W. Mohl, and H. Schimke, "Event-related desynchronization, erd-mapping and hemispheric differences for words and numbers," *International Journal of Psychophysiology*, vol. 8, no. 3, p. 297–308, 1990.

[101] J. Wolpaw and E. W. Wolpaw, eds., *Brain-Computer Interfaces: Principles and Practice*. Oxford university press, 2012.

[102] T. Lin, T. Smith, C. Pinnock, and C. Mowatt, *Fundamentals of Anaesthesia*. Cambridge University Press, Dec. 2016.

[103] P. G. Barash, *Clinical Anesthesia*. Lippincott Williams & Wilkins, 2009.

[104] T. C. Gray, J. F. Nunn, and J. E. Utting, *General Anaesthesia*. Butterworths, 1980.

[105] E. B. Cooper and J. B. Cooper, "Electrical treatment of coma via the median nerve," *Acta Neurochirurgica. Supplement*, vol. 87, pp. 7–10, 2003.

[106] X. Wu, C. Zhang, J. Feng, Q. Mao, G. Gao, and J. Jiang, "Right median nerve electrical stimulation for acute traumatic coma (the Asia Coma Electrical Stimulation trial): study protocol for a randomised controlled trial," *Trials*, vol. 18, July 2017.

[107] A. Maharjan, M. Peng, and Y. O. Cakmak, "Non-invasive High Frequency Median Nerve Stimulation Effectively Suppresses Olfactory Intensity Perception in Healthy Males," *Frontiers in Human Neuroscience*, vol. 12, 2019.

[108] F. P. Ron Kohavi, "Special issue on applications of machine learning and the knowledge discovery process," *Glossary of Terms*, no. 30, pp. 271–274, 1998.

[109] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalography and Clinical Neurophysiology*, vol. 103, pp. 386–394, Sept. 1997.

[110] M. J. Alhaddad, M. Kamel, H. Malibary, K. Thabit, F. Dahlwi, and A. Hadi, "P300 Speller Efficiency with

Common Average Reference," in *Autonomous and Intelligent Systems* (M. Kamel, F. Karray, and H. Hagras, eds.), Lecture Notes in Computer Science, (Berlin, Heidelberg), pp. 234–241, Springer, 2012.

[111] S. Syam, H. Lakany, R. Ahmad, and B. Conway, "Comparing Common Average Referencing to Laplacian Referencing in Detecting Imagination and Intention of Movement for Brain Computer Interface," *MATEC Web of Conferences*, vol. 140, p. 01028, Jan. 2017.

[112] T. Solis-Escalante, G. Müller-Putz, and G. Pfurtscheller, "Overt foot movement detection in one single Laplacian EEG derivation," *Journal of neuroscience methods*, vol. 175, pp. 148–53, Sept. 2008.

[113] J. Lu, D. J. McFarland, and J. R. Wolpaw, "Adaptive Laplacian filtering for sensorimotor rhythm-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 10, p. 016002, Feb. 2013.

[114] S. Butterworth, "On the theory of filter amplifiers," *In Wireless Engineer (also called Experimental Wireless and the Wireless Engineer)*, vol. 7, p. 536–541, 1930.

[115] G. Bianchi and R. Sorrentino, "Electronic filter simulation and design," *McGraw-Hill Professional*, p. 17–20, 2007.

[116] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of Artifacts from EEG Signals: A Review," *Sensors*, vol. 19, p. 987, Feb. 2019.

[117] J. Urigüen and B. Zapirain, "EEG artifact removal – State-of-the-art and guidelines," *Journal of neural engineering*, vol. 12, p. 031001, Apr. 2015.

[118] X. Chen, A. Liu, J. Chiang, Z. J. Wang, M. J. McKeown, and R. K. Ward, "Removing Muscle Artifacts From EEG Data: Multichannel or Single-Channel Techniques?," *IEEE Sensors Journal*, vol. 16, pp. 1986–1997, Apr. 2016. Conference Name: IEEE Sensors Journal.

[119] S. Devuyst, T. Dutoit, P. Stenuit, M. Kerkhofs, and E. Stanus, "Removal of ECG artifacts from EEG using a modified independent component analysis approach," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5204–5207, Aug. 2008. ISSN: 1558-4615.

[120] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.

[121] A. Popov, *Theory of signals: a textbook for applicants for a bachelor's degree in the speciality Micro- and nanosystem technology*. Kyiv, Ukraine: Igor Sikorsky Kyiv Polytechnic Institute, 2019.

[122] I. Daubechies, *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Jan. 1992.

[123] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. Amsterdam ; Boston: Academic Press, 3rd edition ed., Dec. 2008.

[124] R. Quiroga and H. Garcia, "Single-trial event-related potentials with wavelet denoising," *Clinical Neurophysiology*, vol. 114, no. 2, pp. 376–390, 2003.

[125] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, p. 441–446, 2000.

[126] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background eeg," *Brain Topography*, vol. 2, no. 4, p. 275–284, 1990.

[127] M. Grosse-Wentrup and M. Buss, "Multi-class common spatial patterns and information theoretic feature extraction," *IEEE Transactions on Biomedical Engineering*, vol. 8, no. 55, 2008.

[128] F. Lotte, "A tutorial on EEG signal-processing techniques for mental-state recognition in brain–computer interfaces," *Guide to Brain-Computer Music Interfacing*, p. 133–161, 2014.

[129] M. Grosse-Wentrup, C. Liefhold, K. Gramann, and M. Buss, "Beamforming in noninvasive brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, p. 1209–1219, 2009.

[130] B. Reuderink and M. Poel, *Robustness of the Common Spatial Patterns algorithm in the BCI-pipeline*. No. DTR08-9/TR-CTIT-08-52 in CTIT Technical Report Series, Netherlands: Centre for Telematics and Information Technology (CTIT), 7 2008.

[131] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b," *Frontiers in neuroscience*, vol. 6, p. 39, 2012.

[132] M. Congedo, B. Afsari, A. Barachant, and M. Moakher, "Approximate Joint Diagonalization and Geometric Mean of Symmetric Positive Definite Matrices," *PLOS ONE*, vol. 10, p. e0121423, Apr. 2015. Publisher: Public Library of Science.

[133] A. Goh and R. Vidal, "Unsupervised Riemannian Clustering of Probability Density Functions," *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Sept. 2008. Pages: 392.

[134] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol. 4, no. 3, pp. 155–174, 2017. Publisher: Taylor & Francis.

[135] T. H. Gareth James, Daniela Witten and R. Tibshirani, "Statistical learning," *Springer Texts in Statistics An Introduction to Statistical Learning*, p. 15–57, 2013.

[136] M. L. Bermingham, R. Pong-Wong, A. Spiliopoulou, C. Hayward, I. Rudan, H. Campbell, A. F. Wright, J. F. Wilson, F. Agakov, P. Navarro, and et al., "Application of high-dimensional feature selection: evaluation for genomic prediction in man," *Scientific Reports*, vol. 5, no. 1, 2015.

[137] C. C. Aggarwal, *Outlier Analysis*. Springer International Publishing, 2 ed., 2017.

[138] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," *ACM SIGMOD Record*, vol. 29, pp. 93–104, May 2000.

[139] F. T. Liu, K. Ting, and Z.-H. Zhou, "Isolation Forest," *ICDM '08: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, Jan. 2009. Pages: 422.

[140] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the Support of a High-Dimensional Distribution," *Neural Computation*, vol. 13, pp. 1443–1471, July 2001.

[141] P. Rousseeuw and K. Driessen, "A Fast Algorithm for the Minimum Covariance Determinant Estimator," *Technometrics*, vol. 41, pp. 212–223, Aug. 1999.

[142] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165–175, 1989.

[143] J. R. Quinlan, "Induction of decision trees," *MACH. LEARN*, vol. 1, pp. 81–106, 1986.

[144] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001. Publisher: Institute of Mathematical Statistics.

[145] L. Mason, J. Baxter, P. Bartlett, and M. Frean, "Boosting Algorithms as Gradient Descent in Function Space," *NIPS'99: Proceedings of the 12th International Conference on Neural Information Processing Systems*, 1999.

[146] M. R. Fréchet, "Les éléments aléatoires de nature quelconque dans un espace distancié," *Annales de l'institut Henri Poincaré*, vol. 10, no. 4, pp. 215–310, 1948.

[147] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945. Publisher: [International Biometric Society, Wiley].

[148] S. Siegel, *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, 1956.

[149] B. Manly, *Randomization, Bootstrap and Monte Carlo Methods in Biology*, ch. The Generation of Random Permutations. Champman and Hall/CRC, Taylor and Francis Group, 2006.

[150] C. Lindig-Leon, S. Rimbert, O. Avilov, and L. Bougrain, "Scalp EEG activity during simple and combined motor imageries to control a robotic ARM," in *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, pp. 322–327, May 2017.

[151] C. Bishop, *Pattern Recognition and Machine Learning*. Information Science and Statistics, New York: Springer-Verlag, 2006.

[152] M. Hassoun, *Fundamentals of Artificial Neural Networks*. Cambridge, Mass.: A Bradford Book, Jan. 2003.

[153] A. Zell, N. Mache, R. Hübner, G. Mamier, M. Vogt, M. Schmalzl, and K.-U. Herrmann, "SNNS (Stuttgart Neural Network Simulator)," in *Neural Network Simulation Environments* (J. Skrzypek, ed.), The Kluwer International Series in Engineering and Computer Science, pp. 165–186, Boston, MA: Springer US, 1994.

[154] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. D. Jesús, *Neural Network Design*. Wrocław: Martin Hagan, 2nd edition ed., Sept. 2014.

[155] D. J. Livingstone, ed., *Artificial Neural Networks: Methods and Applications*. Methods in Molecular Biology, Humana Press, 2009.

[156] J. L. G. Rosa, ed., *Artificial Neural Networks - Models and Applications*. InTechOpen, Oct. 2016.

[157] Y. Bengio, *Learning Deep Architectures for AI*. Boston: Now Publishers Inc, Oct. 2009.

[158] V. Nair and G. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair," *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, vol. 27, June 2010.

[159] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, p. 3, 2013. Issue: 1.

[160] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)," *arXiv:1511.07289 [cs]*, Feb. 2016. arXiv: 1511.07289.

[161] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-Normalizing Neural Networks," *arXiv:1706.02515 [cs, stat]*, Sept. 2017. arXiv: 1706.02515.

[162] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, "Maxout Networks," *arXiv:1302.4389 [cs, stat]*, Sept. 2013. arXiv: 1302.4389.

[163] S. Behnke, *Hierarchical Neural Networks for Image Interpretation*. Berlin ; New York: Springer, 2003rd edition ed., Aug. 2003.

[164] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, Dec. 1989.

[165] J. Snyman, *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*. Springer Science & Business Media, Mar. 2005.

[166] H. Wu, "Global stability analysis of a general class of discontinuous neural networks with linear growth activation functions," *Inf. Sci.*, vol. 179, pp. 3432–3441, Sept. 2009.

[167] D. Sussillo and L. F. Abbott, "Random Walk Initialization for Training Very Deep Feedforward Networks," *arXiv:1412.6558 [cs, stat]*, Feb. 2015. arXiv: 1412.6558.

[168] Y. LeCun, L. Bottou, G. Orr, and K. Müller, "Efficient BackProp," in *Neural Networks: Tricks of the Trade*, 1998.

[169] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 315–323, JMLR Workshop and Conference Proceedings, June 2011. ISSN: 1938-7228.

[170] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for Activation Functions," *arXiv:1710.05941 [cs]*, Oct. 2017. arXiv: 1710.05941.

[171] L. Bottou, "Stochastic Learning," in *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures* (O. Bousquet, U. von Luxburg, and G. Rätsch, eds.), Lecture Notes in Computer Science, pp. 146–168, Berlin, Heidelberg: Springer, 2004.

[172] L. Bottou and O. Bousquet, "The Tradeoffs of Large Scale Learning.," *Optimization for Machine Learning*, vol. 20, Jan. 2007. Journal Abbreviation: Optimization for Machine Learning Publication Title: Optimization for Machine Learning.

[173] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv:1609.04747 [cs]*, June 2017. arXiv: 1609.04747.

[174] J. Duchi, E. Hazan, and Y. Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, July 2011.

[175] M. D. Zeiler, "ADADELTA: An Adaptive Learning Rate Method," *arXiv:1212.5701 [cs]*, Dec. 2012. arXiv: 1212.5701.

[176] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Dec. 2014. arXiv: 1412.6980.

[177] T. Dozat, "Incorporating nesterov momentum into adam," *ICLR Workshop*, 2016.

[178] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, pp. 541–551, Dec. 1989.

[179] M. A. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2015.

[180] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, Massachusetts: The MIT Press, illustrated edition ed., Nov. 2016.

[181] A. Taspinar, "Using convolutional neural networks to detect features in satellite images," Nov 2018.

[182] A. Golnari, "Convolutional neural network in gpu with opencl," Dec 2017.

[183] C.-F. Wang, "A basic introduction to separable convolutions," Aug 2018.

[184] R. Venkatesan and B. Li, *Convolutional Neural Networks in Visual Computing: A Concise Guide*. Boca Raton London New York: CRC Press, 1st edition ed., Sept. 2017.

[185] E. Bendersky, "Depthwise separable convolutions for machine learning," Apr 2018.

[186] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *arXiv:1610.02357 [cs]*, Apr. 2017. arXiv: 1610.02357.

[187] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv:1704.04861 [cs]*, Apr. 2017. arXiv: 1704.04861.

[188] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.

[189] K. Yamaguchi, K. Sakamoto, T. Akabane, and Y. Fujimoto, "A neural network for speaker-independent

isolated word recognition," *First International Conference on Spoken Language Processing (ICSLP 90)*, nov 1990.

[190] L. Prechelt, "Early Stopping — But When?," in *Neural Networks: Tricks of the Trade: Second Edition* (G. Montavon, G. B. Orr, and K.-R. Müller, eds.), Lecture Notes in Computer Science, pp. 53–67, Berlin, Heidelberg: Springer, 2012.

[191] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, Jan. 2014.

[192] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv:1502.03167 [cs]*, Mar. 2015. arXiv: 1502.03167.

[193] D. Warde-Farley, I. J. Goodfellow, A. Courville, and Y. Bengio, "An empirical analysis of dropout in piecewise linear networks," *arXiv:1312.6197 [cs, stat]*, Jan. 2014. arXiv: 1312.6197.

[194] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[195] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.

[196] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," *arXiv:1709.01507 [cs]*, Sept. 2017. arXiv: 1709.01507.

[197] B. Blankertz, K.-R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, and N. Birbaumer, "The BCI Competition 2003: progress and perspectives in detection and discrimination of EEG single trials," *IEEE transactions on bio-medical engineering*, vol. 51, pp. 1044–1051, June 2004.

[198] B. Blankertz, K.-R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. d. R. Millán, M. Schröder, and N. Birbaumer, "The BCI competition. III: Validating alternative approaches to actual BCI problems," *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 14, pp. 153–159, June 2006.

[199] P. Sajda, A. Gerson, K.-R. Müller, B. Blankertz, and L. Parra, "A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 11, pp. 184–185, July 2003.

[200] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the bci competition iv," *Front Neurosci*, vol. 6, p. 55, 2012.

[201] G. Schalk, D. Mcfarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "BCI2000: a general-purpose Brain-Computer Interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, p. 1034, July 2004.

[202] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13). Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[203] P. Ofner, A. Schwarz, J. Pereira, and G. R. Müller-Putz, "Upper limb movements can be decoded from the time-domain of low-frequency EEG," *PLOS ONE*, vol. 12, p. e0182578, Aug. 2017. Publisher: Public Library of Science.

[204] K. Ha and J. Jeong, "Decoding Two-Class Motor Imagery EEG with Capsule Networks," in *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 1–4, Feb. 2019.

[205] K.-W. Ha and J.-W. Jeong, "Motor Imagery EEG Classification Using Capsule Networks," *Sensors (Basel, Switzerland)*, vol. 19, June 2019.

[206] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *Journal of Neural Engineering*, vol. 14, p. 016003, Feb. 2017.

[207] A. Hassanpour, M. Moradikia, H. Adeli, S. R. Khayami, and P. Shamsinejadbabaki, "A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals," *Expert Systems*, vol. 36, no. 6, p. e12494, 2019. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/exsy.12494.

[208] P. Saha, S. Fels, and M. Abdul-Mageed, "Deep Learning the EEG Manifold for Phonological Categorization from Active Thoughts," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2762–2766, May 2019.

[209] X. Zhang, L. Yao, Q. Z. Sheng, S. S. Kanhere, T. Gu, and D. Zhang, "Converting Your Thoughts to Texts: Enabling Brain Typing via Deep Feature Learning of EEG Signals," in *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–10, Mar. 2018.

[210] M. Dai, D. Zheng, R. Na, S. Wang, and S. Zhang, "EEG Classification of Motor Imagery Using a Novel Deep Learning Framework," *Sensors*, vol. 19, p. 551, Jan. 2019.

[211] T.-j. Luo, C.-l. Zhou, and F. Chao, "Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network," *BMC Bioinformatics*, vol. 19, p. 344, Dec. 2018.

[212] Q. She, B. Hu, Z. Luo, T. Nguyen, and Y. Zhang, "A hierarchical semi-supervised extreme learning machine method for EEG recognition," *Medical & Biological Engineering & Computing*, vol. 57, pp. 147–157, Jan. 2019.

[213] Y. Zhao, S. Yao, S. Hu, S. Chang, R. Ganti, M. Srivatsa, S. Li, and T. Abdelzaher, "On the improvement of classifying EEG recordings using neural networks," in *2017 IEEE International Conference on Big Data (Big Data)*, pp. 1709–1711, Dec. 2017.

[214] A. Kar, S. Bera, S. P. K. Karri, S. Ghosh, M. Mahadevappa, and D. Sheet, "A Deep Convolutional Neural Network Based Classification Of Multi-Class Motor Imagery With Improved Generalization," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5085–5088, July 2018.

[215] E. Jeon, W. Ko, and H. Suk, "Domain Adaptation with Source Selection for Motor-Imagery based BCI," in *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–4, Feb. 2019.

[216] Y. Huang, J. Chang, L. Chen, and Y. Chen, "Deep Neural Network with Attention Mechanism for Classification of Motor Imagery EEG," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 1130–1133, Mar. 2019.

[217] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel Weighted Feature Fusion Using Convolutional Neural Networks for EEG Motor Imagery Classification," *IEEE Access*, vol. 7, pp. 18940–18950, 2019.

[218] D. Zhao, F. Tang, B. Si, and X. Feng, "Learning joint space–time–frequency features for EEG decoding on small labeled data," *Neural Networks*, vol. 114, pp. 67–77, June 2019.

[219] H. Wu, Y. Niu, F. Li, Y. Li, B. Fu, G. Shi, and M. Dong, "A Parallel Multiscale Filter Bank Convolutional Neural Networks for Motor Imagery EEG Classification," *Frontiers in Neuroscience*, vol. 13, p. 1275, Nov. 2019.

[220] G. Dai, J. Zhou, J. Huang, and N. Wang, "HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification," *Journal of Neural Engineering*, vol. 17, p. 016025, Jan. 2020. Publisher: IOP Publishing.

[221] Y. Li, X. Zhang, B. Zhang, M. Lei, W. Cui, and Y. Guo, "A Channel-Projection Mixed-Scale Convolutional Neural Network for Motor Imagery EEG Decoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, pp. 1170–1180, June 2019.

[222] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. Shamim Hossain, "Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion," *Future Generation Computer Systems*, vol. 101, pp. 542–554, Dec. 2019.

[223] Y. Wu, T. H. Huang, C. Y. Lin, S. J. Tsai, and P. Wang, "Classification of EEG Motor Imagery Using Support Vector Machine and Convolutional Neural Network," in *2018 International Automatic Control Conference (CACS)*, pp. 1–4, Nov. 2018.

[224] Z. Tayeb, J. Fedjaev, N. Ghaboosi, C. Richter, L. Everding, X. Qu, Y. Wu, G. Cheng, and J. Conradt, "Validating Deep Neural Networks for Online Decoding of Motor Imagery Movements from EEG Signals," *Sensors*, vol. 19, p. 210, Jan. 2019.

[225] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, and Y. Liu, "Making Sense of Spatio-Temporal Preserving Representations for EEG-Based Human Intention Recognition," *IEEE Transactions on Cybernetics*, vol. PP, Mar. 2019.

[226] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to MI-EEG signal classification for BCIs," *Expert Systems with Applications*, vol. 114, pp. 532–542, Dec. 2018.

[227] W. Ko, J. Yoon, E. Kang, E. Jun, J. Choi, and H. Suk, "Deep recurrent spatio-temporal neural network for motor imagery based BCI," in *2018 6th International Conference on Brain-Computer Interface (BCI)*, pp. 1–3, Jan. 2018.

[228] X. Zhang, L. Yao, C. Huang, Q. Z. Sheng, and X. Wang, "Intent Recognition in Smart Living Through Deep Recurrent Neural Networks," *arXiv:1702.06830 [cs]*, Aug. 2017. arXiv: 1702.06830.

[229] X. Ma, S. Qiu, C. Du, J. Xing, and H. He, "Improving EEG-Based Motor Imagery Classification via Spatial and Temporal Recurrent Neural Networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1903–1906, July 2018.

[230] X.-L. Tang, W.-C. Ma, D.-S. Kong, and W. Li, "Semisupervised Deep Stacking Network with Adaptive Learning Rate Strategy for Motor Imagery EEG Recognition," *Neural Computation*, vol. 31, pp. 919–942, May 2019.

[231] Z. Zhang, F. Duan, J. Sole-Casals, J. Dinares-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A Novel Deep Learning Approach With Data Augmentation to Classify Motor Imagery Signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019.

[232] B. Xu, L. Zhang, A. Song, C. Wu, W. Li, D. Zhang, G. Xu, H. Li, and H. Zeng, "Wavelet Transform Time-Frequency Image and Convolutional Network-Based Motor Imagery EEG Classification," *IEEE Access*, vol. 7, pp. 6084–6093, 2019.

[233] M.-A. Li, M. Zhang, and Y.-J. Sun, "A novel motor imagery EEG recognition method based on deep learning," in *Proceedings of the 2016 International Forum on Management, Education and Information Technology Application*, pp. 728–733, Atlantis Press, Jan. 2016. ISSN: 2352-5398.

[234] J. Yang, S. Yao, and J. Wang, "Deep Fusion Feature Learning Network for MI-EEG Classification," *IEEE Access*, vol. 6, pp. 79050–79059, 2018.

[235] S. Taheri, M. Ezoji, and S. M. Sakhaei, "Convolutional neural network based features for motor imagery EEG signals classification in brain–computer interface system," *SN Applied Sciences*, vol. 2, p. 555, Apr. 2020.

[236] C. J. Ortiz-Echeverri, S. Salazar-Colores, J. Rodríguez-Reséndiz, and R. A. Gómez-Loenzo, "A New Approach for Motor Imagery Classification Based on Sorted Blind Source Separation, Continuous Wavelet Transform, and Convolutional Neural Network," *Sensors*, vol. 19, p. 4541, Oct. 2019.

[237] G. Gao, L. Shang, K. Xiong, J. Fang, C. Zhang, and X. Gu, "EEG Classification Based on Sparse Representation and Deep Learning," *NeuroQuantology*, vol. 16, pp. 789–795, June 2018.

[238] S. Chaudhary, S. Taran, V. Bajaj, and A. Sengur, "Convolutional Neural Network Based Approach Towards Motor Imagery Tasks EEG Signals Classification," *IEEE Sensors Journal*, vol. 19, pp. 4494–4500, June 2019.

[239] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, "A Multi-Branch 3D Convolutional Neural Network for EEG-Based Motor Imagery Classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, pp. 2164–2177, Oct. 2019.

[240] D. Li, J. Wang, J. Xu, and X. Fang, "Densely Feature Fusion Based on Convolutional Neural Networks for Motor Imagery EEG Classification," *IEEE Access*, vol. PP, pp. 1–1, Sept. 2019.

[241] R. Zhang, Q. Zong, L. Dou, and X. Zhao, "A novel hybrid deep learning scheme for four-class motor imagery classification," *Journal of Neural Engineering*, vol. 16, p. 066004, Oct. 2019. Publisher: IOP Publishing.

[242] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "LSTM-Based EEG Classification in Motor Imagery Tasks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, pp. 2086–2095, Nov. 2018.

[243] S. Sakhavi, C. Guan, and S. Yan, "Learning Temporal Information for Brain-Computer Interface Using Convolutional Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, pp. 5619–5629, Nov. 2018.

[244] S. Sakhavi and C. Guan, "Convolutional neural network-based transfer learning and knowledge distillation using multi-subject data in motor imagery BCI," in *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, (Shanghai, China), pp. 588–591, IEEE, May 2017.

[245] S. Sakhavi, C. Guan, and S. Yan, "Parallel convolutional-linear neural network for motor imagery classification," in *2015 23rd European Signal Processing Conference (EUSIPCO)*, (Nice), pp. 2736–2740, IEEE, Aug. 2015.

[246] Huijuan Yang, S. Sakhavi, Kai Keng Ang, and Cuntai Guan, "On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Milan), pp. 2620–2623, IEEE, Aug. 2015.

[247] W. Abbas and N. A. Khan, "DeepMI: Deep Learning for Multiclass Motor Imagery Classification," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 219–222, July 2018.

[248] I. Majidov and T. Whangbo, "Efficient Classification of Motor Imagery Electroencephalography Signals Using Deep Learning Methods," *Sensors (Basel, Switzerland)*, vol. 19, Apr. 2019.

[249] B. E. Olivas-Padilla and M. I. Chacon-Murguia, "Classification of multiple motor imagery using deep convolutional neural networks and spatial filters," *Applied Soft Computing*, vol. 75, pp. 461–472, Feb. 2019.

[250] F. Li, F. He, F. Wang, D. Zhang, Y. Xia, and X. Li, "A Novel Simplified Convolutional Neural Network Classification Algorithm of Motor Imagery EEG Signals Based on Deep Learning," *Applied Sciences*, vol. 10, p. 1605, Feb. 2020.

[251] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, and W. Che, "A Deep Transfer Convolutional Neural Network Framework for EEG Signal Classification," *IEEE Access*, vol. 7, pp. 112767–112776, 2019.

[252] N. Lu, T. Li, X. Ren, and H. Miao, "A Deep Learning Scheme for Motor Imagery Classification based on Restricted Boltzmann Machines," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 566–576, June 2017.

[253] H. K. Lee and Y.-S. Choi, "Application of Continuous Wavelet Transform and Convolutional Neural Network in Decoding Motor Imagery Brain-Computer Interface," *Entropy*, vol. 21, p. 1199, Dec. 2019.

[254] X. Tang, W. Li, X. Li, W. Ma, and X. Dang, "Motor imagery EEG recognition based on conditional optimization empirical mode decomposition and multi-scale convolutional neural network," *Expert Systems with Applications*, vol. 149, p. 113285, July 2020.

[255] X. Zhu, P. Li, C. Li, D. Yao, R. Zhang, and P. Xu, "Separated channel convolutional neural network to realize the training free motor imagery BCI systems," *Biomedical Signal Processing and Control*, vol. 49, pp. 396–403, Mar. 2019.

[256] T. Dharamsi, P. Das, T. Pedapati, G. Bramble, V. Muthusamy, H. Samulowitz, K. R. Varshney, Y. Rajamanickam, J. Thomas, and J. Dauwels, "Neurology-as-a-Service for the Developing World," *arXiv:1711.06195 [cs, stat]*, Nov. 2017. arXiv: 1711.06195.

[257] M. Alomari, A. Samaha, and K. AlKamha, "Automated Classification of L/R Hand Movement EEG Signals using Advanced Feature Extraction and Machine Learning," *International Journal of Advanced Computer Science and Applications*, vol. 4, pp. 207–212, July 2013.

[258] T. C. Major and J. M. Conrad, "The effects of pre-filtering and individualizing components for electroencephalography neural network classification," in *SoutheastCon 2017*, (Concord, NC, USA), pp. 1–6, IEEE, Mar. 2017.

[259] D. Zhang and L. Yao, "Cascade and Parallel Convolutional Recurrent Neural Networks on EEG-based Intention Recognition for Brain Computer Interface," *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, Sept. 2017.

[260] X. Zhang, L. Yao, D. Zhang, X. Wang, Q. Z. Sheng, and T. Gu, "Multi-Person Brain Activity Recognition via Comprehensive EEG Signal Analysis," *arXiv:1709.09077 [cs]*, Sept. 2017. arXiv: 1709.09077.

[261] S. Lee, H. Kim, J. Jeong, I. Wang, S. Lee, and D. Kim, "Recurrent convolutional neural network model based on temporal and spatial feature for motor imagery classification," in *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–4, Feb. 2019.

[262] D. Cheng, Y. Liu, and L. Zhang, "Exploring Motor Imagery Eeg Patterns for Stroke Patients with Deep Neural Networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2561–2565, Apr. 2018.

[263] Z. Wang, L. Cao, Z. Zhang, X. Gong, Y. Sun, and H. Wang, "Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition," *Concurrency and Computation: Practice and Experience*, vol. 30, Jan. 2018.

[264] Y. Chu, X. Zhao, Y. Zou, W. Xu, J. Han, and Y. Zhao, "A Decoding Scheme for Incomplete Motor Imagery EEG With Deep Belief Network," *Frontiers in Neuroscience*, vol. 12, p. 680, Sept. 2018.

[265] O.-Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, "Subject-Independent Brain–Computer Interfaces Based on Deep Convolutional Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, pp. 3839–3852, Oct. 2020.

[266] R. Alazrai, M. Abuhijleh, H. Alwanni, and M. I. Daoud, "A Deep Learning Framework for Decoding Motor Imagery Tasks of the Same Hand Using EEG Signals," *IEEE Access*, vol. 7, pp. 109612–109627, 2019.

[267] X. Ma, S. Qiu, W. Wei, S. Wang, and H. He, "Deep Channel-Correlation Network for Motor Imagery Decoding From the Same Limb," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, pp. 297–306, Jan. 2020.

[268] X. An, D. Kuang, X. Guo, Y. Zhao, and L. He, "A Deep Learning Method for Classification of EEG Data Based on Motor Imagery," in *Intelligent Computing in Bioinformatics* (D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, D. Terzopoulos, D. Tygar, G. Weikum, D.-S. Huang, K. Han, and M. Gromiha, eds.), vol. 8590, pp. 203–210, Cham: Springer International Publishing, 2014. Series Title: Lecture Notes in Computer Science.

[269] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11–18, Feb. 2017.

[270] S. A. C. Yohanandan, I. Kiral-Kornek, J. Tang, B. S. Mshford, U. Asif, and S. Harrer, "A Robust Low-Cost EEG Motor Imagery-Based Brain-Computer Interface," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5089–5092, July 2018.

[271] J. Zhang, C. Yan, and X. Gong, "Deep convolutional neural network for decoding motor imagery based brain computer interface," in *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1–5, Oct. 2017.

[272] N. Mammone, C. Ieracitano, and F. C. Morabito, "A deep CNN approach to decode motor preparation of upper limbs from time–frequency maps of EEG signals at source level," *Neural Networks*, vol. 124, pp. 357–372, Apr. 2020.

[273] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds.), vol. 8689, pp. 818–833, Cham: Springer International Publishing, 2014. Series Title: Lecture Notes in Computer Science.

[274] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[275] D. Balouek, A. Carpen Amarie, G. Charrier, F. Desprez, E. Jeannot, E. Jeanvoine, A. Lèbre, D. Margery, N. Niclausse, L. Nussbaum, O. Richard, C. Pérez, F. Quesnel, C. Rohr, and L. Sarzyniec, "Adding virtualization capabilities to the Grid'5000 testbed," in *Cloud Computing and Services Science* (I. I. Ivanov, M. van Sinderen, F. Leymann, and T. Shan, eds.), vol. 367 of *Communications in Computer and Information Science*, pp. 3–20, Springer International Publishing, 2013.

[276] J. Shin, A. von Lühmann, B. Blankertz, D.-W. Kim, J. Jeong, h.-j. Hwang, and K.-R. Müller, "Open Access Dataset for EEG+NIRS Single-Trial Classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 1735–1745, Oct. 2017.

[277] W. M. Association, "World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects.," *Journal of Postgraduate Medicine*, vol. 48, p. 206, Jan. 2002. Publisher: Medknow Publications.

[278] C. Lindig León, *Classification multilabels à partir de signaux EEG d'imaginations motrices combinées : application au contrôle 3D d'un bras robotique*. These de doctorat, Université de Lorraine, Jan. 2017.

[279] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer, "Openvibe: An open-source software platform to design, test and use brain-computer interfaces in real and virtual environments," *Presence : teleoperators and virtual environments*, vol. 10, no. 1, 2010.

[280] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[281] D. Kumbhare, L. Robinson, and R. Buschbacher, *Buschbacher's Manual of Nerve Conduction Studies, 3rd Edition*, ch. Mediane nerve to the abductor pellicis brevis, p. 10. Demos Medical Publishing New york, 2016.

[282] A. Filgueiras, E. Quintas Conde, and C. Hall, "The neural basis of kinesthetic and visual imagery in sports: an ale meta-analysis," *Brain Imaging Behav*, 2017.

[283] A. Guillot, C. Collet, V. A. Nguyen, F. Malouin, C. Richards, and J. Doyon, "Brain activity during visual versus kinesthetic imagery: an FMRI study.," *Hum Brain Mapp*, vol. 30, pp. 2157–2172, July 2009.

[284] A. Collins and E. Koechlin, "Reasoning, Learning, and Creativity: Frontal Lobe Function and Human Decision-Making," *PLoS Biology*, vol. 10, Mar. 2012.

[285] G. Pfurtscheller, A. Stancák, Jr, and G. Edlinger, "On the existence of different types of central beta rhythms below 30 hz," *Electroencephalogr Clin Neurophysiol*, vol. 102, pp. 316–25, Apr. 1997.

[286] C. Neuper and G. Pfurtscheller, *Handbook of electroencephalography and clinical neurophysiology. Event-related desynchronization*, ch. Motor imagery and ERD, pp. 303–325. Elsevier, 1999.

[287] O. Avilov, A. Popov, V. Timofieiev, L. Bougrain, and P. Henaff, "Estimation of imaginary movements quality based on machine learning for brain computer interface applications," *Microsystems, Electronics and Acoustics*, vol. 23, no. 5, pp. 25–31, 2018.

[288] O. Avilov, A. Popov, V. Timofieiev, L. Bougrain, and P. Henaff, "Combined imaginary movements classification using binary classifiers in brain-computer interfaces," *XVII International scientific and technical conference "Physical processes and fields of technical and biological objects", 2-4 November 2018.: proceedings*, pp. 92–94, 2018.

[289] C. Vidaurre and B. Blankertz, "Towards a Cure for BCI Illiteracy," *Brain Topography*, vol. 23, pp. 194–198, June 2010.

[290] F. Lotte, Cuntai Guan, and Kai Keng Ang, "Comparison of designs towards a subject-independent brain-computer interface based on motor imagery," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (Minneapolis, MN), pp. 4543–4546, IEEE, Sept. 2009.

[291] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871–890, 2015.

[292] A. P. Tu-Chan, N. Natraj, J. Godlove, G. Abrams, and K. Ganguly, "Effects of somatosensory electrical stimulation on motor function and cortical oscillations," *Journal of NeuroEngineering and Rehabilitation*, vol. 14, no. 1, p. 113, 2017.

[293] G. Buzsáki and X.-J. Wang, "Mechanisms of gamma oscillations," *Annual Review of Neuroscience*, vol. 35, pp. 203–225, 2012.

[294] M. Ahn, S. Ahn, J. H. Hong, H. Cho, K. Kim, B. S. Kim, J. W. Chang, and S. C. Jun, "Gamma band activity associated with BCI performance: simultaneous MEG/EEG study," *Frontiers in Human Neuroscience*, vol. 7, 2013.

[295] D. Cheyne and P. Ferrari, "MEG studies of motor cortex gamma oscillations: evidence for a gamma "fingerprint" in the brain?," *Frontiers in Human Neuroscience*, vol. 7, p. 575, 2013.

[296] S. D. Muthukumaraswamy, "Functional properties of human primary motor cortex gamma oscillations," *Journal of Neurophysiology*, vol. 104, pp. 2873–2885, Nov. 2010.

[297] M. Haller, J. Case, N. E. Crone, E. F. Chang, D. King-Stephens, K. D. Laxer, P. B. Weber, J. Parvizi, R. T. Knight, and A. Y. Shestyuk, "Persistent neuronal activity in human prefrontal cortex links perception and action," *Nature Human Behaviour*, vol. 2, pp. 80–91, Jan. 2018. Number: 1 Publisher: Nature Publishing Group.

[298] M. C. Bushnell, G. H. Duncan, R. K. Hofbauer, B. Ha, J.-I. Chen, and B. Carrier, "Pain perception: Is there a role for primary somatosensory cortex?," *Proceedings of the National Academy of Sciences*, vol. 96, pp. 7705–7709, July 1999. Publisher: National Academy of Sciences Section: Colloquium Paper.

[299] A. Guillot, C. Collet, V. A. Nguyen, F. Malouin, C. Richards, and J. Doyon, "Functional neuroanatomical networks associated with expertise in motor imagery," *NeuroImage*, vol. 41, pp. 1471–1483, July 2008.

[300] G. Wei and J. Luo, "Sport expert's motor imagery: functional imaging of professional motor skills and simple motor skills," *Brain Research*, vol. 1341, pp. 52–62, June 2010.

[301] H. Hentschke, A. Raz, B. M. Krause, C. A. Murphy, and M. I. Banks, "Disruption of cortical network activity by the general anaesthetic isoflurane," *British Journal of Anaesthesia*, vol. 119, pp. 685–696, Oct. 2017.

[302] G. A. Mashour and R. Lydic, "Neuroscientific Foundations of Anesthesiology," *European Journal of Anaesthesiology (EJA)*, vol. 29, p. 301, June 2012.

[303] S. J. Peltier, C. Kerssens, S. B. Hamann, P. S. Sebel, M. Byas-Smith, and X. Hu, "Functional connectivity changes with concentration of sevoflurane anesthesia," *NeuroReport*, vol. 16, pp. 285–288, Feb. 2005.

[304] P. L. C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian procrustes analysis: Transfer learning for brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 66, pp. 2390–2401, Aug. 2019.

[305] W. McNicholas, L. Doherty, S. Ryan, J. Garvey, P. Boyle, and E. Chua, "St. Vincent's University Hospital / University College Dublin Sleep Apnea Database," 2004. type: dataset.

[306] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen, and J. J. L. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG," *IEEE Transactions on Biomedical Engineering*, vol. 47, pp. 1185–1194, Sept. 2000. Conference Name: IEEE Transactions on Biomedical Engineering.

[307] M. G. Terzano, L. Parrino, A. Sherieri, R. Chervin, S. Chokroverty, C. Guilleminault, M. Hirshkowitz, M. Mahowald, H. Moldofsky, A. Rosa, R. Thomas, and A. Walters, "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep," *Sleep Medicine*, vol. 2, pp. 537–553, Nov. 2001.

[308] A. Vilamala, K. H. Madsen, and L. K. Hansen, "Deep Convolutional Neural Networks for Interpretable Analysis of EEG Sleep Stage Scoring," *arXiv:1710.00633 [cs, stat]*, Oct. 2017. arXiv: 1710.00633.

[309] T.-j. Luo, Y. Fan, L. Chen, G. Guo, and C. Zhou, "EEG Signal Reconstruction Using a Generative Adversarial Network With Wasserstein Distance and Temporal-Spatial-Frequency Loss," *Frontiers in Neuroinformatics*, vol. 14, 2020. Publisher: Frontiers.

[310] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, "EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals," *arXiv:1806.01875 [cs, eess, q-bio, stat]*, June 2018. arXiv: 1806.01875.

My publications

International journals

1. O. Avilov, S. Rimbert, A. Popov, and L. Bougrain, "Optimizing Motor Intention Detection With Deep Learning: Towards Management Of Intraoperative Awareness", *IEEE Transactions on Biomedical Engineering*, 2021. (accepted for publication)

Ukrainian journals

1. O. Avilov, A. Popov, V. Tymofeev, L. Bougrain, and P. Henaff, "Estimation of Imaginary Movements Quality Based on Machine Learning for Brain Computer Interface Applications," *Microsystems, Electronics and Acoustics,* vol. 23, no 5, pp. 25-31, 2018.

International conferences

1. S. Rimbert, O. Avilov, and L. Bougrain, "Discrete motor imageries can be used to allow a faster detection", *7th Graz Brain-Computer Interface Conference 2017.: proceedings,* Graz, Austria, 2017

2. C. Lindig-Leon, S. Rimbert, O. Avilov, and L. Bougrain, "Scalp EEG activity during simple and combined motor imageries to control a robotic ARM", *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2017.: proceedings,* pp. 322–327, Kyiv, 2017.

3. O. Avilov, "Machine learning in micro- and nanosystem technology", *X International scientific and technical conference ELCONF-2017, 25-27 April 2017: proceedings*, pp. 211-214, Kyiv, 2017.

4. O. Avilov, A. Popov, V. Tymofeev, L. Bougrain, and P. Henaff, "Combined imaginary movements classification using binary classifiers in brain-computer interfaces", *XVII International scientific and technical conference "Physical processes and fields of technical and biological objects", 2-4 November 2018.: proceedings*, pp. 92-94, Kremenchuk, 2018.

5. S. Rimbert, O. Avilov, P. Adam, and L. Bougrain, "Can suggestive hypnosis be used to improve BCI performance?" *8th Graz Brain-Computer Interface Conference 2019, Sep 2019.: proceedings*, Graz, Austria, 2019

6. O. Avilov, S. Rimbert, A. Popov, and L. Bougrain, "Deep Learning Techniques to Improve Intraoperative Awareness Detection from Electroencephalographic Signals", *42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society July 20-24, 2020.: proceedings*, Montreal, Canada, 2020.

Appendix A

Classification results for MNS vs. MI+MNS task using different classifiers

In this section, classification accuracy results obtained by different classifiers (CSP+LDA, MDRM, TS+LR, D-ConvNet, S-ConvNet, EEGNet-2.32, EEGNet-4.8) for 22 subjects from dataset introduced in Section 4.3 are presented. Results are represented for 12 different channel layouts for MNS vs. MNS+MI discriminating task. Signals filtered in 4 different frequency bands, results averaged across 10 folds of cross-validation.

Table A.1: Classification accuracy obtained by different classifiers for 22 subjects using 128 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.817	0.827	0.827	1.000	0.827	0.663	0.933	0.952	0.856	0.615	0.577	0.865	1.000	0.798	0.731	0.990	0.894	0.981	0.846	0.808	0.779	0.865	**0.839**
MDM	0.240	0.817	0.769	1.000	0.779	0.731	0.962	0.933	0.692	0.510	0.615	0.625	0.962	0.769	0.702	0.760	0.827	0.952	0.750	0.827	0.683	0.731	**0.756**
TS+LR	0.779	0.894	0.837	1.000	0.769	0.769	0.942	0.952	0.856	0.692	0.606	0.808	0.981	0.779	0.798	0.904	0.846	0.971	0.942	0.904	0.904	0.865	**0.854**
D-ConvNet	0.823	0.879	0.883	0.980	0.670	0.641	0.885	0.833	0.832	0.645	0.800	0.845	1.000	0.883	0.798	0.950	0.849	0.981	0.842	0.849	0.645	0.857	**0.835**
S-ConvNet	0.691	0.850	0.814	0.893	0.798	0.555	0.864	0.867	0.788	0.777	0.895	0.665	0.854	0.954	0.836	0.932	0.943	0.990	0.905	0.863	0.809	0.805	**0.834**
EEGNet-2.32	0.923	0.954	0.961	0.961	0.825	0.587	1.000	0.971	1.000	0.804	0.923	0.951	0.991	0.962	0.865	0.990	0.915	0.990	0.916	0.901	0.857	0.886	**0.915**
EEGNet-4.8	0.931	0.991	0.970	0.980	0.843	0.548	0.990	0.972	1.000	0.945	0.960	0.950	0.990	0.972	1.000	1.000	0.954	0.990	0.925	0.970	0.926	0.972	**0.945**

Table A.2: Classification accuracy obtained by different classifiers for 22 subjects using 128 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.702	0.779	0.875	1.000	0.788	0.712	0.933	0.933	0.942	0.596	0.471	0.808	0.952	0.769	0.721	0.981	0.865	0.971	0.913	0.817	0.798	0.827	**0.825**
MDM	0.365	0.817	0.788	1.000	0.769	0.740	0.952	0.923	0.683	0.510	0.625	0.606	0.923	0.769	0.702	0.769	0.808	0.962	0.750	0.769	0.712	0.760	**0.759**
TS+LR	0.615	0.885	0.846	1.000	0.769	0.769	0.971	0.933	0.760	0.654	0.654	0.750	0.981	0.779	0.788	0.865	0.827	0.971	0.923	0.904	0.885	0.798	**0.833**
D-ConvNet	0.645	0.859	0.915	0.961	0.672	0.605	0.776	0.829	0.795	0.617	0.770	0.739	0.933	0.943	0.747	0.923	0.878	0.981	0.743	0.836	0.596	0.693	**0.793**
S-ConvNet	0.692	0.803	0.815	0.834	0.788	0.585	0.815	0.848	0.817	0.758	0.913	0.760	0.816	0.913	0.797	0.905	0.944	0.990	0.944	0.882	0.846	0.816	**0.831**
EEGNet-2.32	0.834	0.916	0.941	0.921	0.788	0.585	0.990	0.961	0.964	0.813	0.884	0.895	0.961	0.895	0.846	0.925	0.875	0.970	0.935	0.913	0.885	0.935	**0.892**
EEGNet-4.8	0.923	1.000	0.960	0.950	0.855	0.576	0.980	0.972	0.981	0.831	0.942	0.900	0.981	0.925	0.944	1.000	0.915	0.981	0.971	0.970	0.914	0.955	**0.928**

Table A.3: Classification accuracy obtained by different classifiers for 22 subjects using 128 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.673	0.837	0.788	1.000	0.750	0.788	0.933	0.904	0.856	0.663	0.567	0.894	0.933	0.798	0.846	0.952	0.846	0.962	0.904	0.923	0.808	0.904	**0.842**
MDM	0.260	0.808	0.808	1.000	0.779	0.712	0.923	0.904	0.663	0.510	0.615	0.587	0.942	0.760	0.712	0.769	0.827	0.952	0.750	0.798	0.760	0.788	**0.756**
TS+LR	0.577	0.885	0.846	1.000	0.779	0.769	0.904	0.952	0.769	0.692	0.625	0.750	0.971	0.769	0.788	0.894	0.837	0.971	0.904	0.875	0.885	0.817	**0.830**
D-ConvNet	0.578	0.857	0.846	0.952	0.613	0.555	0.805	0.673	0.633	0.646	0.681	0.745	0.953	0.798	0.731	0.808	0.847	0.972	0.716	0.656	0.510	0.663	**0.738**
S-ConvNet	0.681	0.831	0.798	0.835	0.675	0.595	0.806	0.759	0.737	0.645	0.858	0.575	0.815	0.874	0.770	0.837	0.859	0.990	0.735	0.575	0.692	0.765	**0.759**
EEGNet-2.32	0.768	0.944	0.934	0.940	0.767	0.577	1.000	0.924	0.964	0.810	0.724	0.882	0.980	0.933	0.915	0.913	0.845	0.980	0.876	0.911	0.865	0.896	**0.879**
EEGNet-4.8	0.865	0.982	0.892	0.910	0.810	0.576	0.990	0.944	0.991	0.830	0.884	0.950	0.982	0.963	0.923	0.980	0.866	0.990	0.943	0.960	0.923	0.955	**0.914**

Table A.4: Classification accuracy obtained by different classifiers for 22 subjects using 128 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.779	0.856	0.769	1.000	0.769	0.654	0.942	0.904	0.933	0.567	0.567	0.894	1.000	0.779	0.750	0.971	0.913	0.971	0.913	0.827	0.856	0.904	**0.842**
MDM	0.202	0.808	0.779	1.000	0.788	0.702	0.933	0.923	0.673	0.500	0.596	0.625	0.952	0.769	0.702	0.760	0.837	0.952	0.750	0.837	0.712	0.779	**0.753**
TS+LR	0.760	0.894	0.885	1.000	0.779	0.769	0.894	0.952	0.885	0.731	0.587	0.808	0.971	0.769	0.798	0.923	0.856	0.971	0.904	0.885	0.875	0.885	**0.854**
D-ConvNet	0.718	0.953	0.826	0.971	0.632	0.659	0.960	0.742	0.727	0.733	0.710	0.823	0.990	0.866	0.845	0.970	0.878	0.991	0.704	0.749	0.556	0.815	**0.810**
S-ConvNet	0.663	0.851	0.825	0.864	0.657	0.586	0.932	0.768	0.710	0.759	0.905	0.614	0.972	0.905	0.759	0.854	0.848	0.990	0.795	0.635	0.731	0.837	**0.794**
EEGNet-2.32	0.905	0.954	0.971	0.941	0.826	0.585	1.000	0.972	1.000	0.819	0.800	0.942	0.990	0.972	0.925	0.950	0.925	0.970	0.915	0.912	0.830	0.963	**0.912**
EEGNet-4.8	0.945	0.991	0.961	0.922	0.912	0.548	0.990	0.954	1.000	0.895	0.972	0.971	1.000	0.972	0.971	0.980	0.934	0.990	0.953	0.970	0.933	0.962	**0.942**

Table A.5: Classification accuracy obtained by different classifiers for 22 subjects using 47 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.769	0.827	0.750	1.000	0.471	0.644	0.731	0.885	0.990	0.558	0.510	0.837	0.971	0.894	0.731	0.894	0.856	1.000	0.894	0.808	0.606	0.808	**0.792**
MDM	0.394	0.837	0.760	1.000	0.596	0.548	0.808	0.817	0.923	0.596	0.596	0.779	0.933	0.721	0.750	0.865	0.827	0.904	0.769	0.721	0.721	0.702	**0.753**
TS+LR	0.760	0.837	0.750	1.000	0.692	0.615	0.808	0.885	0.990	0.587	0.587	0.885	0.952	0.933	0.788	0.885	0.865	0.981	0.904	0.827	0.779	0.788	**0.823**
D-ConvNet	0.777	0.898	0.910	0.981	0.713	0.627	0.765	0.970	0.991	0.685	0.751	0.865	0.981	0.922	0.894	0.943	0.923	0.944	0.905	0.934	0.623	0.818	**0.855**
S-ConvNet	0.815	0.887	0.904	0.990	0.753	0.623	0.739	0.943	0.963	0.805	0.885	0.752	0.970	0.924	0.876	0.942	0.925	0.990	0.943	0.885	0.771	0.787	**0.867**
EEGNet-2.32	0.835	0.896	0.915	0.910	0.739	0.547	0.845	0.933	1.000	0.827	0.865	0.940	0.990	0.923	0.895	0.905	0.865	0.990	0.883	0.875	0.762	0.868	**0.873**
EEGNet-4.8	0.922	0.954	0.915	0.970	0.895	0.529	0.990	0.952	1.000	0.790	0.971	0.960	0.972	0.933	0.855	0.942	0.943	0.990	0.980	0.903	0.867	0.932	**0.916**

Table A.6: Classification accuracy obtained by different classifiers for 22 subjects using 47 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.731	0.817	0.788	1.000	0.558	0.606	0.750	0.875	0.981	0.548	0.481	0.769	0.952	0.923	0.769	0.846	0.856	1.000	0.885	0.788	0.625	0.769	**0.787**
MDM	0.510	0.837	0.779	1.000	0.615	0.596	0.808	0.808	0.837	0.606	0.596	0.769	0.904	0.721	0.750	0.865	0.817	0.904	0.769	0.721	0.673	0.673	**0.753**
TS+LR	0.721	0.837	0.740	1.000	0.615	0.596	0.846	0.885	0.952	0.635	0.606	0.875	0.971	0.923	0.779	0.894	0.865	0.952	0.913	0.788	0.788	0.779	**0.816**
D-ConvNet	0.830	0.936	0.903	0.990	0.730	0.567	0.631	0.940	0.991	0.568	0.720	0.734	0.972	0.972	0.805	0.952	0.907	0.971	0.914	0.865	0.635	0.710	**0.829**
S-ConvNet	0.795	0.887	0.864	1.000	0.744	0.545	0.729	0.933	0.888	0.825	0.856	0.702	0.893	0.902	0.914	0.942	0.924	1.000	0.923	0.892	0.771	0.758	**0.849**
EEGNet-2.32	0.823	0.906	0.943	0.960	0.719	0.558	0.748	0.943	1.000	0.778	0.875	0.922	0.972	0.932	0.855	0.915	0.905	0.990	0.894	0.894	0.838	0.821	**0.872**
EEGNet-4.8	0.862	0.954	0.893	0.940	0.758	0.548	0.815	0.952	1.000	0.802	0.961	0.932	0.953	0.933	0.934	0.924	0.971	0.990	0.970	0.921	0.887	0.944	**0.902**

Table A.7: Classification accuracy obtained by different classifiers for 22 subjects using 47 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.702	0.760	0.750	1.000	0.529	0.615	0.692	0.817	0.981	0.558	0.471	0.769	0.904	0.865	0.673	0.885	0.750	1.000	0.913	0.683	0.692	0.827	**0.765**
MDM	0.394	0.846	0.760	1.000	0.606	0.548	0.817	0.837	0.856	0.538	0.558	0.750	0.913	0.721	0.721	0.885	0.846	0.904	0.760	0.721	0.654	0.731	**0.744**
TS+LR	0.644	0.827	0.769	1.000	0.635	0.644	0.827	0.846	0.962	0.587	0.519	0.913	0.971	0.923	0.702	0.894	0.865	0.942	0.894	0.817	0.760	0.817	**0.807**
D-ConvNet	0.617	0.890	0.901	0.990	0.595	0.586	0.598	0.875	0.962	0.616	0.645	0.885	0.971	0.837	0.786	0.756	0.838	0.981	0.850	0.694	0.556	0.682	**0.778**
S-ConvNet	0.779	0.888	0.865	0.971	0.707	0.584	0.718	0.865	0.876	0.681	0.877	0.750	0.962	0.789	0.729	0.876	0.807	0.961	0.838	0.741	0.703	0.738	**0.805**
EEGNet-2.32	0.765	0.886	0.952	0.892	0.693	0.519	0.834	0.865	1.000	0.846	0.762	0.924	0.990	0.893	0.785	0.863	0.857	0.990	0.943	0.865	0.808	0.845	**0.854**
EEGNet-4.8	0.855	0.935	0.892	0.940	0.751	0.538	0.882	0.867	1.000	0.714	0.820	0.960	0.962	0.931	0.855	0.902	0.903	0.990	0.943	0.852	0.859	0.924	**0.876**

Table A.8: Classification accuracy obtained by different classifiers for 22 subjects using 47 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.760	0.808	0.760	1.000	0.587	0.625	0.712	0.808	0.990	0.452	0.433	0.904	0.971	0.827	0.596	0.894	0.788	1.000	0.913	0.596	0.740	0.875	**0.774**
MDM	0.231	0.846	0.740	1.000	0.606	0.558	0.817	0.846	0.913	0.538	0.548	0.779	0.952	0.712	0.731	0.865	0.846	0.904	0.750	0.712	0.712	0.721	**0.742**
TS+LR	0.731	0.827	0.750	1.000	0.731	0.673	0.798	0.837	0.981	0.606	0.519	0.923	0.962	0.923	0.721	0.885	0.865	0.981	0.894	0.837	0.750	0.827	**0.819**
D-ConvNet	0.675	0.936	0.921	0.971	0.666	0.608	0.794	0.776	0.991	0.688	0.685	0.923	0.972	0.874	0.757	0.826	0.778	0.981	0.935	0.749	0.585	0.838	**0.815**
S-ConvNet	0.749	0.889	0.885	0.981	0.689	0.605	0.729	0.836	0.951	0.713	0.904	0.816	0.944	0.815	0.755	0.873	0.857	0.980	0.809	0.817	0.683	0.779	**0.821**
EEGNet-2.32	0.845	0.925	0.950	0.930	0.819	0.528	0.882	0.904	1.000	0.700	0.791	0.942	0.990	0.905	0.806	0.865	0.865	1.000	0.923	0.835	0.799	0.884	**0.868**
EEGNet-4.8	0.915	0.935	0.922	0.940	0.896	0.529	0.961	0.886	1.000	0.794	0.868	0.980	0.972	0.924	0.845	0.895	0.895	0.990	0.914	0.855	0.877	0.953	**0.898**

Table A.9: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.808	0.788	0.702	1.000	0.567	0.635	0.673	0.808	0.673	0.615	0.558	0.577	0.538	0.856	0.702	0.846	0.721	0.885	0.894	0.673	0.683	0.817	**0.728**
MDM	0.596	0.817	0.769	1.000	0.462	0.596	0.673	0.798	0.615	0.654	0.577	0.654	0.490	0.779	0.702	0.817	0.779	0.817	0.817	0.635	0.654	0.654	**0.698**
TS+LR	0.760	0.837	0.769	1.000	0.548	0.673	0.779	0.885	0.683	0.596	0.558	0.731	0.606	0.885	0.740	0.827	0.788	0.885	0.894	0.750	0.740	0.769	**0.759**
D-ConvNet	0.766	0.897	0.891	1.000	0.789	0.566	0.626	0.921	0.753	0.616	0.686	0.752	0.824	0.885	0.895	0.855	0.845	0.943	0.933	0.817	0.567	0.885	**0.805**
S-ConvNet	0.825	0.898	0.845	0.951	0.722	0.575	0.692	0.951	0.715	0.785	0.790	0.767	0.747	0.874	0.788	0.894	0.903	0.970	0.905	0.797	0.695	0.796	**0.813**
EEGNet-2.32	0.815	0.887	0.873	0.940	0.789	0.500	0.701	0.931	0.751	0.726	0.884	0.837	0.860	0.896	0.885	0.903	0.905	0.980	0.885	0.844	0.767	0.885	**0.838**
EEGNet-4.8	0.873	0.906	0.843	0.960	0.789	0.482	0.718	0.951	0.739	0.772	0.953	0.932	0.973	0.925	0.934	0.886	0.895	0.981	0.952	0.893	0.820	0.874	**0.866**

Table A.10: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.837	0.817	0.721	1.000	0.596	0.654	0.683	0.846	0.692	0.548	0.596	0.635	0.558	0.894	0.683	0.837	0.731	0.856	0.856	0.673	0.692	0.779	**0.736**
MDM	0.663	0.808	0.798	1.000	0.433	0.587	0.683	0.798	0.615	0.644	0.587	0.673	0.442	0.808	0.702	0.808	0.760	0.808	0.817	0.673	0.673	0.625	**0.700**
TS+LR	0.740	0.827	0.769	1.000	0.529	0.692	0.750	0.885	0.663	0.615	0.577	0.731	0.558	0.894	0.721	0.846	0.798	0.875	0.894	0.760	0.740	0.712	**0.753**
D-ConvNet	0.805	0.907	0.902	0.990	0.769	0.604	0.615	0.960	0.799	0.684	0.732	0.730	0.575	0.875	0.864	0.845	0.895	0.915	0.933	0.817	0.530	0.730	**0.794**
S-ConvNet	0.845	0.867	0.806	0.980	0.712	0.635	0.679	0.923	0.715	0.776	0.790	0.757	0.690	0.855	0.778	0.865	0.893	0.952	0.923	0.807	0.724	0.800	**0.808**
EEGNet-2.32	0.854	0.897	0.854	0.970	0.760	0.473	0.672	0.940	0.770	0.745	0.875	0.826	0.727	0.924	0.875	0.884	0.865	0.971	0.903	0.864	0.758	0.837	**0.829**
EEGNet-4.8	0.844	0.925	0.872	0.941	0.818	0.492	0.686	0.942	0.750	0.799	0.943	0.876	0.824	0.914	0.924	0.903	0.864	0.971	0.951	0.864	0.846	0.895	**0.856**

Table A.11: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.683	0.788	0.721	1.000	0.500	0.663	0.683	0.750	0.606	0.606	0.538	0.721	0.490	0.846	0.644	0.856	0.740	0.846	0.865	0.606	0.635	0.740	**0.706**
MDM	0.587	0.788	0.779	1.000	0.452	0.587	0.644	0.798	0.625	0.606	0.587	0.625	0.462	0.769	0.702	0.846	0.769	0.846	0.798	0.615	0.673	0.683	**0.693**
TS+LR	0.606	0.808	0.827	1.000	0.500	0.692	0.740	0.798	0.683	0.625	0.538	0.721	0.462	0.846	0.740	0.885	0.808	0.865	0.894	0.673	0.788	0.750	**0.739**
D-ConvNet	0.714	0.896	0.903	1.000	0.646	0.535	0.590	0.839	0.752	0.619	0.724	0.737	0.668	0.790	0.747	0.653	0.786	0.944	0.887	0.743	0.606	0.730	**0.750**
S-ConvNet	0.777	0.905	0.835	0.990	0.731	0.556	0.721	0.866	0.674	0.644	0.796	0.703	0.755	0.855	0.767	0.778	0.808	0.932	0.830	0.656	0.715	0.787	**0.776**
EEGNet-2.32	0.785	0.877	0.854	0.980	0.626	0.509	0.720	0.943	0.694	0.769	0.745	0.855	0.766	0.807	0.836	0.785	0.762	0.971	0.885	0.805	0.740	0.826	**0.797**
EEGNet-4.8	0.851	0.926	0.853	0.960	0.729	0.500	0.688	0.895	0.721	0.731	0.814	0.895	0.808	0.873	0.855	0.885	0.827	0.980	0.914	0.749	0.797	0.865	**0.824**

Table A.12: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.683	0.788	0.750	1.000	0.433	0.625	0.721	0.750	0.615	0.606	0.471	0.683	0.558	0.827	0.712	0.865	0.740	0.923	0.875	0.644	0.673	0.769	**0.714**
MDM	0.548	0.808	0.750	1.000	0.433	0.567	0.644	0.798	0.606	0.625	0.587	0.606	0.462	0.750	0.702	0.837	0.750	0.846	0.788	0.606	0.635	0.692	**0.684**
TS+LR	0.606	0.798	0.798	1.000	0.433	0.663	0.731	0.798	0.683	0.663	0.490	0.740	0.577	0.856	0.750	0.865	0.808	0.865	0.894	0.683	0.788	0.760	**0.739**
D-ConvNet	0.693	0.915	0.912	1.000	0.627	0.589	0.643	0.752	0.766	0.658	0.665	0.769	0.855	0.836	0.700	0.778	0.779	0.952	0.868	0.713	0.520	0.778	**0.762**
S-ConvNet	0.786	0.895	0.846	0.980	0.731	0.539	0.702	0.865	0.675	0.684	0.778	0.732	0.747	0.847	0.729	0.862	0.858	0.961	0.829	0.685	0.769	0.818	**0.787**
EEGNet-2.32	0.755	0.898	0.835	0.970	0.693	0.509	0.632	0.933	0.741	0.675	0.783	0.922	0.981	0.875	0.827	0.925	0.847	0.990	0.905	0.826	0.769	0.874	**0.826**
EEGNet-4.8	0.805	0.916	0.874	0.970	0.759	0.482	0.738	0.905	0.758	0.733	0.842	0.904	0.970	0.884	0.856	0.855	0.856	0.990	0.914	0.809	0.808	0.884	**0.841**

Table A.13: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.702	0.827	0.740	0.654	0.702	0.635	0.740	0.856	0.721	0.750	0.452	0.683	0.933	0.913	0.750	0.808	0.875	1.000	0.875	0.654	0.740	0.663	**0.758**
MDM	0.538	0.721	0.769	0.817	0.779	0.635	0.683	0.808	0.683	0.692	0.577	0.702	0.779	0.779	0.769	0.769	0.856	0.990	0.760	0.625	0.663	0.654	**0.729**
TS+LR	0.692	0.846	0.779	0.769	0.750	0.740	0.769	0.894	0.760	0.788	0.471	0.702	0.942	0.933	0.760	0.769	0.904	1.000	0.865	0.731	0.779	0.673	**0.787**
D-ConvNet	0.650	0.906	0.881	0.875	0.711	0.632	0.605	0.770	0.770	0.616	0.779	0.657	0.982	0.961	0.854	0.769	0.895	0.971	0.963	0.794	0.570	0.673	**0.786**
S-ConvNet	0.728	0.866	0.864	0.894	0.846	0.599	0.750	0.855	0.792	0.778	0.866	0.701	0.962	0.941	0.722	0.856	0.945	1.000	0.933	0.845	0.798	0.770	**0.832**
EEGNet-2.32	0.777	0.925	0.885	0.863	0.759	0.643	0.746	0.911	0.820	0.741	0.864	0.885	0.981	0.933	0.837	0.874	0.876	1.000	0.925	0.865	0.811	0.799	**0.851**
EEGNet-4.8	0.838	0.935	0.894	0.893	0.798	0.665	0.769	0.893	0.850	0.810	0.952	0.940	0.972	0.953	0.884	0.882	0.935	1.000	0.981	0.834	0.916	0.780	**0.880**

Table A.14: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.702	0.808	0.788	0.702	0.692	0.625	0.721	0.827	0.712	0.740	0.500	0.663	0.865	0.865	0.740	0.817	0.885	1.000	0.875	0.635	0.731	0.683	**0.753**
MDM	0.625	0.740	0.750	0.817	0.769	0.625	0.673	0.808	0.702	0.673	0.587	0.673	0.740	0.817	0.779	0.779	0.846	0.990	0.760	0.606	0.702	0.625	**0.731**
TS+LR	0.712	0.808	0.788	0.769	0.760	0.721	0.779	0.904	0.769	0.760	0.500	0.712	0.817	0.913	0.740	0.817	0.885	0.990	0.865	0.731	0.788	0.683	**0.782**
D-ConvNet	0.722	0.865	0.892	0.941	0.730	0.653	0.612	0.770	0.768	0.615	0.824	0.704	0.963	0.971	0.873	0.681	0.943	0.953	0.952	0.784	0.581	0.672	**0.794**
S-ConvNet	0.709	0.743	0.845	0.904	0.865	0.609	0.759	0.855	0.791	0.787	0.885	0.682	0.885	0.934	0.752	0.875	0.934	0.991	0.933	0.815	0.749	0.777	**0.822**
EEGNet-2.32	0.750	0.828	0.874	0.844	0.816	0.577	0.815	0.882	0.819	0.712	0.854	0.817	0.953	0.895	0.856	0.874	0.895	1.000	0.945	0.865	0.788	0.759	**0.837**
EEGNet-4.8	0.840	0.934	0.854	0.904	0.818	0.662	0.709	0.892	0.841	0.790	0.952	0.845	0.962	0.925	0.875	0.864	0.924	1.000	0.961	0.805	0.885	0.809	**0.866**

Table A.15: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.606	0.731	0.788	0.654	0.740	0.721	0.587	0.817	0.663	0.702	0.490	0.692	0.894	0.865	0.692	0.712	0.798	1.000	0.865	0.615	0.788	0.692	**0.733**
MDM	0.462	0.731	0.788	0.808	0.779	0.673	0.673	0.779	0.606	0.663	0.606	0.673	0.760	0.788	0.808	0.750	0.750	0.990	0.760	0.587	0.721	0.673	**0.719**
TS+LR	0.519	0.788	0.788	0.740	0.760	0.750	0.769	0.837	0.740	0.760	0.538	0.740	0.865	0.933	0.779	0.769	0.875	0.990	0.904	0.615	0.827	0.673	**0.771**
D-ConvNet	0.675	0.732	0.844	0.757	0.672	0.595	0.518	0.596	0.713	0.667	0.663	0.674	0.944	0.839	0.597	0.510	0.817	0.962	0.925	0.605	0.587	0.664	**0.707**
S-ConvNet	0.779	0.839	0.817	0.711	0.807	0.598	0.691	0.770	0.772	0.652	0.790	0.702	0.916	0.912	0.738	0.749	0.780	0.971	0.877	0.683	0.816	0.711	**0.776**
EEGNet-2.32	0.700	0.869	0.814	0.825	0.788	0.556	0.709	0.790	0.724	0.685	0.915	0.848	0.962	0.923	0.854	0.795	0.837	1.000	0.907	0.691	0.799	0.625	**0.801**
EEGNet-4.8	0.733	0.895	0.823	0.816	0.836	0.615	0.682	0.788	0.783	0.705	0.961	0.835	0.943	0.895	0.826	0.738	0.894	1.000	0.945	0.777	0.850	0.838	**0.826**

Table A.16: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.606	0.788	0.731	0.673	0.731	0.760	0.625	0.827	0.625	0.702	0.490	0.663	0.952	0.865	0.702	0.673	0.837	0.990	0.865	0.673	0.846	0.683	**0.741**
MDM	0.346	0.731	0.769	0.817	0.788	0.673	0.673	0.798	0.606	0.673	0.596	0.663	0.846	0.779	0.798	0.721	0.788	0.990	0.750	0.606	0.683	0.702	**0.718**
TS+LR	0.567	0.808	0.779	0.740	0.760	0.760	0.731	0.827	0.731	0.760	0.529	0.740	0.942	0.913	0.798	0.760	0.904	1.000	0.894	0.635	0.827	0.673	**0.776**
D-ConvNet	0.653	0.905	0.862	0.700	0.748	0.536	0.627	0.664	0.660	0.618	0.638	0.796	0.981	0.791	0.716	0.585	0.738	0.971	0.905	0.568	0.615	0.625	**0.723**
S-ConvNet	0.800	0.888	0.863	0.732	0.835	0.628	0.680	0.771	0.790	0.719	0.838	0.712	0.981	0.863	0.748	0.786	0.836	1.000	0.905	0.671	0.800	0.741	**0.799**
EEGNet-2.32	0.703	0.867	0.902	0.786	0.715	0.565	0.729	0.758	0.705	0.705	0.850	0.856	0.981	0.913	0.815	0.844	0.846	1.000	0.917	0.750	0.800	0.762	**0.808**
EEGNet-4.8	0.755	0.915	0.892	0.807	0.768	0.558	0.739	0.799	0.781	0.754	0.904	0.950	0.990	0.933	0.828	0.805	0.835	1.000	0.924	0.747	0.866	0.773	**0.833**

Table A.17: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.769	0.750	0.692	0.731	0.606	0.519	0.644	0.760	0.615	0.644	0.510	0.702	0.442	0.798	0.740	0.721	0.548	0.865	0.913	0.663	0.606	0.673	**0.678**
MDM	0.712	0.721	0.702	0.798	0.462	0.548	0.692	0.769	0.615	0.683	0.567	0.615	0.413	0.760	0.721	0.740	0.615	0.750	0.837	0.567	0.644	0.529	**0.657**
TS+LR	0.798	0.817	0.702	0.740	0.567	0.606	0.731	0.875	0.596	0.654	0.548	0.663	0.452	0.865	0.740	0.769	0.692	0.846	0.885	0.615	0.692	0.673	**0.706**
D-ConvNet	0.787	0.830	0.910	0.683	0.768	0.636	0.672	0.871	0.673	0.625	0.637	0.731	0.834	0.925	0.893	0.689	0.787	0.924	0.885	0.786	0.565	0.740	**0.766**
S-ConvNet	0.805	0.868	0.815	0.765	0.780	0.570	0.759	0.895	0.694	0.760	0.781	0.729	0.787	0.835	0.768	0.836	0.837	0.951	0.914	0.777	0.732	0.779	**0.793**
EEGNet-2.32	0.845	0.905	0.895	0.796	0.733	0.616	0.634	0.903	0.753	0.691	0.864	0.765	0.922	0.845	0.865	0.798	0.865	0.970	0.893	0.748	0.836	0.732	**0.812**
EEGNet-4.8	0.817	0.935	0.853	0.815	0.838	0.617	0.796	0.875	0.750	0.713	0.913	0.854	0.953	0.934	0.855	0.854	0.847	0.990	0.941	0.845	0.808	0.808	**0.846**

Table A.18: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.808	0.692	0.731	0.731	0.587	0.529	0.635	0.779	0.596	0.644	0.577	0.712	0.394	0.827	0.692	0.750	0.625	0.865	0.923	0.673	0.702	0.654	**0.687**
MDM	0.731	0.654	0.760	0.798	0.452	0.558	0.692	0.779	0.615	0.673	0.567	0.635	0.365	0.827	0.721	0.769	0.663	0.769	0.837	0.596	0.683	0.519	**0.667**
TS+LR	0.788	0.779	0.740	0.740	0.529	0.587	0.712	0.875	0.606	0.625	0.587	0.663	0.433	0.856	0.750	0.798	0.712	0.827	0.894	0.635	0.702	0.644	**0.704**
D-ConvNet	0.778	0.839	0.911	0.644	0.730	0.645	0.664	0.827	0.615	0.614	0.724	0.666	0.641	0.883	0.865	0.643	0.777	0.904	0.905	0.776	0.606	0.696	**0.743**
S-ConvNet	0.795	0.839	0.796	0.775	0.752	0.561	0.767	0.905	0.694	0.770	0.781	0.749	0.795	0.834	0.750	0.836	0.778	0.941	0.923	0.797	0.751	0.749	**0.788**
EEGNet-2.32	0.797	0.857	0.835	0.815	0.743	0.645	0.721	0.894	0.714	0.711	0.895	0.756	0.769	0.855	0.875	0.835	0.788	0.961	0.894	0.733	0.758	0.727	**0.799**
EEGNet-4.8	0.805	0.886	0.872	0.824	0.828	0.599	0.758	0.885	0.744	0.750	0.922	0.825	0.837	0.895	0.894	0.807	0.791	0.952	0.971	0.806	0.837	0.779	**0.830**

Table A.19: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.538	0.721	0.692	0.760	0.519	0.635	0.663	0.769	0.702	0.654	0.538	0.596	0.538	0.779	0.663	0.827	0.596	0.808	0.894	0.635	0.615	0.587	**0.670**
MDM	0.625	0.692	0.788	0.808	0.462	0.587	0.644	0.750	0.606	0.625	0.596	0.606	0.394	0.750	0.692	0.798	0.615	0.779	0.827	0.587	0.673	0.587	**0.659**
TS+LR	0.548	0.731	0.798	0.798	0.462	0.692	0.702	0.798	0.683	0.654	0.490	0.692	0.471	0.846	0.721	0.817	0.663	0.817	0.885	0.567	0.721	0.673	**0.692**
D-ConvNet	0.771	0.809	0.881	0.728	0.627	0.597	0.541	0.595	0.682	0.588	0.620	0.692	0.662	0.675	0.767	0.631	0.662	0.943	0.820	0.576	0.605	0.684	**0.689**
S-ConvNet	0.767	0.826	0.864	0.795	0.810	0.519	0.740	0.826	0.655	0.656	0.761	0.782	0.710	0.855	0.720	0.787	0.724	0.924	0.810	0.686	0.725	0.769	**0.760**
EEGNet-2.32	0.753	0.857	0.816	0.797	0.710	0.618	0.732	0.865	0.714	0.730	0.727	0.836	0.712	0.806	0.838	0.759	0.684	0.953	0.914	0.586	0.807	0.662	**0.767**
EEGNet-4.8	0.806	0.866	0.883	0.844	0.807	0.673	0.719	0.827	0.733	0.752	0.800	0.766	0.768	0.828	0.837	0.750	0.763	0.960	0.925	0.740	0.836	0.754	**0.802**

Table A.20: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.510	0.760	0.654	0.779	0.529	0.615	0.683	0.769	0.644	0.644	0.462	0.596	0.471	0.760	0.683	0.798	0.644	0.894	0.894	0.548	0.625	0.606	**0.662**
MDM	0.577	0.760	0.740	0.798	0.452	0.587	0.644	0.750	0.606	0.635	0.587	0.606	0.404	0.692	0.702	0.769	0.635	0.788	0.817	0.548	0.635	0.596	**0.651**
TS+LR	0.538	0.750	0.769	0.798	0.452	0.663	0.702	0.788	0.654	0.663	0.481	0.654	0.567	0.856	0.731	0.769	0.692	0.856	0.885	0.558	0.673	0.683	**0.690**
D-ConvNet	0.741	0.858	0.893	0.718	0.550	0.645	0.682	0.605	0.675	0.596	0.628	0.653	0.875	0.665	0.641	0.595	0.797	0.933	0.801	0.674	0.520	0.722	**0.703**
S-ConvNet	0.767	0.858	0.875	0.785	0.790	0.520	0.741	0.836	0.694	0.647	0.731	0.773	0.796	0.837	0.766	0.776	0.810	0.961	0.801	0.684	0.725	0.797	**0.771**
EEGNet-2.32	0.789	0.925	0.874	0.776	0.692	0.646	0.749	0.856	0.724	0.695	0.755	0.816	0.921	0.772	0.829	0.796	0.797	0.980	0.932	0.635	0.750	0.701	**0.791**
EEGNet-4.8	0.735	0.926	0.892	0.863	0.756	0.645	0.805	0.807	0.721	0.752	0.810	0.815	0.953	0.895	0.845	0.835	0.856	0.953	0.913	0.758	0.769	0.790	**0.823**

Table A.21: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.808	0.760	0.712	0.702	0.683	0.683	0.692	0.808	0.721	0.769	0.442	0.692	0.433	0.856	0.750	0.856	0.750	0.990	0.885	0.654	0.750	0.644	**0.729**
MDM	0.731	0.740	0.760	0.760	0.788	0.644	0.712	0.760	0.673	0.702	0.548	0.721	0.529	0.731	0.769	0.760	0.625	0.962	0.769	0.606	0.654	0.606	**0.707**
TS+LR	0.779	0.798	0.750	0.692	0.731	0.654	0.740	0.798	0.625	0.779	0.490	0.740	0.577	0.894	0.769	0.769	0.760	0.981	0.856	0.673	0.788	0.596	**0.738**
D-ConvNet	0.727	0.855	0.891	0.805	0.751	0.699	0.595	0.750	0.779	0.635	0.883	0.733	0.931	0.954	0.834	0.738	0.934	0.954	0.972	0.735	0.699	0.626	**0.795**
S-ConvNet	0.748	0.811	0.825	0.864	0.807	0.635	0.691	0.788	0.839	0.719	0.817	0.734	0.719	0.942	0.796	0.865	0.903	1.000	0.934	0.734	0.767	0.759	**0.804**
EEGNet-2.32	0.845	0.885	0.885	0.854	0.778	0.617	0.752	0.883	0.800	0.703	0.951	0.749	0.884	0.944	0.876	0.825	0.904	1.000	0.934	0.797	0.835	0.673	**0.835**
EEGNet-4.8	0.806	0.867	0.884	0.835	0.788	0.623	0.767	0.884	0.840	0.811	0.932	0.809	0.933	0.944	0.875	0.854	0.914	1.000	0.981	0.756	0.855	0.757	**0.851**

Table A.22: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.760	0.692	0.721	0.721	0.663	0.683	0.721	0.788	0.663	0.721	0.442	0.712	0.481	0.846	0.760	0.808	0.740	1.000	0.894	0.654	0.712	0.596	**0.717**
MDM	0.731	0.712	0.750	0.760	0.750	0.663	0.712	0.760	0.683	0.702	0.567	0.769	0.510	0.740	0.769	0.760	0.654	0.971	0.769	0.606	0.692	0.577	**0.709**
TS+LR	0.779	0.740	0.740	0.721	0.721	0.654	0.760	0.798	0.663	0.760	0.529	0.740	0.567	0.885	0.769	0.798	0.779	0.971	0.875	0.692	0.760	0.577	**0.740**
D-ConvNet	0.799	0.745	0.872	0.802	0.770	0.625	0.630	0.690	0.694	0.502	0.883	0.797	0.678	0.952	0.854	0.777	0.943	0.924	0.952	0.710	0.699	0.645	**0.770**
S-ConvNet	0.729	0.770	0.815	0.864	0.826	0.645	0.700	0.788	0.822	0.718	0.828	0.763	0.732	0.932	0.798	0.875	0.865	0.971	0.934	0.714	0.739	0.729	**0.798**
EEGNet-2.32	0.835	0.838	0.863	0.825	0.777	0.635	0.762	0.873	0.839	0.719	0.941	0.768	0.729	0.934	0.895	0.836	0.865	1.000	0.954	0.789	0.816	0.683	**0.826**
EEGNet-4.8	0.808	0.865	0.874	0.882	0.827	0.612	0.757	0.875	0.849	0.760	0.932	0.826	0.789	0.933	0.885	0.864	0.903	1.000	0.961	0.804	0.798	0.690	**0.841**

Table A.23: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.654	0.750	0.702	0.702	0.683	0.692	0.654	0.817	0.606	0.731	0.433	0.683	0.462	0.856	0.808	0.750	0.702	0.981	0.885	0.654	0.837	0.577	**0.710**
MDM	0.615	0.731	0.779	0.760	0.760	0.663	0.702	0.740	0.635	0.683	0.577	0.625	0.471	0.731	0.779	0.760	0.538	0.962	0.760	0.567	0.702	0.615	**0.689**
TS+LR	0.635	0.740	0.731	0.721	0.712	0.740	0.721	0.808	0.673	0.760	0.529	0.683	0.558	0.894	0.769	0.760	0.635	0.981	0.875	0.587	0.798	0.635	**0.725**
D-ConvNet	0.684	0.693	0.834	0.690	0.722	0.664	0.624	0.665	0.638	0.576	0.676	0.699	0.602	0.815	0.669	0.510	0.790	0.952	0.903	0.539	0.777	0.585	**0.696**
S-ConvNet	0.711	0.827	0.824	0.760	0.785	0.559	0.633	0.798	0.763	0.702	0.817	0.731	0.665	0.846	0.796	0.826	0.749	0.980	0.925	0.605	0.771	0.673	**0.761**
EEGNet-2.32	0.759	0.771	0.807	0.807	0.768	0.595	0.739	0.790	0.753	0.683	0.914	0.742	0.737	0.876	0.857	0.786	0.828	1.000	0.963	0.656	0.825	0.603	**0.785**
EEGNet-4.8	0.787	0.847	0.872	0.837	0.788	0.639	0.755	0.771	0.784	0.762	0.912	0.788	0.799	0.914	0.855	0.796	0.865	1.000	0.945	0.655	0.826	0.680	**0.813**

Table A.24: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.654	0.779	0.712	0.692	0.683	0.702	0.673	0.837	0.577	0.721	0.423	0.654	0.519	0.865	0.779	0.712	0.750	0.981	0.885	0.625	0.760	0.663	**0.711**
MDM	0.615	0.750	0.740	0.779	0.798	0.673	0.721	0.740	0.635	0.692	0.567	0.615	0.490	0.702	0.779	0.731	0.538	0.962	0.760	0.548	0.673	0.606	**0.687**
TS+LR	0.635	0.779	0.721	0.712	0.740	0.740	0.712	0.837	0.673	0.750	0.510	0.673	0.615	0.904	0.769	0.740	0.721	0.971	0.865	0.577	0.817	0.663	**0.733**
D-ConvNet	0.735	0.876	0.842	0.730	0.759	0.613	0.723	0.643	0.547	0.586	0.695	0.694	0.894	0.730	0.682	0.528	0.775	0.954	0.876	0.483	0.728	0.655	**0.716**
S-ConvNet	0.729	0.887	0.827	0.797	0.795	0.557	0.661	0.809	0.781	0.713	0.790	0.721	0.750	0.875	0.712	0.789	0.827	0.991	0.915	0.566	0.769	0.665	**0.769**
EEGNet-2.32	0.788	0.887	0.846	0.826	0.730	0.595	0.748	0.789	0.725	0.751	0.855	0.779	0.943	0.886	0.857	0.776	0.875	1.000	0.963	0.686	0.806	0.663	**0.808**
EEGNet-4.8	0.799	0.867	0.863	0.856	0.798	0.600	0.796	0.829	0.837	0.800	0.913	0.845	0.942	0.915	0.854	0.784	0.855	1.000	0.935	0.673	0.807	0.771	**0.834**

Table A.25: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.760	0.808	0.663	0.798	0.510	0.587	0.558	0.731	0.615	0.644	0.433	0.587	0.510	0.712	0.721	0.731	0.577	0.904	0.846	0.625	0.625	0.663	**0.664**
MDM	0.721	0.702	0.663	0.779	0.462	0.596	0.654	0.750	0.596	0.712	0.548	0.644	0.510	0.702	0.740	0.760	0.596	0.740	0.750	0.538	0.644	0.596	**0.655**
TS+LR	0.750	0.798	0.654	0.750	0.558	0.558	0.692	0.750	0.654	0.644	0.442	0.558	0.567	0.760	0.740	0.779	0.606	0.856	0.827	0.587	0.673	0.587	**0.672**
D-ConvNet	0.816	0.810	0.825	0.663	0.752	0.568	0.645	0.635	0.674	0.559	0.610	0.634	0.875	0.655	0.913	0.635	0.817	0.932	0.829	0.654	0.702	0.594	**0.718**
S-ConvNet	0.833	0.877	0.747	0.740	0.751	0.549	0.749	0.798	0.721	0.694	0.733	0.682	0.815	0.845	0.789	0.826	0.821	0.951	0.887	0.635	0.733	0.739	**0.769**
EEGNet-2.32	0.815	0.907	0.749	0.778	0.752	0.577	0.727	0.787	0.655	0.685	0.779	0.805	0.846	0.808	0.817	0.758	0.817	0.971	0.876	0.617	0.818	0.722	**0.776**
EEGNet-4.8	0.815	0.916	0.748	0.837	0.770	0.558	0.765	0.808	0.781	0.725	0.866	0.740	0.932	0.865	0.875	0.865	0.829	0.980	0.878	0.827	0.847	0.740	**0.817**

Table A.26: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.808	0.721	0.654	0.769	0.538	0.558	0.596	0.750	0.606	0.644	0.481	0.606	0.442	0.702	0.740	0.760	0.567	0.875	0.856	0.644	0.615	0.606	**0.661**
MDM	0.721	0.673	0.673	0.779	0.462	0.587	0.635	0.760	0.606	0.683	0.558	0.644	0.433	0.721	0.740	0.779	0.587	0.769	0.750	0.567	0.654	0.529	**0.650**
TS+LR	0.769	0.788	0.644	0.731	0.558	0.558	0.721	0.750	0.644	0.663	0.471	0.567	0.490	0.750	0.740	0.808	0.567	0.846	0.827	0.615	0.712	0.577	**0.673**
D-ConvNet	0.736	0.702	0.825	0.711	0.780	0.626	0.635	0.636	0.693	0.578	0.724	0.690	0.623	0.720	0.895	0.585	0.854	0.924	0.847	0.654	0.662	0.644	**0.716**
S-ConvNet	0.793	0.782	0.739	0.730	0.732	0.540	0.718	0.760	0.682	0.674	0.704	0.652	0.759	0.845	0.819	0.779	0.778	0.905	0.869	0.610	0.743	0.728	**0.743**
EEGNet-2.32	0.835	0.856	0.719	0.797	0.714	0.567	0.718	0.797	0.705	0.692	0.787	0.729	0.748	0.864	0.846	0.758	0.673	0.962	0.865	0.575	0.828	0.690	**0.760**
EEGNet-4.8	0.795	0.866	0.759	0.807	0.798	0.569	0.689	0.806	0.743	0.743	0.858	0.692	0.755	0.885	0.845	0.808	0.789	0.961	0.898	0.750	0.847	0.691	**0.789**

Table A.27: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.577	0.740	0.683	0.798	0.500	0.577	0.625	0.721	0.587	0.635	0.538	0.635	0.452	0.673	0.760	0.750	0.510	0.865	0.875	0.615	0.683	0.625	**0.656**
MDM	0.606	0.702	0.721	0.788	0.462	0.615	0.673	0.721	0.606	0.663	0.606	0.673	0.385	0.702	0.731	0.798	0.481	0.779	0.779	0.567	0.663	0.615	**0.652**
TS+LR	0.567	0.779	0.702	0.769	0.500	0.577	0.654	0.750	0.663	0.606	0.510	0.692	0.567	0.788	0.750	0.827	0.510	0.865	0.837	0.615	0.683	0.577	**0.672**
D-ConvNet	0.729	0.714	0.815	0.625	0.623	0.557	0.635	0.692	0.739	0.560	0.587	0.603	0.492	0.615	0.702	0.585	0.663	0.932	0.742	0.510	0.654	0.626	**0.654**
S-ConvNet	0.785	0.812	0.738	0.709	0.692	0.519	0.684	0.780	0.596	0.605	0.724	0.642	0.710	0.764	0.730	0.798	0.675	0.924	0.800	0.577	0.676	0.701	**0.711**
EEGNet-2.32	0.738	0.838	0.765	0.720	0.663	0.595	0.602	0.751	0.683	0.655	0.704	0.710	0.688	0.702	0.786	0.728	0.587	0.961	0.868	0.507	0.779	0.636	**0.712**
EEGNet-4.8	0.824	0.884	0.748	0.826	0.732	0.578	0.706	0.710	0.722	0.665	0.779	0.751	0.698	0.730	0.779	0.768	0.761	0.980	0.838	0.613	0.790	0.674	**0.753**

Table A.28: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.577	0.779	0.644	0.808	0.538	0.587	0.654	0.721	0.567	0.673	0.433	0.587	0.567	0.702	0.740	0.712	0.606	0.837	0.865	0.596	0.654	0.635	**0.658**
MDM	0.596	0.760	0.712	0.779	0.462	0.625	0.663	0.721	0.625	0.692	0.596	0.654	0.471	0.673	0.740	0.769	0.519	0.750	0.769	0.558	0.635	0.615	**0.654**
TS+LR	0.529	0.788	0.721	0.769	0.462	0.587	0.644	0.760	0.635	0.625	0.490	0.683	0.606	0.798	0.760	0.788	0.558	0.856	0.837	0.577	0.692	0.625	**0.672**
D-ConvNet	0.699	0.788	0.760	0.586	0.605	0.597	0.740	0.596	0.700	0.598	0.628	0.652	0.885	0.635	0.693	0.598	0.770	0.914	0.821	0.609	0.682	0.587	**0.688**
S-ConvNet	0.757	0.887	0.785	0.720	0.663	0.538	0.683	0.753	0.625	0.606	0.694	0.651	0.854	0.751	0.693	0.787	0.723	0.924	0.800	0.693	0.676	0.729	**0.727**
EEGNet-2.32	0.787	0.896	0.759	0.711	0.674	0.605	0.729	0.741	0.665	0.695	0.789	0.679	0.892	0.779	0.786	0.769	0.790	0.961	0.869	0.641	0.788	0.722	**0.760**
EEGNet-4.8	0.778	0.886	0.776	0.836	0.712	0.589	0.778	0.751	0.763	0.734	0.878	0.777	0.912	0.845	0.788	0.845	0.791	0.980	0.896	0.698	0.782	0.700	**0.795**

Table A.29: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.721	0.827	0.740	0.673	0.750	0.692	0.731	0.788	0.779	0.788	0.587	0.692	0.452	0.846	0.750	0.798	0.750	0.971	0.865	0.529	0.769	0.490	**0.727**
MDM	0.760	0.760	0.760	0.740	0.769	0.635	0.673	0.798	0.663	0.673	0.558	0.683	0.558	0.721	0.779	0.769	0.596	0.971	0.750	0.567	0.712	0.433	**0.697**
TS+LR	0.798	0.808	0.740	0.721	0.769	0.712	0.663	0.856	0.712	0.779	0.558	0.731	0.538	0.856	0.731	0.817	0.740	0.971	0.894	0.558	0.788	0.577	**0.742**
D-ConvNet	0.751	0.866	0.863	0.719	0.741	0.667	0.555	0.874	0.789	0.705	0.753	0.655	0.718	0.925	0.577	0.652	0.944	0.943	0.963	0.586	0.746	0.596	**0.754**
S-ConvNet	0.778	0.857	0.903	0.786	0.834	0.645	0.625	0.807	0.782	0.645	0.762	0.645	0.758	0.911	0.787	0.826	0.866	1.000	0.924	0.693	0.796	0.722	**0.789**
EEGNet-2.32	0.765	0.867	0.835	0.797	0.750	0.770	0.661	0.894	0.789	0.752	0.825	0.808	0.894	0.886	0.767	0.795	0.885	1.000	0.955	0.675	0.816	0.681	**0.812**
EEGNet-4.8	0.808	0.905	0.875	0.845	0.795	0.698	0.633	0.894	0.877	0.774	0.818	0.769	0.902	0.935	0.815	0.795	0.925	1.000	0.915	0.738	0.806	0.791	**0.832**

Table A.30: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.683	0.769	0.769	0.721	0.740	0.712	0.760	0.788	0.769	0.740	0.587	0.683	0.442	0.865	0.779	0.827	0.740	0.971	0.885	0.606	0.740	0.548	**0.733**
MDM	0.731	0.731	0.779	0.721	0.750	0.635	0.673	0.798	0.663	0.673	0.558	0.663	0.529	0.731	0.760	0.788	0.644	0.971	0.750	0.567	0.721	0.433	**0.694**
TS+LR	0.817	0.779	0.779	0.731	0.769	0.712	0.683	0.865	0.750	0.779	0.548	0.721	0.519	0.846	0.721	0.837	0.760	0.971	0.904	0.558	0.740	0.587	**0.744**
D-ConvNet	0.752	0.817	0.844	0.797	0.710	0.706	0.546	0.854	0.817	0.619	0.758	0.657	0.725	0.923	0.567	0.625	0.905	0.904	0.954	0.528	0.710	0.576	**0.741**
S-ConvNet	0.759	0.818	0.893	0.785	0.816	0.616	0.606	0.797	0.790	0.681	0.750	0.654	0.768	0.931	0.729	0.865	0.835	0.990	0.905	0.673	0.758	0.693	**0.778**
EEGNet-2.32	0.815	0.857	0.844	0.796	0.815	0.759	0.661	0.894	0.840	0.722	0.816	0.779	0.721	0.885	0.767	0.795	0.894	1.000	0.954	0.724	0.825	0.675	**0.811**
EEGNet-4.8	0.797	0.817	0.864	0.826	0.789	0.737	0.644	0.884	0.849	0.696	0.834	0.785	0.755	0.916	0.767	0.805	0.857	1.000	0.953	0.711	0.807	0.722	**0.810**

Table A.31: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.654	0.740	0.712	0.663	0.692	0.760	0.692	0.856	0.692	0.779	0.471	0.683	0.404	0.846	0.798	0.750	0.644	0.942	0.856	0.587	0.788	0.644	**0.712**
MDM	0.625	0.760	0.692	0.798	0.769	0.731	0.683	0.760	0.625	0.663	0.625	0.635	0.404	0.712	0.779	0.740	0.548	0.942	0.760	0.538	0.740	0.615	**0.688**
TS+LR	0.702	0.769	0.731	0.740	0.731	0.769	0.683	0.837	0.712	0.750	0.510	0.721	0.394	0.827	0.788	0.760	0.663	0.971	0.856	0.596	0.827	0.644	**0.726**
D-ConvNet	0.646	0.855	0.783	0.808	0.781	0.659	0.541	0.761	0.682	0.695	0.647	0.721	0.531	0.827	0.558	0.540	0.788	0.913	0.886	0.589	0.645	0.595	**0.702**
S-ConvNet	0.682	0.847	0.866	0.758	0.785	0.655	0.649	0.825	0.771	0.651	0.831	0.750	0.645	0.856	0.708	0.817	0.749	0.991	0.866	0.664	0.767	0.690	**0.765**
EEGNet-2.32	0.709	0.827	0.741	0.780	0.809	0.756	0.653	0.840	0.724	0.645	0.799	0.799	0.641	0.897	0.739	0.795	0.808	1.000	0.896	0.645	0.863	0.644	**0.773**
EEGNet-4.8	0.825	0.750	0.807	0.759	0.799	0.737	0.691	0.846	0.742	0.783	0.816	0.789	0.710	0.935	0.747	0.815	0.807	1.000	0.857	0.712	0.796	0.680	**0.791**

Table A.32: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.712	0.788	0.692	0.673	0.721	0.721	0.692	0.865	0.692	0.750	0.471	0.702	0.356	0.808	0.779	0.702	0.625	0.981	0.817	0.567	0.808	0.606	**0.706**
MDM	0.635	0.779	0.692	0.788	0.779	0.731	0.663	0.769	0.635	0.673	0.625	0.635	0.413	0.692	0.788	0.721	0.538	0.962	0.760	0.548	0.740	0.606	**0.690**
TS+LR	0.731	0.808	0.712	0.769	0.760	0.750	0.663	0.846	0.692	0.750	0.500	0.731	0.404	0.827	0.750	0.740	0.683	0.971	0.846	0.606	0.827	0.635	**0.727**
D-ConvNet	0.627	0.855	0.806	0.700	0.760	0.641	0.559	0.790	0.691	0.734	0.665	0.703	0.767	0.754	0.518	0.520	0.748	0.952	0.868	0.596	0.765	0.615	**0.711**
S-ConvNet	0.681	0.896	0.857	0.775	0.835	0.684	0.670	0.874	0.788	0.673	0.819	0.712	0.732	0.876	0.682	0.798	0.855	1.000	0.856	0.661	0.733	0.652	**0.778**
EEGNet-2.32	0.729	0.847	0.796	0.771	0.732	0.768	0.702	0.906	0.715	0.675	0.759	0.788	0.835	0.867	0.769	0.785	0.883	1.000	0.896	0.596	0.798	0.618	**0.784**
EEGNet-4.8	0.806	0.858	0.885	0.796	0.816	0.727	0.683	0.780	0.769	0.764	0.827	0.831	0.885	0.907	0.717	0.797	0.845	1.000	0.830	0.699	0.796	0.661	**0.804**

Table A.33: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe from left side of the brain for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.760	0.750	0.740	0.644	0.750	0.712	0.740	0.827	0.837	0.635	0.490	0.760	0.346	0.721	0.702	0.769	0.663	0.923	0.913	0.625	0.731	0.442	**0.704**
MDM	0.702	0.740	0.721	0.663	0.788	0.654	0.702	0.827	0.663	0.692	0.548	0.712	0.500	0.817	0.712	0.769	0.500	0.865	0.760	0.615	0.683	0.481	**0.687**
TS+LR	0.750	0.760	0.712	0.663	0.769	0.750	0.740	0.865	0.808	0.683	0.558	0.721	0.433	0.827	0.721	0.788	0.596	0.942	0.913	0.663	0.798	0.413	**0.722**
D-ConvNet	0.789	0.779	0.892	0.819	0.683	0.614	0.720	0.710	0.818	0.624	0.923	0.790	0.661	0.887	0.873	0.699	0.875	0.944	0.912	0.872	0.695	0.615	**0.781**
S-ConvNet	0.710	0.768	0.873	0.663	0.796	0.663	0.665	0.844	0.810	0.662	0.758	0.839	0.632	0.903	0.855	0.845	0.816	0.973	0.981	0.741	0.761	0.798	**0.789**
EEGNet-2.32	0.721	0.846	0.883	0.813	0.751	0.719	0.761	0.894	0.887	0.779	0.914	0.816	0.625	0.933	0.875	0.834	0.885	1.000	0.963	0.788	0.806	0.719	**0.828**
EEGNet-4.8	0.836	0.885	0.901	0.870	0.771	0.717	0.655	0.845	0.905	0.730	0.902	0.855	0.726	0.944	0.905	0.844	0.895	1.000	0.953	0.795	0.847	0.702	**0.840**

Table A.34: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe from left side of the brain for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.721	0.731	0.731	0.596	0.663	0.702	0.731	0.817	0.788	0.654	0.490	0.731	0.346	0.760	0.702	0.779	0.702	0.923	0.942	0.596	0.750	0.481	**0.697**
MDM	0.750	0.673	0.731	0.635	0.769	0.654	0.702	0.808	0.673	0.712	0.558	0.712	0.481	0.817	0.721	0.788	0.500	0.837	0.769	0.625	0.721	0.481	**0.687**
TS+LR	0.721	0.721	0.721	0.606	0.740	0.750	0.731	0.875	0.798	0.683	0.587	0.702	0.394	0.817	0.721	0.788	0.654	0.913	0.913	0.702	0.798	0.452	**0.718**
D-ConvNet	0.866	0.776	0.863	0.807	0.709	0.621	0.702	0.757	0.732	0.602	0.913	0.865	0.651	0.865	0.915	0.757	0.886	0.907	0.961	0.843	0.707	0.624	**0.788**
S-ConvNet	0.662	0.779	0.882	0.644	0.806	0.633	0.665	0.826	0.803	0.683	0.768	0.840	0.615	0.894	0.855	0.807	0.835	0.973	0.971	0.720	0.741	0.759	**0.780**
EEGNet-2.32	0.758	0.825	0.873	0.829	0.720	0.750	0.752	0.855	0.868	0.789	0.865	0.825	0.662	0.933	0.875	0.825	0.904	1.000	0.944	0.817	0.836	0.756	**0.830**
EEGNet-4.8	0.730	0.815	0.882	0.811	0.827	0.708	0.685	0.855	0.878	0.757	0.902	0.834	0.784	0.935	0.933	0.845	0.913	1.000	0.962	0.824	0.835	0.711	**0.837**

Table A.35: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe from left side of the brain for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.663	0.712	0.721	0.702	0.712	0.654	0.769	0.808	0.692	0.663	0.452	0.712	0.490	0.731	0.712	0.750	0.644	0.913	0.904	0.683	0.798	0.452	**0.697**
MDM	0.654	0.721	0.769	0.663	0.769	0.606	0.721	0.808	0.625	0.615	0.567	0.663	0.452	0.837	0.779	0.788	0.500	0.846	0.760	0.567	0.712	0.567	**0.681**
TS+LR	0.683	0.692	0.760	0.654	0.740	0.712	0.760	0.808	0.740	0.654	0.529	0.663	0.529	0.798	0.712	0.779	0.548	0.904	0.885	0.615	0.817	0.500	**0.704**
D-ConvNet	0.700	0.826	0.853	0.682	0.689	0.613	0.721	0.703	0.789	0.623	0.808	0.694	0.587	0.728	0.732	0.582	0.616	0.962	0.884	0.569	0.700	0.595	**0.712**
S-ConvNet	0.721	0.769	0.798	0.711	0.739	0.625	0.672	0.798	0.760	0.649	0.798	0.701	0.674	0.875	0.815	0.844	0.692	0.963	0.963	0.745	0.818	0.714	**0.766**
EEGNet-2.32	0.670	0.809	0.786	0.810	0.760	0.726	0.655	0.762	0.789	0.721	0.816	0.760	0.653	0.904	0.877	0.874	0.856	1.000	0.925	0.658	0.819	0.749	**0.790**
EEGNet-4.8	0.741	0.786	0.835	0.712	0.731	0.612	0.685	0.808	0.869	0.701	0.884	0.787	0.670	0.933	0.828	0.845	0.702	1.000	0.935	0.705	0.864	0.675	**0.787**

Table A.36: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe from left side of the brain for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.721	0.798	0.721	0.712	0.750	0.644	0.779	0.827	0.750	0.635	0.510	0.712	0.490	0.702	0.702	0.721	0.596	0.933	0.875	0.635	0.760	0.462	**0.701**
MDM	0.606	0.740	0.750	0.692	0.798	0.615	0.731	0.798	0.635	0.615	0.558	0.654	0.442	0.808	0.731	0.769	0.500	0.846	0.760	0.577	0.683	0.577	**0.677**
TS+LR	0.702	0.769	0.731	0.740	0.769	0.683	0.740	0.837	0.769	0.615	0.510	0.663	0.587	0.808	0.740	0.740	0.577	0.942	0.885	0.606	0.846	0.490	**0.716**
D-ConvNet	0.675	0.808	0.844	0.799	0.683	0.642	0.740	0.695	0.828	0.692	0.810	0.618	0.651	0.660	0.710	0.624	0.724	0.945	0.954	0.645	0.717	0.617	**0.731**
S-ConvNet	0.699	0.836	0.845	0.741	0.723	0.625	0.663	0.770	0.770	0.622	0.798	0.750	0.634	0.855	0.795	0.844	0.702	0.972	0.944	0.685	0.809	0.697	**0.763**
EEGNet-2.32	0.733	0.855	0.835	0.821	0.734	0.697	0.675	0.752	0.866	0.770	0.808	0.721	0.595	0.893	0.887	0.854	0.820	1.000	0.954	0.659	0.819	0.682	**0.792**
EEGNet-4.8	0.877	0.885	0.817	0.830	0.752	0.583	0.714	0.807	0.894	0.767	0.903	0.819	0.644	0.933	0.847	0.875	0.749	1.000	0.944	0.752	0.829	0.663	**0.813**

Table A.37: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.683	0.740	0.683	0.760	0.519	0.558	0.577	0.827	0.625	0.625	0.577	0.644	0.433	0.750	0.731	0.740	0.558	0.875	0.769	0.558	0.635	0.538	**0.655**
MDM	0.702	0.644	0.673	0.740	0.490	0.490	0.606	0.750	0.596	0.644	0.596	0.625	0.433	0.663	0.712	0.779	0.510	0.885	0.750	0.538	0.683	0.385	**0.632**
TS+LR	0.779	0.731	0.683	0.750	0.510	0.548	0.596	0.798	0.606	0.635	0.567	0.663	0.413	0.721	0.712	0.779	0.538	0.885	0.731	0.606	0.683	0.510	**0.656**
D-ConvNet	0.809	0.869	0.835	0.767	0.663	0.589	0.589	0.864	0.693	0.605	0.638	0.722	0.737	0.799	0.692	0.661	0.837	0.895	0.883	0.615	0.586	0.674	**0.728**
S-ConvNet	0.749	0.866	0.836	0.767	0.683	0.587	0.642	0.874	0.667	0.589	0.734	0.722	0.719	0.815	0.702	0.768	0.827	0.905	0.816	0.683	0.646	0.645	**0.738**
EEGNet-2.32	0.775	0.867	0.815	0.806	0.644	0.548	0.634	0.845	0.684	0.665	0.744	0.721	0.843	0.835	0.709	0.795	0.808	0.953	0.882	0.639	0.696	0.635	**0.752**
EEGNet-4.8	0.750	0.916	0.866	0.804	0.701	0.624	0.652	0.856	0.683	0.685	0.712	0.749	0.863	0.827	0.670	0.767	0.818	0.962	0.893	0.670	0.664	0.665	**0.764**

Table A.38: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.673	0.731	0.740	0.750	0.500	0.567	0.577	0.846	0.596	0.615	0.596	0.644	0.413	0.798	0.712	0.721	0.529	0.875	0.740	0.567	0.663	0.538	**0.654**
MDM	0.702	0.625	0.721	0.750	0.490	0.490	0.615	0.750	0.596	0.654	0.596	0.635	0.433	0.712	0.721	0.779	0.519	0.875	0.750	0.538	0.712	0.394	**0.639**
TS+LR	0.779	0.721	0.721	0.750	0.519	0.548	0.615	0.808	0.596	0.663	0.567	0.654	0.413	0.750	0.712	0.788	0.558	0.865	0.769	0.606	0.683	0.510	**0.663**
D-ConvNet	0.827	0.799	0.825	0.775	0.635	0.607	0.558	0.844	0.703	0.605	0.675	0.665	0.650	0.798	0.682	0.690	0.758	0.905	0.846	0.587	0.566	0.653	**0.712**
S-ConvNet	0.760	0.828	0.767	0.750	0.674	0.588	0.585	0.865	0.678	0.609	0.675	0.694	0.682	0.814	0.691	0.779	0.781	0.884	0.826	0.605	0.696	0.645	**0.722**
EEGNet-2.32	0.757	0.808	0.835	0.777	0.644	0.557	0.633	0.874	0.731	0.643	0.732	0.741	0.699	0.855	0.767	0.777	0.702	0.942	0.853	0.633	0.638	0.702	**0.741**
EEGNet-4.8	0.761	0.827	0.836	0.815	0.691	0.604	0.623	0.846	0.693	0.695	0.722	0.741	0.717	0.827	0.727	0.777	0.769	0.961	0.864	0.595	0.664	0.664	**0.746**

Table A.39: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.654	0.702	0.683	0.788	0.404	0.567	0.625	0.779	0.692	0.635	0.529	0.635	0.462	0.712	0.712	0.740	0.510	0.827	0.721	0.500	0.654	0.644	**0.644**
MDM	0.606	0.606	0.673	0.760	0.423	0.481	0.635	0.750	0.596	0.625	0.596	0.615	0.365	0.596	0.721	0.779	0.481	0.856	0.702	0.567	0.692	0.529	**0.621**
TS+LR	0.663	0.654	0.692	0.779	0.471	0.548	0.606	0.760	0.654	0.635	0.490	0.663	0.413	0.635	0.702	0.788	0.529	0.846	0.750	0.567	0.654	0.596	**0.641**
D-ConvNet	0.643	0.750	0.864	0.698	0.587	0.575	0.595	0.828	0.625	0.598	0.608	0.625	0.575	0.704	0.661	0.671	0.604	0.922	0.785	0.491	0.588	0.720	**0.669**
S-ConvNet	0.748	0.828	0.816	0.729	0.605	0.595	0.701	0.865	0.687	0.596	0.745	0.692	0.645	0.771	0.672	0.770	0.596	0.856	0.808	0.635	0.610	0.656	**0.710**
EEGNet-2.32	0.796	0.779	0.749	0.723	0.615	0.585	0.664	0.801	0.723	0.547	0.684	0.702	0.635	0.685	0.712	0.732	0.711	0.923	0.768	0.555	0.674	0.569	**0.697**
EEGNet-4.8	0.770	0.731	0.846	0.733	0.645	0.605	0.680	0.770	0.703	0.635	0.695	0.714	0.635	0.720	0.748	0.769	0.732	0.951	0.749	0.571	0.665	0.711	**0.717**

Table A.40: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.654	0.721	0.673	0.788	0.404	0.587	0.625	0.760	0.663	0.644	0.519	0.635	0.462	0.615	0.702	0.692	0.558	0.885	0.712	0.529	0.654	0.615	**0.641**
MDM	0.567	0.625	0.644	0.740	0.433	0.500	0.625	0.760	0.606	0.635	0.615	0.596	0.356	0.606	0.721	0.760	0.538	0.875	0.702	0.558	0.683	0.529	**0.622**
TS+LR	0.644	0.692	0.673	0.769	0.452	0.548	0.606	0.750	0.625	0.644	0.500	0.663	0.394	0.644	0.712	0.769	0.577	0.865	0.740	0.567	0.673	0.606	**0.642**
D-ConvNet	0.703	0.831	0.797	0.669	0.519	0.604	0.647	0.781	0.610	0.586	0.599	0.645	0.787	0.694	0.613	0.682	0.713	0.895	0.728	0.618	0.598	0.678	**0.682**
S-ConvNet	0.747	0.849	0.788	0.748	0.595	0.578	0.692	0.826	0.725	0.637	0.752	0.703	0.730	0.760	0.680	0.760	0.788	0.905	0.771	0.675	0.607	0.683	**0.727**
EEGNet-2.32	0.778	0.859	0.846	0.721	0.588	0.615	0.634	0.820	0.683	0.584	0.655	0.701	0.854	0.740	0.700	0.720	0.829	0.943	0.825	0.632	0.636	0.655	**0.728**
EEGNet-4.8	0.808	0.898	0.872	0.762	0.655	0.614	0.711	0.750	0.703	0.615	0.638	0.725	0.835	0.792	0.680	0.807	0.817	0.933	0.785	0.621	0.605	0.703	**0.742**

Table A.41: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.615	0.510	0.577	0.442	0.740	0.654	0.644	0.740	0.673	0.712	0.538	0.538	0.519	0.788	0.663	0.683	0.548	0.981	0.721	0.529	0.760	0.346	**0.633**
MDM	0.567	0.462	0.577	0.375	0.760	0.673	0.635	0.731	0.635	0.692	0.548	0.558	0.519	0.663	0.673	0.712	0.519	0.990	0.750	0.587	0.769	0.500	**0.632**
TS+LR	0.577	0.529	0.587	0.385	0.760	0.740	0.692	0.731	0.654	0.712	0.577	0.567	0.558	0.760	0.683	0.692	0.519	0.990	0.731	0.538	0.731	0.452	**0.644**
D-ConvNet	0.615	0.539	0.575	0.490	0.722	0.544	0.538	0.672	0.635	0.724	0.560	0.433	0.529	0.766	0.718	0.615	0.549	1.000	0.738	0.440	0.634	0.633	**0.621**
S-ConvNet	0.481	0.728	0.596	0.442	0.721	0.689	0.575	0.578	0.655	0.598	0.616	0.435	0.497	0.815	0.661	0.786	0.528	0.991	0.706	0.575	0.691	0.606	**0.635**
EEGNet-2.32	0.567	0.624	0.579	0.546	0.732	0.688	0.623	0.605	0.664	0.734	0.525	0.490	0.537	0.817	0.788	0.716	0.645	1.000	0.768	0.613	0.672	0.693	**0.665**
EEGNet-4.8	0.577	0.618	0.508	0.548	0.732	0.736	0.587	0.645	0.684	0.732	0.641	0.587	0.462	0.865	0.778	0.728	0.548	1.000	0.758	0.559	0.645	0.614	**0.662**

Table A.42: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.577	0.548	0.596	0.452	0.740	0.683	0.635	0.731	0.683	0.692	0.548	0.538	0.510	0.788	0.673	0.740	0.548	0.981	0.721	0.519	0.712	0.356	**0.635**
MDM	0.548	0.471	0.596	0.404	0.740	0.673	0.635	0.721	0.635	0.673	0.548	0.567	0.510	0.673	0.683	0.721	0.519	0.990	0.750	0.587	0.769	0.510	**0.633**
TS+LR	0.596	0.538	0.567	0.394	0.740	0.721	0.683	0.721	0.673	0.702	0.606	0.567	0.558	0.740	0.654	0.721	0.500	0.971	0.721	0.529	0.760	0.433	**0.641**
D-ConvNet	0.537	0.556	0.516	0.451	0.663	0.630	0.510	0.576	0.578	0.610	0.570	0.395	0.548	0.743	0.661	0.584	0.576	0.991	0.670	0.510	0.693	0.585	**0.598**
S-ConvNet	0.500	0.644	0.573	0.403	0.681	0.698	0.537	0.578	0.617	0.523	0.588	0.425	0.506	0.805	0.700	0.718	0.537	0.972	0.726	0.624	0.701	0.597	**0.621**
EEGNet-2.32	0.595	0.653	0.558	0.586	0.761	0.688	0.653	0.571	0.666	0.685	0.557	0.481	0.516	0.808	0.796	0.678	0.685	0.963	0.768	0.424	0.719	0.665	**0.658**
EEGNet-4.8	0.519	0.638	0.539	0.556	0.741	0.746	0.595	0.615	0.663	0.666	0.623	0.539	0.480	0.845	0.780	0.708	0.568	0.962	0.748	0.495	0.683	0.644	**0.652**

Table A.43: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.635	0.500	0.538	0.529	0.731	0.750	0.596	0.692	0.577	0.673	0.548	0.567	0.404	0.769	0.721	0.731	0.538	0.962	0.740	0.683	0.808	0.519	**0.646**
MDM	0.615	0.433	0.558	0.394	0.760	0.721	0.625	0.673	0.596	0.558	0.587	0.558	0.442	0.683	0.721	0.683	0.510	0.962	0.760	0.596	0.779	0.625	**0.629**
TS+LR	0.644	0.519	0.587	0.442	0.750	0.750	0.587	0.692	0.606	0.663	0.567	0.587	0.423	0.779	0.712	0.731	0.500	0.952	0.712	0.596	0.846	0.606	**0.648**
D-ConvNet	0.570	0.550	0.595	0.540	0.635	0.575	0.508	0.517	0.518	0.665	0.454	0.537	0.499	0.768	0.585	0.540	0.675	0.963	0.679	0.546	0.692	0.585	**0.600**
S-ConvNet	0.605	0.637	0.570	0.528	0.704	0.718	0.536	0.560	0.632	0.645	0.604	0.425	0.532	0.808	0.708	0.640	0.559	0.982	0.707	0.602	0.747	0.672	**0.642**
EEGNet-2.32	0.615	0.565	0.605	0.635	0.731	0.632	0.471	0.679	0.615	0.748	0.598	0.529	0.461	0.836	0.632	0.688	0.749	0.982	0.757	0.602	0.673	0.663	**0.658**
EEGNet-4.8	0.531	0.567	0.570	0.548	0.681	0.705	0.520	0.623	0.605	0.731	0.538	0.500	0.475	0.866	0.672	0.727	0.548	0.991	0.777	0.584	0.722	0.662	**0.643**

Table A.44: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.683	0.510	0.548	0.529	0.731	0.750	0.596	0.702	0.567	0.731	0.529	0.577	0.423	0.779	0.702	0.683	0.529	0.971	0.740	0.635	0.798	0.529	**0.647**
MDM	0.615	0.433	0.567	0.394	0.769	0.712	0.635	0.673	0.596	0.596	0.587	0.538	0.462	0.673	0.731	0.683	0.510	0.971	0.760	0.596	0.798	0.635	**0.633**
TS+LR	0.673	0.500	0.596	0.423	0.760	0.750	0.625	0.683	0.596	0.692	0.567	0.567	0.442	0.808	0.731	0.692	0.529	0.971	0.731	0.596	0.827	0.615	**0.653**
D-ConvNet	0.568	0.499	0.587	0.541	0.713	0.566	0.547	0.624	0.645	0.724	0.494	0.492	0.529	0.749	0.625	0.574	0.605	1.000	0.718	0.558	0.662	0.584	**0.618**
S-ConvNet	0.595	0.662	0.606	0.498	0.713	0.698	0.518	0.615	0.585	0.664	0.645	0.455	0.550	0.818	0.711	0.740	0.530	0.991	0.685	0.621	0.680	0.665	**0.648**
EEGNet-2.32	0.583	0.655	0.539	0.556	0.713	0.623	0.461	0.615	0.605	0.753	0.565	0.500	0.488	0.828	0.653	0.677	0.693	1.000	0.749	0.594	0.655	0.674	**0.644**
EEGNet-4.8	0.587	0.557	0.559	0.577	0.701	0.706	0.497	0.683	0.692	0.742	0.575	0.490	0.510	0.895	0.747	0.696	0.596	1.000	0.739	0.538	0.665	0.662	**0.655**

Table A.45: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe from left side of the brain for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.606	0.538	0.577	0.663	0.779	0.673	0.673	0.702	0.750	0.683	0.548	0.490	0.365	0.731	0.644	0.721	0.558	0.971	0.731	0.577	0.731	0.385	**0.641**
MDM	0.587	0.519	0.606	0.577	0.788	0.654	0.673	0.673	0.663	0.654	0.558	0.548	0.423	0.779	0.663	0.712	0.500	0.894	0.750	0.567	0.702	0.538	**0.638**
TS+LR	0.587	0.529	0.596	0.615	0.769	0.673	0.654	0.702	0.721	0.663	0.596	0.471	0.356	0.788	0.663	0.712	0.510	0.952	0.760	0.577	0.740	0.490	**0.642**
D-ConvNet	0.626	0.557	0.558	0.715	0.733	0.572	0.531	0.557	0.733	0.725	0.565	0.539	0.557	0.707	0.671	0.755	0.605	0.991	0.719	0.577	0.722	0.587	**0.650**
S-ConvNet	0.655	0.623	0.555	0.676	0.768	0.677	0.519	0.540	0.818	0.606	0.615	0.526	0.567	0.787	0.691	0.807	0.575	0.990	0.678	0.475	0.685	0.695	**0.660**
EEGNet-2.32	0.635	0.560	0.499	0.754	0.752	0.612	0.655	0.596	0.869	0.738	0.612	0.441	0.534	0.825	0.653	0.785	0.723	1.000	0.758	0.462	0.663	0.583	**0.669**
EEGNet-4.8	0.685	0.680	0.595	0.744	0.742	0.612	0.619	0.543	0.819	0.781	0.574	0.606	0.557	0.826	0.651	0.834	0.605	1.000	0.776	0.633	0.721	0.653	**0.693**

Table A.46: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe from left side of the brain for MNS vs. MNS+MI discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.606	0.490	0.615	0.615	0.788	0.692	0.692	0.721	0.721	0.673	0.548	0.510	0.394	0.740	0.625	0.721	0.519	0.923	0.740	0.567	0.731	0.462	**0.641**
MDM	0.577	0.519	0.625	0.548	0.769	0.654	0.673	0.692	0.654	0.635	0.567	0.567	0.394	0.779	0.654	0.721	0.500	0.894	0.750	0.567	0.731	0.548	**0.637**
TS+LR	0.596	0.500	0.615	0.587	0.750	0.673	0.654	0.673	0.683	0.654	0.577	0.529	0.356	0.769	0.663	0.712	0.519	0.942	0.760	0.577	0.731	0.490	**0.637**
D-ConvNet	0.597	0.540	0.549	0.647	0.712	0.552	0.539	0.547	0.532	0.575	0.519	0.541	0.548	0.641	0.691	0.688	0.568	0.972	0.688	0.580	0.722	0.595	**0.616**
S-ConvNet	0.577	0.637	0.572	0.639	0.757	0.630	0.479	0.519	0.721	0.628	0.546	0.499	0.586	0.795	0.719	0.755	0.644	0.981	0.657	0.407	0.732	0.694	**0.644**
EEGNet-2.32	0.585	0.530	0.576	0.780	0.752	0.612	0.625	0.527	0.714	0.672	0.634	0.449	0.515	0.816	0.642	0.710	0.694	1.000	0.756	0.434	0.654	0.593	**0.649**
EEGNet-4.8	0.550	0.622	0.584	0.753	0.771	0.632	0.618	0.562	0.625	0.694	0.593	0.576	0.488	0.875	0.701	0.824	0.599	1.000	0.795	0.606	0.694	0.653	**0.673**

Table A.47: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe from left side of the brain for MNS vs. MNS+MI discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.702	0.510	0.606	0.663	0.779	0.644	0.644	0.663	0.721	0.635	0.567	0.625	0.442	0.731	0.635	0.673	0.500	0.952	0.731	0.692	0.779	0.567	**0.657**
MDM	0.615	0.471	0.548	0.567	0.779	0.596	0.635	0.644	0.596	0.577	0.577	0.673	0.413	0.779	0.673	0.750	0.500	0.837	0.750	0.587	0.779	0.606	**0.634**
TS+LR	0.625	0.538	0.567	0.663	0.760	0.663	0.683	0.654	0.683	0.606	0.548	0.625	0.404	0.760	0.654	0.673	0.510	0.913	0.740	0.635	0.769	0.606	**0.649**
D-ConvNet	0.569	0.518	0.480	0.696	0.662	0.497	0.491	0.595	0.576	0.628	0.606	0.548	0.526	0.630	0.625	0.706	0.606	0.971	0.717	0.550	0.684	0.556	**0.611**
S-ConvNet	0.615	0.615	0.624	0.715	0.625	0.621	0.506	0.481	0.710	0.548	0.662	0.546	0.558	0.806	0.615	0.708	0.605	0.963	0.678	0.482	0.713	0.626	**0.637**
EEGNet-2.32	0.655	0.607	0.567	0.763	0.740	0.646	0.500	0.516	0.771	0.769	0.555	0.498	0.488	0.788	0.586	0.718	0.695	1.000	0.766	0.595	0.684	0.642	**0.661**
EEGNet-4.8	0.591	0.606	0.595	0.666	0.702	0.554	0.470	0.474	0.703	0.732	0.632	0.569	0.507	0.807	0.579	0.804	0.587	1.000	0.785	0.606	0.684	0.641	**0.650**

Table A.48: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe from left side of the brain for MNS vs. MNS+MI discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.702	0.490	0.587	0.702	0.769	0.673	0.635	0.654	0.760	0.596	0.567	0.596	0.433	0.702	0.625	0.692	0.529	0.971	0.731	0.721	0.760	0.587	**0.658**
MDM	0.567	0.490	0.500	0.587	0.798	0.596	0.625	0.663	0.606	0.567	0.577	0.644	0.423	0.788	0.673	0.731	0.500	0.846	0.750	0.587	0.779	0.606	**0.632**
TS+LR	0.644	0.529	0.548	0.702	0.760	0.673	0.673	0.663	0.712	0.606	0.538	0.644	0.413	0.779	0.673	0.683	0.519	0.942	0.740	0.635	0.779	0.587	**0.656**
D-ConvNet	0.675	0.595	0.575	0.724	0.683	0.476	0.520	0.608	0.740	0.695	0.595	0.572	0.537	0.667	0.577	0.746	0.578	0.990	0.758	0.557	0.683	0.575	**0.642**
S-ConvNet	0.695	0.662	0.615	0.754	0.675	0.609	0.527	0.473	0.768	0.585	0.605	0.587	0.568	0.836	0.605	0.815	0.528	0.990	0.686	0.598	0.734	0.646	**0.662**
EEGNet-2.32	0.625	0.625	0.587	0.773	0.712	0.617	0.535	0.534	0.800	0.762	0.603	0.486	0.507	0.780	0.519	0.833	0.733	1.000	0.746	0.624	0.715	0.660	**0.672**
EEGNet-4.8	0.685	0.699	0.645	0.714	0.742	0.584	0.527	0.546	0.801	0.818	0.574	0.628	0.481	0.837	0.530	0.815	0.605	1.000	0.738	0.635	0.702	0.613	**0.678**

Appendix B

Classification results for MI vs. rest task using different classifiers

In this section, classification accuracy results obtained by different classifiers (CSP+LDA, MDRM, TS+LR, D-ConvNet, S-ConvNet, EEGNet-2.32, EEGNet-4.8) for 22 subjects from dataset introduced in Section 4.3 are presented. Results are represented for 12 different channel layouts for MI vs. rest discriminating task. Signals filtered in 4 different frequency bands, results averaged across 10 folds of cross-validation.

Table B.1: Classification accuracy obtained by different classifiers for 22 subjects using 128 channel layout for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.615	0.702	0.760	0.779	0.577	0.500	0.673	0.750	0.615	0.769	0.692	0.740	0.529	0.779	0.615	0.683	0.663	0.587	0.923	0.808	0.779	0.558	**0.686**
MDM	0.548	0.692	0.644	0.587	0.596	0.538	0.654	0.683	0.663	0.731	0.663	0.663	0.615	0.644	0.683	0.663	0.663	0.548	0.894	0.663	0.779	0.577	**0.654**
TS+LR	0.558	0.721	0.769	0.808	0.606	0.587	0.644	0.788	0.721	0.808	0.692	0.683	0.587	0.740	0.827	0.731	0.740	0.567	0.923	0.721	0.760	0.558	**0.706**
D-ConvNet	0.471	0.614	0.651	0.663	0.597	0.535	0.501	0.584	0.616	0.752	0.672	0.693	0.547	0.575	0.511	0.709	0.592	0.615	0.895	0.625	0.520	0.595	**0.615**
S-ConvNet	0.540	0.711	0.703	0.522	0.556	0.567	0.635	0.540	0.684	0.624	0.651	0.647	0.530	0.724	0.636	0.825	0.655	0.598	0.827	0.556	0.797	0.537	**0.639**
EEGNet-2.32	0.501	0.658	0.746	0.510	0.749	0.490	0.713	0.598	0.635	0.784	0.637	0.749	0.655	0.731	0.714	0.789	0.585	0.684	0.932	0.706	0.777	0.542	**0.677**
EEGNet-4.8	0.461	0.700	0.789	0.547	0.683	0.609	0.730	0.710	0.640	0.743	0.740	0.730	0.664	0.778	0.790	0.815	0.645	0.675	0.942	0.698	0.777	0.502	**0.699**

Table B.2: Classification accuracy obtained by different classifiers for 22 subjects using 128 channel layout for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.615	0.673	0.712	0.769	0.587	0.510	0.558	0.731	0.587	0.798	0.654	0.740	0.538	0.673	0.625	0.644	0.615	0.538	0.923	0.817	0.769	0.500	**0.663**
MDM	0.529	0.654	0.654	0.606	0.596	0.529	0.654	0.683	0.654	0.731	0.663	0.635	0.615	0.606	0.712	0.683	0.663	0.587	0.856	0.673	0.779	0.587	**0.652**
TS+LR	0.500	0.702	0.740	0.817	0.596	0.567	0.654	0.769	0.721	0.769	0.712	0.625	0.558	0.712	0.808	0.702	0.731	0.654	0.875	0.712	0.808	0.538	**0.694**
D-ConvNet	0.541	0.645	0.693	0.644	0.589	0.500	0.552	0.595	0.609	0.742	0.614	0.673	0.566	0.599	0.528	0.623	0.579	0.597	0.914	0.637	0.626	0.546	**0.619**
S-ConvNet	0.531	0.711	0.685	0.482	0.558	0.567	0.655	0.561	0.714	0.577	0.623	0.665	0.578	0.705	0.683	0.826	0.684	0.635	0.845	0.644	0.788	0.557	**0.649**
EEGNet-2.32	0.501	0.665	0.740	0.519	0.732	0.463	0.690	0.625	0.654	0.742	0.696	0.760	0.665	0.766	0.646	0.738	0.692	0.704	0.952	0.777	0.818	0.519	**0.685**
EEGNet-4.8	0.577	0.759	0.751	0.548	0.693	0.596	0.705	0.672	0.636	0.743	0.694	0.750	0.693	0.731	0.742	0.785	0.724	0.714	0.922	0.708	0.766	0.635	**0.706**

Table B.3: Classification accuracy obtained by different classifiers for 22 subjects using 128 channel layout for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.490	0.663	0.673	0.779	0.625	0.538	0.596	0.769	0.567	0.750	0.606	0.673	0.548	0.702	0.615	0.692	0.635	0.596	0.923	0.808	0.750	0.519	**0.660**
MDM	0.596	0.702	0.673	0.683	0.663	0.529	0.654	0.702	0.683	0.731	0.683	0.692	0.615	0.615	0.740	0.683	0.692	0.615	0.875	0.654	0.750	0.558	**0.672**
TS+LR	0.606	0.712	0.760	0.837	0.673	0.558	0.654	0.760	0.750	0.760	0.731	0.721	0.606	0.721	0.788	0.750	0.702	0.635	0.904	0.721	0.750	0.606	**0.714**
D-ConvNet	0.509	0.602	0.627	0.606	0.550	0.567	0.543	0.546	0.668	0.712	0.596	0.599	0.624	0.586	0.561	0.594	0.552	0.597	0.883	0.654	0.627	0.538	**0.606**
S-ConvNet	0.518	0.695	0.617	0.548	0.634	0.607	0.686	0.605	0.605	0.702	0.635	0.615	0.675	0.692	0.643	0.827	0.614	0.663	0.818	0.595	0.758	0.537	**0.649**
EEGNet-2.32	0.599	0.683	0.789	0.559	0.692	0.526	0.733	0.710	0.541	0.750	0.585	0.751	0.663	0.718	0.720	0.768	0.643	0.702	0.912	0.790	0.795	0.558	**0.690**
EEGNet-4.8	0.558	0.722	0.802	0.550	0.645	0.550	0.778	0.663	0.655	0.759	0.624	0.720	0.705	0.731	0.722	0.779	0.600	0.728	0.951	0.729	0.856	0.548	**0.699**

Table B.4: Classification accuracy obtained by different classifiers for 22 subjects using 128 channel layout for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.587	0.740	0.654	0.817	0.644	0.548	0.615	0.760	0.635	0.740	0.644	0.663	0.596	0.779	0.577	0.740	0.635	0.596	0.933	0.808	0.760	0.510	**0.681**
MDM	0.558	0.692	0.673	0.635	0.587	0.538	0.654	0.692	0.663	0.712	0.663	0.663	0.596	0.654	0.692	0.683	0.663	0.558	0.894	0.683	0.779	0.548	**0.658**
TS+LR	0.567	0.702	0.769	0.856	0.654	0.538	0.663	0.750	0.721	0.779	0.712	0.750	0.596	0.769	0.817	0.702	0.692	0.548	0.913	0.731	0.788	0.615	**0.711**
D-ConvNet	0.532	0.647	0.605	0.684	0.599	0.604	0.512	0.507	0.668	0.759	0.577	0.654	0.644	0.588	0.510	0.672	0.580	0.586	0.874	0.645	0.586	0.595	**0.620**
S-ConvNet	0.575	0.703	0.694	0.530	0.586	0.526	0.625	0.685	0.656	0.713	0.641	0.673	0.617	0.763	0.663	0.790	0.578	0.664	0.836	0.558	0.749	0.635	**0.657**
EEGNet-2.32	0.519	0.685	0.767	0.539	0.711	0.595	0.730	0.674	0.615	0.684	0.595	0.715	0.694	0.719	0.722	0.775	0.566	0.587	0.970	0.693	0.704	0.523	**0.672**
EEGNet-4.8	0.551	0.675	0.780	0.548	0.666	0.549	0.691	0.663	0.578	0.762	0.691	0.721	0.701	0.723	0.639	0.844	0.633	0.637	0.951	0.710	0.867	0.577	**0.689**

Table B.5: Classification accuracy obtained by different classifiers for 22 subjects using 47 channel layout for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.615	0.712	0.692	0.788	0.635	0.500	0.625	0.779	0.587	0.712	0.779	0.606	0.567	0.692	0.625	0.615	0.606	0.558	0.952	0.837	0.731	0.433	**0.666**
MDM	0.558	0.702	0.731	0.577	0.625	0.471	0.663	0.817	0.625	0.644	0.673	0.702	0.606	0.721	0.702	0.721	0.702	0.500	0.971	0.750	0.769	0.519	**0.670**
TS+LR	0.654	0.731	0.750	0.846	0.654	0.587	0.702	0.788	0.635	0.731	0.721	0.731	0.548	0.808	0.769	0.721	0.740	0.558	0.971	0.846	0.788	0.606	**0.722**
D-ConvNet	0.471	0.733	0.800	0.607	0.615	0.510	0.541	0.559	0.589	0.752	0.597	0.722	0.623	0.531	0.540	0.742	0.498	0.614	0.931	0.729	0.857	0.566	**0.642**
S-ConvNet	0.626	0.790	0.778	0.479	0.655	0.478	0.668	0.642	0.635	0.729	0.700	0.657	0.673	0.674	0.587	0.838	0.666	0.627	0.951	0.740	0.758	0.626	**0.681**
EEGNet-2.32	0.565	0.712	0.751	0.490	0.750	0.549	0.760	0.703	0.625	0.829	0.682	0.759	0.664	0.751	0.750	0.795	0.634	0.597	0.931	0.816	0.887	0.606	**0.709**
EEGNet-4.8	0.603	0.778	0.770	0.502	0.644	0.558	0.770	0.711	0.567	0.753	0.656	0.700	0.692	0.731	0.733	0.816	0.611	0.704	0.941	0.815	0.827	0.590	**0.703**

Table B.6: Classification accuracy obtained by different classifiers for 22 subjects using 47 channel layout for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.596	0.731	0.663	0.808	0.692	0.490	0.654	0.740	0.596	0.712	0.769	0.683	0.567	0.673	0.644	0.654	0.625	0.529	0.971	0.856	0.750	0.471	**0.676**
MDM	0.538	0.731	0.712	0.577	0.654	0.500	0.673	0.817	0.644	0.721	0.663	0.721	0.615	0.798	0.692	0.731	0.731	0.500	0.971	0.750	0.779	0.529	**0.684**
TS+LR	0.635	0.750	0.740	0.856	0.692	0.615	0.731	0.798	0.625	0.769	0.740	0.740	0.529	0.808	0.750	0.731	0.702	0.519	0.971	0.827	0.779	0.644	**0.725**
D-ConvNet	0.501	0.716	0.789	0.665	0.605	0.537	0.539	0.575	0.625	0.742	0.606	0.705	0.641	0.607	0.558	0.699	0.474	0.613	0.961	0.766	0.867	0.624	**0.655**
S-ConvNet	0.565	0.791	0.779	0.537	0.684	0.450	0.676	0.624	0.606	0.699	0.682	0.667	0.645	0.662	0.647	0.808	0.636	0.645	0.941	0.760	0.776	0.598	**0.676**
EEGNet-2.32	0.555	0.742	0.722	0.500	0.712	0.521	0.695	0.712	0.615	0.769	0.693	0.731	0.665	0.750	0.720	0.808	0.597	0.644	0.950	0.826	0.858	0.598	**0.699**
EEGNet-4.8	0.566	0.772	0.761	0.510	0.624	0.558	0.758	0.750	0.639	0.743	0.704	0.712	0.645	0.806	0.736	0.825	0.675	0.675	0.951	0.805	0.818	0.712	**0.716**

Table B.7: Classification accuracy obtained by different classifiers for 22 subjects using 47 channel layout for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.577	0.740	0.644	0.827	0.644	0.538	0.654	0.731	0.567	0.721	0.587	0.769	0.558	0.663	0.625	0.654	0.596	0.567	0.952	0.856	0.731	0.587	**0.672**
MDM	0.529	0.721	0.788	0.606	0.606	0.510	0.692	0.808	0.644	0.712	0.663	0.692	0.625	0.788	0.644	0.712	0.654	0.500	0.962	0.798	0.760	0.529	**0.679**
TS+LR	0.625	0.760	0.740	0.846	0.596	0.548	0.702	0.779	0.615	0.769	0.712	0.779	0.587	0.817	0.760	0.731	0.683	0.567	0.962	0.875	0.808	0.567	**0.719**
D-ConvNet	0.562	0.765	0.750	0.674	0.595	0.511	0.570	0.633	0.585	0.772	0.605	0.617	0.613	0.615	0.580	0.760	0.519	0.606	0.951	0.742	0.799	0.670	**0.659**
S-ConvNet	0.622	0.780	0.737	0.490	0.655	0.471	0.585	0.645	0.624	0.680	0.616	0.714	0.616	0.593	0.646	0.849	0.548	0.608	0.882	0.644	0.817	0.667	**0.659**
EEGNet-2.32	0.645	0.724	0.818	0.527	0.653	0.539	0.731	0.693	0.617	0.713	0.684	0.693	0.671	0.740	0.780	0.835	0.489	0.709	0.931	0.875	0.886	0.606	**0.707**
EEGNet-4.8	0.691	0.743	0.760	0.519	0.702	0.540	0.801	0.681	0.598	0.771	0.587	0.799	0.681	0.786	0.725	0.797	0.555	0.779	0.941	0.814	0.828	0.637	**0.715**

Table B.8: Classification accuracy obtained by different classifiers for 22 subjects using 47 channel layout for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.635	0.750	0.721	0.827	0.644	0.577	0.625	0.702	0.558	0.779	0.606	0.740	0.529	0.635	0.625	0.692	0.596	0.567	0.942	0.837	0.654	0.558	**0.673**
MDM	0.567	0.712	0.740	0.606	0.558	0.481	0.683	0.788	0.654	0.692	0.663	0.702	0.596	0.760	0.654	0.692	0.654	0.500	0.962	0.798	0.779	0.538	**0.672**
TS+LR	0.683	0.760	0.779	0.856	0.587	0.529	0.673	0.779	0.596	0.769	0.673	0.740	0.596	0.837	0.788	0.721	0.692	0.615	0.962	0.875	0.760	0.558	**0.719**
D-ConvNet	0.515	0.697	0.733	0.718	0.588	0.492	0.569	0.623	0.645	0.753	0.566	0.588	0.615	0.566	0.530	0.786	0.577	0.603	0.901	0.721	0.760	0.616	**0.644**
S-ConvNet	0.647	0.771	0.739	0.491	0.586	0.548	0.578	0.607	0.615	0.730	0.560	0.732	0.636	0.584	0.626	0.878	0.561	0.635	0.911	0.683	0.826	0.609	**0.662**
EEGNet-2.32	0.642	0.733	0.779	0.519	0.653	0.566	0.647	0.674	0.675	0.762	0.585	0.723	0.691	0.752	0.769	0.805	0.521	0.646	0.933	0.798	0.887	0.555	**0.696**
EEGNet-4.8	0.681	0.788	0.791	0.529	0.683	0.559	0.780	0.731	0.655	0.752	0.627	0.779	0.710	0.721	0.704	0.796	0.585	0.685	0.951	0.814	0.878	0.673	**0.721**

Table B.9: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.558	0.740	0.673	0.769	0.644	0.587	0.587	0.760	0.596	0.625	0.683	0.673	0.635	0.702	0.702	0.702	0.625	0.577	0.913	0.817	0.788	0.587	**0.679**
MDM	0.567	0.740	0.663	0.548	0.654	0.519	0.692	0.808	0.606	0.635	0.635	0.692	0.625	0.702	0.673	0.692	0.654	0.500	0.981	0.663	0.760	0.471	**0.658**
TS+LR	0.615	0.817	0.721	0.817	0.692	0.654	0.721	0.817	0.577	0.644	0.692	0.798	0.596	0.779	0.692	0.712	0.625	0.625	0.942	0.750	0.731	0.567	**0.708**
D-ConvNet	0.489	0.693	0.800	0.482	0.634	0.547	0.533	0.604	0.579	0.683	0.587	0.713	0.634	0.559	0.519	0.755	0.519	0.537	0.931	0.805	0.849	0.625	**0.640**
S-ConvNet	0.616	0.770	0.827	0.529	0.662	0.528	0.705	0.645	0.615	0.658	0.593	0.692	0.674	0.675	0.605	0.808	0.550	0.614	0.863	0.664	0.798	0.617	**0.669**
EEGNet-2.32	0.614	0.714	0.791	0.500	0.751	0.494	0.797	0.684	0.658	0.768	0.594	0.751	0.682	0.701	0.734	0.767	0.577	0.732	0.932	0.825	0.848	0.614	**0.706**
EEGNet-4.8	0.606	0.780	0.876	0.482	0.674	0.575	0.770	0.635	0.649	0.724	0.622	0.788	0.691	0.739	0.675	0.817	0.597	0.722	0.941	0.767	0.865	0.701	**0.713**

Table B.10: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.558	0.808	0.635	0.740	0.654	0.558	0.615	0.788	0.587	0.577	0.683	0.683	0.615	0.702	0.683	0.683	0.548	0.558	0.942	0.837	0.779	0.577	**0.673**
MDM	0.587	0.731	0.654	0.558	0.673	0.519	0.692	0.808	0.606	0.654	0.635	0.692	0.644	0.740	0.663	0.712	0.683	0.500	0.962	0.663	0.788	0.471	**0.665**
TS+LR	0.587	0.846	0.702	0.798	0.702	0.663	0.731	0.817	0.577	0.644	0.712	0.779	0.558	0.740	0.721	0.721	0.635	0.615	0.923	0.760	0.740	0.558	**0.706**
D-ConvNet	0.538	0.755	0.789	0.481	0.701	0.491	0.531	0.636	0.618	0.722	0.599	0.732	0.567	0.539	0.579	0.747	0.508	0.578	0.922	0.777	0.867	0.585	**0.648**
S-ConvNet	0.585	0.760	0.759	0.548	0.664	0.567	0.676	0.643	0.625	0.649	0.564	0.692	0.665	0.681	0.615	0.808	0.541	0.589	0.884	0.751	0.767	0.617	**0.666**
EEGNet-2.32	0.554	0.789	0.771	0.500	0.751	0.523	0.807	0.703	0.657	0.815	0.594	0.742	0.712	0.730	0.724	0.835	0.537	0.637	0.942	0.805	0.858	0.585	**0.708**
EEGNet-4.8	0.596	0.810	0.856	0.491	0.692	0.584	0.780	0.664	0.657	0.806	0.673	0.749	0.692	0.796	0.695	0.817	0.578	0.692	0.942	0.805	0.837	0.693	**0.723**

Table B.11: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.567	0.769	0.769	0.788	0.673	0.538	0.596	0.788	0.683	0.673	0.692	0.692	0.538	0.712	0.663	0.702	0.673	0.567	0.933	0.808	0.712	0.567	**0.687**
MDM	0.538	0.760	0.683	0.567	0.615	0.510	0.702	0.779	0.625	0.635	0.663	0.644	0.615	0.788	0.654	0.712	0.683	0.500	0.942	0.702	0.808	0.548	**0.667**
TS+LR	0.673	0.817	0.702	0.798	0.625	0.635	0.731	0.817	0.625	0.731	0.663	0.702	0.606	0.817	0.712	0.683	0.635	0.663	0.933	0.760	0.779	0.587	**0.713**
D-ConvNet	0.549	0.713	0.796	0.550	0.624	0.470	0.571	0.612	0.654	0.671	0.539	0.585	0.651	0.569	0.529	0.757	0.577	0.634	0.961	0.766	0.868	0.652	**0.650**
S-ConvNet	0.563	0.818	0.776	0.520	0.644	0.528	0.598	0.683	0.694	0.750	0.521	0.693	0.674	0.673	0.616	0.826	0.585	0.663	0.845	0.682	0.846	0.655	**0.675**
EEGNet-2.32	0.568	0.848	0.788	0.529	0.739	0.528	0.817	0.685	0.696	0.778	0.587	0.694	0.623	0.778	0.744	0.817	0.577	0.672	0.952	0.815	0.859	0.636	**0.715**
EEGNet-4.8	0.519	0.799	0.828	0.499	0.661	0.540	0.733	0.721	0.684	0.695	0.597	0.759	0.701	0.769	0.724	0.827	0.535	0.720	0.951	0.785	0.866	0.671	**0.708**

Table B.12: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.625	0.760	0.769	0.827	0.625	0.510	0.625	0.779	0.702	0.740	0.644	0.712	0.529	0.702	0.644	0.740	0.606	0.577	0.933	0.827	0.740	0.635	**0.693**
MDM	0.519	0.769	0.692	0.577	0.606	0.519	0.702	0.779	0.615	0.663	0.663	0.625	0.606	0.750	0.654	0.702	0.673	0.500	0.942	0.702	0.779	0.510	**0.661**
TS+LR	0.683	0.798	0.683	0.808	0.625	0.577	0.721	0.827	0.625	0.731	0.673	0.702	0.587	0.779	0.712	0.712	0.644	0.683	0.933	0.769	0.779	0.587	**0.711**
D-ConvNet	0.569	0.734	0.672	0.549	0.674	0.549	0.569	0.605	0.627	0.711	0.607	0.606	0.624	0.625	0.500	0.809	0.529	0.521	0.961	0.682	0.847	0.662	**0.647**
S-ConvNet	0.550	0.778	0.749	0.559	0.692	0.540	0.617	0.692	0.673	0.684	0.535	0.683	0.684	0.625	0.637	0.836	0.575	0.592	0.893	0.682	0.817	0.685	**0.672**
EEGNet-2.32	0.529	0.798	0.791	0.519	0.740	0.538	0.750	0.665	0.656	0.789	0.586	0.713	0.691	0.711	0.734	0.769	0.516	0.700	0.942	0.813	0.865	0.598	**0.701**
EEGNet-4.8	0.628	0.734	0.868	0.499	0.654	0.520	0.740	0.684	0.637	0.707	0.579	0.721	0.710	0.787	0.684	0.816	0.548	0.732	0.951	0.795	0.875	0.709	**0.708**

Table B.13: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.663	0.760	0.587	0.740	0.721	0.712	0.644	0.769	0.538	0.740	0.577	0.683	0.519	0.615	0.635	0.635	0.548	0.596	0.913	0.779	0.712	0.577	**0.667**
MDM	0.625	0.721	0.654	0.779	0.731	0.587	0.721	0.788	0.596	0.740	0.635	0.596	0.577	0.615	0.692	0.644	0.538	0.500	0.952	0.644	0.712	0.500	**0.661**
TS+LR	0.663	0.702	0.615	0.817	0.712	0.673	0.702	0.769	0.538	0.779	0.625	0.635	0.490	0.760	0.750	0.702	0.625	0.663	0.923	0.731	0.740	0.615	**0.692**
D-ConvNet	0.598	0.724	0.675	0.781	0.605	0.548	0.547	0.673	0.549	0.741	0.596	0.595	0.615	0.614	0.559	0.759	0.520	0.709	0.923	0.702	0.733	0.585	**0.652**
S-ConvNet	0.623	0.750	0.712	0.680	0.655	0.566	0.690	0.600	0.699	0.765	0.708	0.567	0.592	0.700	0.624	0.836	0.647	0.683	0.874	0.695	0.789	0.585	**0.684**
EEGNet-2.32	0.520	0.735	0.759	0.819	0.712	0.550	0.797	0.695	0.642	0.722	0.633	0.664	0.664	0.741	0.664	0.789	0.541	0.759	0.942	0.776	0.770	0.645	**0.706**
EEGNet-4.8	0.568	0.742	0.798	0.848	0.625	0.585	0.799	0.704	0.683	0.817	0.702	0.735	0.729	0.731	0.714	0.775	0.655	0.811	0.961	0.745	0.791	0.713	**0.738**

Table B.14: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.673	0.721	0.567	0.779	0.740	0.731	0.673	0.788	0.529	0.750	0.615	0.702	0.471	0.654	0.663	0.702	0.548	0.635	0.904	0.769	0.769	0.577	**0.680**
MDM	0.635	0.712	0.654	0.779	0.731	0.587	0.721	0.808	0.577	0.692	0.635	0.615	0.587	0.683	0.692	0.654	0.548	0.500	0.952	0.644	0.740	0.500	**0.666**
TS+LR	0.654	0.702	0.577	0.817	0.721	0.663	0.712	0.760	0.548	0.721	0.625	0.635	0.471	0.740	0.721	0.654	0.663	0.644	0.923	0.721	0.817	0.606	**0.686**
D-ConvNet	0.539	0.645	0.713	0.800	0.625	0.549	0.577	0.702	0.575	0.711	0.565	0.664	0.614	0.684	0.541	0.712	0.490	0.598	0.924	0.712	0.761	0.585	**0.649**
S-ConvNet	0.606	0.740	0.739	0.633	0.635	0.568	0.692	0.609	0.720	0.673	0.688	0.576	0.620	0.709	0.615	0.835	0.723	0.645	0.894	0.665	0.788	0.585	**0.680**
EEGNet-2.32	0.575	0.761	0.741	0.809	0.741	0.587	0.806	0.647	0.682	0.700	0.614	0.645	0.653	0.732	0.675	0.845	0.591	0.672	0.952	0.815	0.759	0.558	**0.707**
EEGNet-4.8	0.559	0.752	0.750	0.839	0.656	0.645	0.779	0.665	0.681	0.806	0.700	0.735	0.768	0.730	0.705	0.757	0.676	0.799	0.951	0.775	0.790	0.693	**0.737**

Table B.15: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.587	0.740	0.683	0.817	0.692	0.654	0.673	0.769	0.500	0.798	0.587	0.673	0.548	0.654	0.683	0.654	0.654	0.519	0.923	0.808	0.721	0.596	**0.679**
MDM	0.615	0.721	0.673	0.817	0.683	0.587	0.721	0.779	0.596	0.692	0.644	0.654	0.606	0.692	0.692	0.663	0.558	0.500	0.942	0.654	0.731	0.519	**0.670**
TS+LR	0.673	0.750	0.606	0.827	0.673	0.644	0.721	0.817	0.558	0.798	0.615	0.615	0.529	0.750	0.779	0.702	0.683	0.635	0.952	0.798	0.721	0.663	**0.705**
D-ConvNet	0.530	0.704	0.672	0.771	0.586	0.558	0.585	0.681	0.606	0.720	0.615	0.579	0.595	0.634	0.540	0.722	0.530	0.551	0.931	0.775	0.724	0.653	**0.648**
S-ConvNet	0.551	0.769	0.640	0.713	0.615	0.508	0.615	0.665	0.646	0.729	0.570	0.665	0.614	0.691	0.665	0.808	0.664	0.667	0.865	0.762	0.816	0.606	**0.675**
EEGNet-2.32	0.558	0.753	0.752	0.780	0.635	0.565	0.730	0.692	0.701	0.729	0.661	0.673	0.713	0.703	0.685	0.836	0.645	0.801	0.933	0.806	0.781	0.634	**0.717**
EEGNet-4.8	0.577	0.768	0.692	0.802	0.642	0.576	0.724	0.712	0.646	0.790	0.615	0.760	0.740	0.665	0.667	0.749	0.589	0.782	0.970	0.805	0.801	0.708	**0.717**

Table B.16: Classification accuracy obtained by different classifiers for 22 subjects using 13 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.683	0.712	0.663	0.808	0.740	0.635	0.663	0.760	0.471	0.750	0.644	0.644	0.567	0.663	0.731	0.663	0.596	0.558	0.933	0.788	0.692	0.548	**0.678**
MDM	0.635	0.712	0.702	0.817	0.683	0.587	0.740	0.779	0.596	0.673	0.644	0.654	0.606	0.615	0.702	0.673	0.548	0.500	0.942	0.644	0.692	0.519	**0.667**
TS+LR	0.702	0.740	0.644	0.827	0.673	0.663	0.712	0.808	0.548	0.779	0.606	0.644	0.567	0.740	0.798	0.702	0.692	0.644	0.952	0.808	0.731	0.635	**0.710**
D-ConvNet	0.636	0.751	0.625	0.742	0.521	0.518	0.624	0.777	0.706	0.716	0.567	0.561	0.596	0.645	0.550	0.770	0.485	0.732	0.922	0.737	0.713	0.596	**0.659**
S-ConvNet	0.571	0.740	0.621	0.652	0.619	0.527	0.634	0.665	0.665	0.750	0.571	0.664	0.614	0.672	0.655	0.809	0.596	0.724	0.865	0.791	0.797	0.626	**0.674**
EEGNet-2.32	0.556	0.770	0.799	0.809	0.616	0.539	0.702	0.751	0.634	0.742	0.643	0.674	0.691	0.760	0.715	0.788	0.556	0.800	0.932	0.797	0.802	0.650	**0.715**
EEGNet-4.8	0.577	0.731	0.779	0.858	0.595	0.585	0.744	0.693	0.648	0.808	0.577	0.759	0.681	0.675	0.695	0.718	0.576	0.840	0.951	0.783	0.829	0.711	**0.719**

Table B.17: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.596	0.750	0.635	0.769	0.654	0.577	0.635	0.798	0.567	0.673	0.702	0.625	0.587	0.702	0.692	0.577	0.702	0.625	0.933	0.750	0.712	0.481	**0.670**
MDM	0.606	0.731	0.625	0.769	0.635	0.567	0.712	0.740	0.596	0.663	0.635	0.635	0.625	0.596	0.692	0.663	0.683	0.500	0.952	0.615	0.721	0.413	**0.653**
TS+LR	0.635	0.798	0.673	0.721	0.663	0.587	0.692	0.788	0.587	0.702	0.683	0.644	0.558	0.654	0.683	0.692	0.654	0.644	0.952	0.702	0.731	0.490	**0.679**
D-ConvNet	0.531	0.703	0.743	0.829	0.634	0.529	0.548	0.615	0.530	0.673	0.599	0.714	0.643	0.645	0.569	0.672	0.425	0.501	0.923	0.752	0.825	0.558	**0.644**
S-ConvNet	0.575	0.771	0.779	0.672	0.662	0.547	0.696	0.634	0.635	0.624	0.700	0.722	0.682	0.664	0.645	0.856	0.608	0.639	0.855	0.702	0.787	0.627	**0.685**
EEGNet-2.32	0.481	0.790	0.809	0.741	0.661	0.741	0.732	0.685	0.667	0.780	0.557	0.695	0.714	0.709	0.675	0.818	0.567	0.655	0.962	0.825	0.855	0.674	**0.718**
EEGNet-4.8	0.484	0.780	0.847	0.772	0.712	0.682	0.733	0.684	0.643	0.712	0.650	0.763	0.751	0.787	0.656	0.805	0.615	0.732	0.951	0.786	0.855	0.670	**0.730**

Table B.18: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.644	0.769	0.625	0.769	0.673	0.615	0.615	0.798	0.587	0.663	0.683	0.625	0.567	0.702	0.692	0.577	0.673	0.692	0.933	0.740	0.702	0.548	**0.677**
MDM	0.606	0.750	0.615	0.769	0.683	0.567	0.702	0.760	0.587	0.606	0.635	0.635	0.635	0.644	0.702	0.663	0.702	0.500	0.942	0.635	0.740	0.423	**0.659**
TS+LR	0.625	0.808	0.663	0.721	0.654	0.615	0.721	0.769	0.606	0.692	0.673	0.644	0.596	0.654	0.692	0.712	0.635	0.615	0.942	0.702	0.750	0.490	**0.681**
D-ConvNet	0.540	0.710	0.787	0.793	0.654	0.593	0.589	0.633	0.561	0.682	0.597	0.656	0.623	0.663	0.590	0.683	0.517	0.608	0.924	0.731	0.847	0.540	**0.660**
S-ConvNet	0.594	0.752	0.750	0.685	0.692	0.557	0.639	0.663	0.618	0.634	0.710	0.731	0.643	0.665	0.683	0.827	0.597	0.632	0.836	0.731	0.788	0.615	**0.684**
EEGNet-2.32	0.502	0.781	0.835	0.750	0.634	0.625	0.741	0.695	0.626	0.798	0.625	0.705	0.705	0.720	0.675	0.807	0.587	0.634	0.953	0.826	0.867	0.606	**0.714**
EEGNet-4.8	0.484	0.798	0.828	0.742	0.729	0.642	0.750	0.714	0.614	0.769	0.680	0.733	0.771	0.818	0.675	0.816	0.625	0.655	0.952	0.825	0.848	0.653	**0.733**

Table B.19: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.606	0.740	0.692	0.808	0.644	0.615	0.644	0.731	0.596	0.702	0.644	0.683	0.587	0.663	0.721	0.635	0.702	0.596	0.942	0.750	0.702	0.567	**0.681**
MDM	0.519	0.760	0.663	0.769	0.587	0.587	0.692	0.702	0.625	0.606	0.663	0.625	0.625	0.702	0.663	0.683	0.683	0.510	0.962	0.635	0.740	0.558	**0.662**
TS+LR	0.635	0.788	0.683	0.769	0.625	0.644	0.702	0.808	0.654	0.760	0.635	0.663	0.596	0.712	0.692	0.712	0.606	0.635	0.971	0.731	0.798	0.558	**0.699**
D-ConvNet	0.475	0.723	0.710	0.763	0.635	0.586	0.578	0.684	0.694	0.691	0.571	0.606	0.612	0.617	0.577	0.673	0.502	0.608	0.961	0.725	0.780	0.536	**0.650**
S-ConvNet	0.552	0.781	0.758	0.655	0.652	0.625	0.657	0.644	0.663	0.617	0.626	0.712	0.703	0.636	0.695	0.827	0.566	0.643	0.864	0.733	0.787	0.644	**0.684**
EEGNet-2.32	0.551	0.799	0.830	0.753	0.660	0.615	0.739	0.723	0.619	0.664	0.635	0.732	0.702	0.780	0.666	0.798	0.519	0.617	0.953	0.786	0.857	0.636	**0.711**
EEGNet-4.8	0.522	0.835	0.838	0.772	0.615	0.644	0.733	0.704	0.643	0.711	0.578	0.771	0.692	0.768	0.694	0.825	0.605	0.655	0.942	0.796	0.875	0.621	**0.720**

Table B.20: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.615	0.740	0.721	0.808	0.606	0.577	0.635	0.740	0.635	0.721	0.663	0.702	0.558	0.673	0.712	0.635	0.673	0.567	0.962	0.779	0.702	0.529	**0.680**
MDM	0.538	0.788	0.673	0.779	0.587	0.577	0.712	0.702	0.625	0.625	0.663	0.615	0.615	0.702	0.663	0.683	0.683	0.510	0.962	0.635	0.721	0.510	**0.662**
TS+LR	0.635	0.798	0.673	0.769	0.596	0.625	0.692	0.798	0.635	0.779	0.635	0.663	0.606	0.731	0.712	0.683	0.644	0.654	0.971	0.721	0.779	0.567	**0.698**
D-ConvNet	0.570	0.693	0.751	0.772	0.557	0.539	0.558	0.654	0.646	0.634	0.570	0.600	0.654	0.622	0.546	0.748	0.505	0.549	0.923	0.748	0.782	0.671	**0.650**
S-ConvNet	0.544	0.732	0.795	0.723	0.661	0.595	0.617	0.615	0.673	0.646	0.617	0.703	0.713	0.740	0.704	0.845	0.635	0.685	0.883	0.769	0.777	0.684	**0.698**
EEGNet-2.32	0.513	0.798	0.827	0.762	0.661	0.604	0.769	0.714	0.666	0.703	0.588	0.723	0.712	0.721	0.658	0.779	0.538	0.664	0.972	0.796	0.849	0.635	**0.711**
EEGNet-4.8	0.561	0.788	0.857	0.772	0.672	0.661	0.724	0.655	0.624	0.672	0.588	0.761	0.741	0.779	0.685	0.805	0.564	0.810	0.942	0.834	0.865	0.654	**0.728**

Table B.21: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.635	0.721	0.644	0.760	0.615	0.683	0.692	0.808	0.577	0.731	0.615	0.606	0.548	0.635	0.673	0.625	0.577	0.663	0.913	0.779	0.683	0.510	**0.668**
MDM	0.663	0.750	0.625	0.769	0.635	0.625	0.721	0.798	0.596	0.721	0.635	0.596	0.587	0.635	0.702	0.635	0.490	0.500	0.923	0.625	0.702	0.442	**0.653**
TS+LR	0.750	0.760	0.673	0.779	0.644	0.673	0.712	0.798	0.577	0.817	0.567	0.558	0.577	0.644	0.731	0.692	0.673	0.663	0.952	0.750	0.750	0.548	**0.695**
D-ConvNet	0.628	0.654	0.672	0.763	0.681	0.578	0.527	0.634	0.568	0.711	0.625	0.508	0.655	0.635	0.567	0.655	0.501	0.587	0.912	0.730	0.752	0.575	**0.642**
S-ConvNet	0.567	0.701	0.682	0.701	0.585	0.522	0.674	0.656	0.635	0.650	0.658	0.546	0.598	0.664	0.635	0.846	0.685	0.700	0.922	0.694	0.827	0.604	**0.670**
EEGNet-2.32	0.527	0.725	0.768	0.770	0.634	0.615	0.761	0.734	0.645	0.769	0.625	0.540	0.662	0.641	0.656	0.740	0.582	0.622	0.961	0.817	0.801	0.596	**0.690**
EEGNet-4.8	0.588	0.750	0.779	0.810	0.625	0.598	0.768	0.704	0.633	0.684	0.634	0.626	0.750	0.760	0.684	0.747	0.591	0.656	0.952	0.824	0.831	0.643	**0.711**

Table B.22: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.615	0.750	0.625	0.769	0.654	0.625	0.663	0.788	0.567	0.683	0.635	0.635	0.538	0.635	0.673	0.644	0.567	0.606	0.913	0.760	0.721	0.519	**0.663**
MDM	0.654	0.750	0.596	0.769	0.663	0.596	0.702	0.808	0.587	0.721	0.635	0.596	0.558	0.596	0.673	0.654	0.500	0.500	0.942	0.625	0.712	0.442	**0.649**
TS+LR	0.750	0.769	0.654	0.788	0.654	0.683	0.712	0.798	0.558	0.808	0.577	0.587	0.577	0.625	0.712	0.683	0.644	0.663	0.952	0.750	0.760	0.529	**0.692**
D-ConvNet	0.551	0.693	0.675	0.745	0.679	0.546	0.577	0.653	0.609	0.673	0.564	0.624	0.664	0.648	0.530	0.733	0.549	0.604	0.895	0.795	0.783	0.593	**0.654**
S-ConvNet	0.557	0.723	0.693	0.721	0.619	0.550	0.665	0.714	0.673	0.633	0.679	0.555	0.616	0.683	0.625	0.828	0.705	0.673	0.913	0.720	0.826	0.604	**0.681**
EEGNet-2.32	0.518	0.770	0.749	0.762	0.693	0.615	0.703	0.733	0.645	0.777	0.637	0.559	0.704	0.662	0.647	0.758	0.630	0.685	0.951	0.815	0.828	0.616	**0.703**
EEGNet-4.8	0.623	0.729	0.750	0.839	0.696	0.617	0.769	0.675	0.575	0.711	0.654	0.675	0.749	0.730	0.665	0.805	0.677	0.692	0.943	0.824	0.821	0.612	**0.720**

Table B.23: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.606	0.731	0.625	0.798	0.692	0.625	0.673	0.817	0.606	0.788	0.712	0.683	0.606	0.712	0.663	0.673	0.567	0.577	0.962	0.788	0.721	0.519	**0.688**
MDM	0.519	0.769	0.654	0.779	0.625	0.587	0.712	0.788	0.596	0.654	0.644	0.663	0.587	0.663	0.683	0.654	0.490	0.500	0.962	0.654	0.721	0.519	**0.656**
TS+LR	0.625	0.779	0.673	0.788	0.673	0.606	0.644	0.846	0.548	0.779	0.683	0.577	0.587	0.702	0.750	0.692	0.615	0.615	0.971	0.750	0.721	0.587	**0.691**
D-ConvNet	0.648	0.691	0.743	0.764	0.613	0.595	0.568	0.684	0.655	0.779	0.585	0.529	0.625	0.635	0.550	0.789	0.525	0.647	0.933	0.776	0.779	0.625	**0.670**
S-ConvNet	0.598	0.741	0.588	0.749	0.590	0.583	0.647	0.664	0.714	0.694	0.625	0.666	0.615	0.712	0.635	0.753	0.666	0.645	0.924	0.771	0.828	0.635	**0.684**
EEGNet-2.32	0.511	0.799	0.807	0.754	0.664	0.530	0.702	0.703	0.675	0.739	0.548	0.692	0.722	0.683	0.695	0.807	0.607	0.695	0.961	0.805	0.821	0.643	**0.707**
EEGNet-4.8	0.567	0.761	0.724	0.849	0.672	0.550	0.732	0.724	0.598	0.703	0.528	0.752	0.749	0.711	0.675	0.740	0.705	0.675	0.952	0.795	0.793	0.668	**0.710**

Table B.24: Classification accuracy obtained by different classifiers for 22 subjects using 9 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.663	0.731	0.673	0.817	0.673	0.577	0.702	0.788	0.577	0.798	0.692	0.673	0.558	0.712	0.692	0.673	0.538	0.538	0.962	0.760	0.692	0.548	**0.684**
MDM	0.538	0.769	0.673	0.779	0.606	0.567	0.721	0.769	0.606	0.663	0.654	0.663	0.596	0.635	0.683	0.654	0.481	0.500	0.952	0.654	0.702	0.538	**0.655**
TS+LR	0.673	0.750	0.692	0.798	0.644	0.596	0.673	0.837	0.558	0.779	0.673	0.615	0.615	0.702	0.779	0.692	0.644	0.635	0.962	0.750	0.683	0.606	**0.698**
D-ConvNet	0.645	0.761	0.682	0.743	0.605	0.606	0.586	0.624	0.605	0.731	0.586	0.600	0.635	0.605	0.570	0.759	0.558	0.713	0.931	0.738	0.778	0.653	**0.669**
S-ConvNet	0.580	0.761	0.635	0.742	0.503	0.584	0.645	0.673	0.703	0.742	0.617	0.655	0.605	0.751	0.684	0.781	0.705	0.723	0.923	0.753	0.779	0.675	**0.692**
EEGNet-2.32	0.501	0.743	0.799	0.743	0.599	0.578	0.740	0.722	0.626	0.759	0.558	0.647	0.663	0.654	0.675	0.737	0.646	0.665	0.942	0.846	0.800	0.631	**0.694**
EEGNet-4.8	0.565	0.768	0.740	0.801	0.704	0.606	0.760	0.685	0.599	0.711	0.548	0.733	0.710	0.748	0.665	0.739	0.647	0.733	0.971	0.814	0.783	0.658	**0.713**

Table B.25: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.587	0.750	0.721	0.769	0.673	0.577	0.683	0.760	0.558	0.635	0.644	0.606	0.577	0.615	0.731	0.721	0.625	0.673	0.913	0.702	0.731	0.538	**0.672**
MDM	0.538	0.769	0.654	0.760	0.606	0.615	0.702	0.740	0.596	0.596	0.635	0.587	0.558	0.587	0.683	0.663	0.663	0.500	0.933	0.625	0.692	0.394	**0.641**
TS+LR	0.567	0.779	0.663	0.760	0.625	0.606	0.740	0.817	0.606	0.663	0.567	0.635	0.558	0.644	0.712	0.731	0.644	0.625	0.952	0.673	0.683	0.442	**0.668**
D-ConvNet	0.576	0.654	0.722	0.782	0.625	0.587	0.617	0.654	0.586	0.595	0.599	0.586	0.615	0.712	0.511	0.683	0.454	0.537	0.914	0.728	0.838	0.512	**0.640**
S-ConvNet	0.566	0.751	0.795	0.673	0.664	0.515	0.665	0.614	0.653	0.633	0.671	0.635	0.646	0.624	0.705	0.798	0.501	0.606	0.874	0.671	0.771	0.540	**0.662**
EEGNet-2.32	0.564	0.817	0.771	0.725	0.730	0.615	0.730	0.735	0.633	0.691	0.692	0.607	0.634	0.721	0.618	0.778	0.489	0.694	0.923	0.788	0.837	0.542	**0.697**
EEGNet-4.8	0.559	0.787	0.723	0.781	0.722	0.650	0.741	0.694	0.655	0.684	0.634	0.683	0.673	0.712	0.666	0.749	0.552	0.735	0.962	0.729	0.770	0.542	**0.700**

Table B.26: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.615	0.750	0.702	0.769	0.663	0.596	0.692	0.769	0.567	0.644	0.625	0.596	0.587	0.587	0.721	0.702	0.635	0.663	0.913	0.692	0.721	0.490	**0.668**
MDM	0.529	0.769	0.635	0.769	0.615	0.615	0.692	0.760	0.596	0.625	0.635	0.587	0.548	0.615	0.683	0.673	0.712	0.500	0.933	0.625	0.692	0.413	**0.646**
TS+LR	0.538	0.788	0.683	0.750	0.644	0.596	0.750	0.817	0.615	0.673	0.577	0.635	0.548	0.625	0.731	0.731	0.635	0.625	0.952	0.692	0.702	0.471	**0.672**
D-ConvNet	0.558	0.704	0.719	0.803	0.605	0.604	0.606	0.635	0.520	0.643	0.665	0.674	0.645	0.703	0.557	0.663	0.495	0.539	0.913	0.672	0.904	0.518	**0.652**
S-ConvNet	0.556	0.742	0.739	0.702	0.636	0.505	0.667	0.605	0.615	0.681	0.679	0.643	0.635	0.566	0.685	0.779	0.528	0.579	0.863	0.699	0.790	0.500	**0.654**
EEGNet-2.32	0.576	0.779	0.740	0.744	0.691	0.625	0.721	0.755	0.644	0.728	0.694	0.598	0.634	0.711	0.608	0.796	0.463	0.637	0.923	0.788	0.847	0.560	**0.694**
EEGNet-4.8	0.522	0.798	0.780	0.771	0.645	0.619	0.751	0.703	0.609	0.730	0.699	0.714	0.703	0.693	0.666	0.777	0.514	0.665	0.952	0.689	0.798	0.538	**0.697**

Table B.27: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.481	0.788	0.712	0.779	0.654	0.558	0.683	0.769	0.567	0.587	0.625	0.663	0.615	0.644	0.721	0.692	0.635	0.644	0.933	0.740	0.740	0.519	**0.670**
MDM	0.481	0.788	0.635	0.760	0.615	0.577	0.731	0.712	0.615	0.606	0.644	0.635	0.596	0.606	0.683	0.673	0.692	0.500	0.962	0.596	0.712	0.481	**0.650**
TS+LR	0.558	0.798	0.683	0.760	0.587	0.606	0.712	0.788	0.596	0.596	0.635	0.615	0.644	0.702	0.673	0.721	0.635	0.663	0.971	0.721	0.740	0.500	**0.677**
D-ConvNet	0.466	0.730	0.722	0.812	0.645	0.595	0.607	0.715	0.577	0.637	0.616	0.547	0.595	0.655	0.588	0.731	0.498	0.559	0.903	0.653	0.856	0.595	**0.650**
S-ConvNet	0.620	0.759	0.625	0.732	0.635	0.585	0.664	0.693	0.645	0.614	0.550	0.626	0.615	0.671	0.740	0.810	0.546	0.562	0.931	0.741	0.744	0.548	**0.666**
EEGNet-2.32	0.530	0.816	0.763	0.734	0.682	0.605	0.731	0.733	0.685	0.673	0.635	0.664	0.633	0.673	0.725	0.826	0.477	0.657	0.933	0.769	0.819	0.605	**0.698**
EEGNet-4.8	0.491	0.780	0.734	0.781	0.651	0.604	0.722	0.752	0.638	0.693	0.550	0.741	0.654	0.710	0.657	0.768	0.510	0.659	0.942	0.739	0.877	0.527	**0.690**

Table B.28: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.558	0.750	0.712	0.740	0.673	0.558	0.692	0.760	0.587	0.548	0.625	0.654	0.567	0.625	0.740	0.712	0.587	0.673	0.942	0.740	0.740	0.500	**0.667**
MDM	0.490	0.798	0.673	0.769	0.606	0.577	0.721	0.712	0.606	0.606	0.644	0.644	0.644	0.577	0.673	0.721	0.692	0.500	0.962	0.587	0.712	0.490	**0.655**
TS+LR	0.596	0.798	0.692	0.760	0.635	0.606	0.712	0.779	0.587	0.606	0.635	0.615	0.635	0.673	0.702	0.731	0.644	0.644	0.971	0.721	0.702	0.510	**0.680**
D-ConvNet	0.525	0.674	0.712	0.801	0.633	0.606	0.577	0.751	0.560	0.645	0.608	0.597	0.665	0.615	0.548	0.670	0.512	0.578	0.884	0.665	0.787	0.578	**0.645**
S-ConvNet	0.571	0.742	0.682	0.713	0.653	0.604	0.655	0.675	0.656	0.585	0.542	0.636	0.614	0.720	0.730	0.797	0.501	0.607	0.931	0.722	0.783	0.538	**0.666**
EEGNet-2.32	0.557	0.797	0.734	0.733	0.722	0.605	0.740	0.752	0.693	0.615	0.614	0.665	0.614	0.663	0.715	0.797	0.491	0.685	0.942	0.788	0.788	0.540	**0.693**
EEGNet-4.8	0.584	0.800	0.705	0.789	0.718	0.615	0.742	0.761	0.639	0.645	0.560	0.741	0.642	0.731	0.723	0.749	0.529	0.735	0.952	0.776	0.867	0.608	**0.710**

Table B.29: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.615	0.721	0.615	0.683	0.663	0.625	0.625	0.731	0.519	0.750	0.577	0.654	0.548	0.721	0.750	0.606	0.635	0.529	0.865	0.654	0.760	0.587	**0.656**
MDM	0.587	0.644	0.596	0.731	0.577	0.625	0.663	0.788	0.596	0.663	0.596	0.567	0.615	0.731	0.740	0.587	0.558	0.500	0.904	0.596	0.692	0.404	**0.635**
TS+LR	0.606	0.721	0.615	0.702	0.683	0.606	0.615	0.769	0.567	0.721	0.538	0.596	0.577	0.769	0.760	0.596	0.644	0.558	0.923	0.615	0.779	0.567	**0.660**
D-ConvNet	0.509	0.584	0.635	0.743	0.554	0.528	0.547	0.626	0.586	0.711	0.568	0.674	0.594	0.588	0.539	0.643	0.549	0.623	0.761	0.628	0.577	0.549	**0.605**
S-ConvNet	0.635	0.679	0.741	0.655	0.607	0.557	0.665	0.596	0.641	0.709	0.631	0.570	0.672	0.626	0.607	0.827	0.617	0.625	0.837	0.607	0.757	0.661	**0.665**
EEGNet-2.32	0.557	0.685	0.720	0.762	0.617	0.575	0.661	0.674	0.545	0.807	0.633	0.691	0.672	0.701	0.703	0.767	0.723	0.712	0.905	0.702	0.778	0.730	**0.696**
EEGNet-4.8	0.550	0.655	0.751	0.772	0.606	0.616	0.633	0.547	0.568	0.820	0.635	0.751	0.690	0.701	0.705	0.795	0.627	0.674	0.924	0.719	0.749	0.674	**0.689**

Table B.30: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.615	0.750	0.587	0.692	0.644	0.587	0.615	0.731	0.510	0.692	0.577	0.625	0.558	0.721	0.760	0.615	0.635	0.606	0.894	0.663	0.760	0.558	**0.654**
MDM	0.625	0.635	0.596	0.731	0.596	0.596	0.654	0.779	0.596	0.654	0.596	0.558	0.606	0.740	0.740	0.596	0.577	0.500	0.904	0.577	0.721	0.471	**0.639**
TS+LR	0.644	0.769	0.606	0.702	0.692	0.606	0.587	0.769	0.577	0.702	0.538	0.606	0.558	0.788	0.721	0.596	0.625	0.596	0.923	0.625	0.750	0.587	**0.662**
D-ConvNet	0.499	0.653	0.703	0.714	0.527	0.557	0.569	0.594	0.568	0.713	0.530	0.683	0.584	0.625	0.550	0.653	0.529	0.559	0.800	0.608	0.683	0.664	**0.616**
S-ConvNet	0.645	0.689	0.742	0.625	0.598	0.565	0.684	0.605	0.584	0.682	0.640	0.560	0.671	0.605	0.617	0.826	0.588	0.664	0.836	0.617	0.795	0.662	**0.659**
EEGNet-2.32	0.507	0.774	0.731	0.743	0.606	0.556	0.635	0.692	0.595	0.729	0.625	0.663	0.691	0.653	0.703	0.768	0.656	0.706	0.924	0.711	0.790	0.654	**0.687**
EEGNet-4.8	0.509	0.724	0.769	0.752	0.635	0.605	0.624	0.576	0.643	0.789	0.635	0.789	0.738	0.701	0.705	0.807	0.655	0.732	0.904	0.756	0.827	0.701	**0.708**

Table B.31: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.654	0.721	0.606	0.712	0.596	0.615	0.654	0.750	0.587	0.808	0.538	0.606	0.577	0.731	0.740	0.625	0.615	0.538	0.913	0.635	0.721	0.596	**0.661**
MDM	0.587	0.654	0.625	0.721	0.587	0.625	0.683	0.750	0.615	0.615	0.635	0.587	0.587	0.760	0.731	0.596	0.519	0.500	0.923	0.625	0.721	0.567	**0.646**
TS+LR	0.654	0.760	0.663	0.712	0.625	0.577	0.625	0.779	0.587	0.798	0.625	0.635	0.558	0.769	0.779	0.635	0.692	0.596	0.933	0.673	0.740	0.683	**0.686**
D-ConvNet	0.529	0.674	0.722	0.722	0.616	0.566	0.567	0.685	0.558	0.695	0.566	0.645	0.681	0.615	0.530	0.672	0.558	0.586	0.845	0.558	0.647	0.674	**0.632**
S-ConvNet	0.531	0.760	0.644	0.670	0.648	0.583	0.635	0.566	0.606	0.703	0.588	0.684	0.701	0.665	0.625	0.798	0.580	0.646	0.874	0.615	0.739	0.713	**0.662**
EEGNet-2.32	0.531	0.753	0.759	0.752	0.614	0.555	0.615	0.664	0.635	0.691	0.526	0.673	0.673	0.701	0.704	0.739	0.714	0.668	0.952	0.708	0.780	0.672	**0.685**
EEGNet-4.8	0.575	0.701	0.749	0.762	0.674	0.626	0.674	0.711	0.626	0.730	0.685	0.778	0.671	0.711	0.685	0.837	0.652	0.597	0.922	0.708	0.816	0.712	**0.709**

Table B.32: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.712	0.702	0.606	0.683	0.577	0.635	0.635	0.731	0.587	0.798	0.558	0.654	0.567	0.750	0.750	0.615	0.663	0.606	0.904	0.635	0.702	0.587	**0.666**
MDM	0.567	0.635	0.625	0.731	0.587	0.615	0.692	0.740	0.625	0.654	0.635	0.596	0.587	0.721	0.750	0.625	0.519	0.500	0.933	0.625	0.721	0.548	**0.647**
TS+LR	0.673	0.740	0.683	0.731	0.587	0.577	0.625	0.769	0.577	0.798	0.615	0.654	0.577	0.750	0.788	0.615	0.654	0.548	0.942	0.673	0.721	0.635	**0.679**
D-ConvNet	0.581	0.675	0.721	0.741	0.595	0.614	0.567	0.673	0.539	0.692	0.555	0.596	0.654	0.588	0.558	0.690	0.501	0.549	0.835	0.576	0.587	0.653	**0.625**
S-ConvNet	0.605	0.681	0.670	0.708	0.592	0.545	0.633	0.604	0.606	0.759	0.608	0.656	0.681	0.674	0.588	0.798	0.616	0.598	0.854	0.558	0.758	0.760	**0.662**
EEGNet-2.32	0.549	0.723	0.767	0.742	0.559	0.537	0.645	0.644	0.538	0.731	0.565	0.643	0.664	0.741	0.724	0.710	0.646	0.655	0.932	0.708	0.781	0.730	**0.679**
EEGNet-4.8	0.607	0.661	0.711	0.772	0.643	0.664	0.674	0.614	0.627	0.761	0.655	0.779	0.692	0.692	0.725	0.786	0.625	0.624	0.923	0.690	0.780	0.712	**0.701**

Table B.33: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe from left side of the brain for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.635	0.760	0.625	0.721	0.587	0.558	0.673	0.712	0.548	0.673	0.644	0.663	0.558	0.615	0.731	0.702	0.577	0.577	0.933	0.760	0.644	0.471	**0.653**
MDM	0.596	0.740	0.577	0.760	0.625	0.548	0.683	0.625	0.596	0.702	0.663	0.587	0.510	0.510	0.731	0.663	0.519	0.500	0.942	0.673	0.644	0.423	**0.628**
TS+LR	0.702	0.721	0.663	0.750	0.635	0.615	0.702	0.779	0.519	0.731	0.635	0.683	0.567	0.635	0.712	0.663	0.654	0.519	0.962	0.721	0.654	0.423	**0.666**
D-ConvNet	0.521	0.545	0.565	0.745	0.536	0.527	0.695	0.615	0.635	0.674	0.589	0.615	0.615	0.657	0.567	0.575	0.520	0.595	0.895	0.699	0.645	0.575	**0.619**
S-ConvNet	0.575	0.732	0.690	0.770	0.590	0.488	0.661	0.605	0.611	0.720	0.585	0.575	0.615	0.777	0.771	0.836	0.655	0.703	0.855	0.702	0.749	0.625	**0.677**
EEGNet-2.32	0.463	0.700	0.703	0.821	0.587	0.531	0.722	0.577	0.576	0.740	0.645	0.606	0.598	0.694	0.674	0.798	0.585	0.771	0.962	0.817	0.741	0.571	**0.676**
EEGNet-4.8	0.535	0.681	0.671	0.811	0.548	0.510	0.788	0.527	0.559	0.682	0.684	0.600	0.637	0.781	0.759	0.815	0.635	0.753	0.934	0.756	0.811	0.615	**0.686**

Table B.34: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe from left side of the brain for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.654	0.750	0.654	0.683	0.615	0.587	0.635	0.731	0.558	0.683	0.654	0.635	0.587	0.635	0.702	0.692	0.596	0.606	0.942	0.760	0.654	0.413	**0.656**
MDM	0.606	0.740	0.587	0.769	0.663	0.538	0.683	0.625	0.587	0.721	0.663	0.615	0.500	0.558	0.731	0.654	0.500	0.510	0.942	0.673	0.663	0.404	**0.633**
TS+LR	0.692	0.750	0.663	0.750	0.644	0.596	0.692	0.769	0.548	0.760	0.635	0.663	0.567	0.615	0.712	0.654	0.635	0.615	0.962	0.702	0.644	0.413	**0.667**
D-ConvNet	0.490	0.594	0.624	0.753	0.558	0.517	0.664	0.569	0.595	0.681	0.604	0.585	0.648	0.627	0.557	0.729	0.579	0.672	0.885	0.726	0.697	0.500	**0.630**
S-ConvNet	0.473	0.721	0.701	0.789	0.551	0.507	0.682	0.586	0.574	0.623	0.595	0.546	0.626	0.760	0.730	0.791	0.666	0.692	0.846	0.682	0.749	0.615	**0.659**
EEGNet-2.32	0.463	0.694	0.713	0.830	0.693	0.495	0.732	0.595	0.650	0.683	0.685	0.540	0.569	0.683	0.683	0.845	0.634	0.770	0.972	0.798	0.741	0.618	**0.686**
EEGNet-4.8	0.432	0.702	0.709	0.866	0.599	0.537	0.807	0.557	0.615	0.693	0.673	0.553	0.655	0.742	0.719	0.805	0.637	0.771	0.943	0.746	0.812	0.635	**0.691**

Table B.35: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe from left side of the brain for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.625	0.779	0.683	0.788	0.615	0.625	0.644	0.740	0.567	0.712	0.644	0.615	0.510	0.673	0.731	0.692	0.596	0.606	0.942	0.760	0.683	0.519	**0.670**
MDM	0.615	0.750	0.635	0.769	0.615	0.587	0.702	0.635	0.606	0.702	0.644	0.615	0.596	0.548	0.712	0.635	0.500	0.500	0.942	0.692	0.692	0.538	**0.647**
TS+LR	0.663	0.769	0.712	0.779	0.635	0.587	0.683	0.808	0.538	0.760	0.663	0.644	0.558	0.702	0.692	0.635	0.625	0.625	0.933	0.760	0.712	0.481	**0.680**
D-ConvNet	0.506	0.681	0.615	0.791	0.559	0.568	0.694	0.625	0.566	0.654	0.596	0.538	0.586	0.606	0.596	0.674	0.566	0.684	0.905	0.679	0.666	0.508	**0.630**
S-ConvNet	0.530	0.758	0.652	0.733	0.625	0.585	0.645	0.597	0.645	0.585	0.586	0.615	0.656	0.630	0.729	0.791	0.703	0.676	0.865	0.731	0.735	0.657	**0.670**
EEGNet-2.32	0.518	0.685	0.685	0.857	0.701	0.537	0.750	0.605	0.632	0.672	0.597	0.587	0.550	0.694	0.700	0.827	0.665	0.763	0.962	0.778	0.752	0.519	**0.683**
EEGNet-4.8	0.557	0.749	0.692	0.810	0.623	0.605	0.722	0.684	0.576	0.683	0.568	0.658	0.636	0.695	0.739	0.797	0.645	0.771	0.933	0.756	0.752	0.566	**0.692**

Table B.36: Classification accuracy obtained by different classifiers for 22 subjects using 6 channel layout over motor cortex and frontal lobe from left side of the brain for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.635	0.788	0.702	0.808	0.615	0.587	0.663	0.750	0.538	0.731	0.683	0.625	0.577	0.731	0.740	0.692	0.615	0.567	0.942	0.750	0.635	0.529	**0.677**
MDM	0.673	0.750	0.663	0.779	0.587	0.538	0.702	0.635	0.625	0.712	0.644	0.625	0.577	0.529	0.740	0.644	0.500	0.500	0.942	0.683	0.673	0.519	**0.647**
TS+LR	0.663	0.760	0.731	0.808	0.625	0.577	0.683	0.788	0.548	0.769	0.654	0.644	0.596	0.683	0.721	0.654	0.673	0.596	0.952	0.750	0.692	0.500	**0.685**
D-ConvNet	0.472	0.585	0.597	0.744	0.490	0.509	0.684	0.644	0.616	0.674	0.596	0.547	0.577	0.598	0.568	0.701	0.559	0.567	0.894	0.708	0.615	0.577	**0.615**
S-ConvNet	0.543	0.714	0.615	0.685	0.617	0.538	0.655	0.605	0.635	0.613	0.633	0.627	0.665	0.687	0.738	0.806	0.673	0.627	0.865	0.712	0.759	0.695	**0.669**
EEGNet-2.32	0.517	0.710	0.675	0.847	0.542	0.596	0.760	0.606	0.635	0.701	0.567	0.568	0.578	0.712	0.711	0.728	0.557	0.781	0.953	0.798	0.764	0.502	**0.673**
EEGNet-4.8	0.597	0.711	0.665	0.830	0.588	0.565	0.712	0.664	0.560	0.730	0.626	0.656	0.617	0.732	0.738	0.806	0.648	0.769	0.924	0.756	0.839	0.615	**0.698**

Table B.37: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.567	0.769	0.635	0.712	0.615	0.596	0.596	0.750	0.567	0.558	0.606	0.471	0.548	0.740	0.731	0.635	0.644	0.490	0.885	0.471	0.769	0.481	**0.629**
MDM	0.548	0.615	0.606	0.692	0.625	0.596	0.663	0.683	0.606	0.596	0.606	0.500	0.615	0.712	0.692	0.606	0.654	0.510	0.846	0.423	0.673	0.404	**0.612**
TS+LR	0.529	0.750	0.625	0.692	0.615	0.635	0.625	0.740	0.548	0.548	0.587	0.500	0.615	0.731	0.731	0.644	0.683	0.529	0.865	0.490	0.740	0.423	**0.629**
D-ConvNet	0.519	0.654	0.711	0.754	0.576	0.623	0.558	0.570	0.568	0.634	0.549	0.715	0.623	0.599	0.521	0.539	0.501	0.492	0.751	0.516	0.748	0.666	**0.608**
S-ConvNet	0.577	0.731	0.641	0.702	0.558	0.568	0.605	0.570	0.551	0.596	0.641	0.749	0.583	0.615	0.635	0.663	0.510	0.569	0.895	0.546	0.727	0.645	**0.631**
EEGNet-2.32	0.499	0.742	0.778	0.773	0.516	0.566	0.597	0.625	0.579	0.721	0.625	0.749	0.711	0.654	0.703	0.720	0.585	0.702	0.865	0.555	0.788	0.665	**0.669**
EEGNet-4.8	0.490	0.665	0.767	0.741	0.586	0.681	0.695	0.590	0.567	0.629	0.557	0.741	0.652	0.679	0.674	0.690	0.526	0.707	0.836	0.470	0.797	0.724	**0.657**

Table B.38: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.558	0.750	0.615	0.712	0.587	0.606	0.606	0.740	0.567	0.577	0.615	0.471	0.538	0.740	0.731	0.644	0.654	0.519	0.885	0.471	0.779	0.529	**0.632**
MDM	0.548	0.644	0.587	0.692	0.625	0.596	0.663	0.683	0.606	0.596	0.606	0.500	0.606	0.712	0.692	0.606	0.654	0.519	0.846	0.423	0.692	0.404	**0.614**
TS+LR	0.529	0.740	0.625	0.702	0.596	0.635	0.625	0.721	0.548	0.587	0.587	0.500	0.606	0.721	0.721	0.654	0.663	0.538	0.856	0.481	0.731	0.462	**0.628**
D-ConvNet	0.536	0.644	0.729	0.800	0.656	0.616	0.569	0.560	0.532	0.635	0.567	0.704	0.623	0.559	0.549	0.559	0.445	0.530	0.799	0.547	0.686	0.682	**0.615**
S-ConvNet	0.588	0.730	0.629	0.655	0.513	0.559	0.636	0.522	0.560	0.665	0.670	0.731	0.632	0.539	0.636	0.672	0.498	0.530	0.865	0.525	0.719	0.645	**0.624**
EEGNet-2.32	0.549	0.731	0.777	0.764	0.555	0.566	0.617	0.615	0.559	0.701	0.635	0.730	0.681	0.692	0.713	0.701	0.528	0.644	0.836	0.575	0.836	0.645	**0.666**
EEGNet-4.8	0.550	0.683	0.777	0.743	0.567	0.615	0.685	0.617	0.559	0.609	0.635	0.732	0.663	0.671	0.674	0.721	0.531	0.654	0.855	0.510	0.759	0.714	**0.660**

Table B.39: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.538	0.750	0.673	0.712	0.548	0.587	0.663	0.683	0.596	0.712	0.558	0.577	0.625	0.712	0.712	0.577	0.663	0.510	0.894	0.490	0.740	0.558	**0.640**
MDM	0.481	0.673	0.606	0.712	0.615	0.558	0.635	0.663	0.596	0.615	0.635	0.587	0.596	0.702	0.712	0.558	0.654	0.510	0.846	0.423	0.740	0.548	**0.621**
TS+LR	0.558	0.731	0.712	0.721	0.577	0.587	0.625	0.712	0.587	0.654	0.596	0.596	0.567	0.692	0.731	0.615	0.673	0.538	0.885	0.510	0.760	0.577	**0.646**
D-ConvNet	0.549	0.644	0.768	0.734	0.616	0.614	0.530	0.488	0.618	0.644	0.597	0.712	0.663	0.596	0.558	0.604	0.512	0.510	0.856	0.567	0.753	0.673	**0.628**
S-ConvNet	0.569	0.712	0.669	0.693	0.674	0.576	0.617	0.557	0.477	0.574	0.548	0.655	0.625	0.625	0.684	0.665	0.547	0.528	0.855	0.545	0.750	0.701	**0.629**
EEGNet-2.32	0.510	0.741	0.776	0.723	0.607	0.617	0.616	0.503	0.559	0.635	0.519	0.692	0.712	0.702	0.732	0.672	0.503	0.681	0.863	0.480	0.788	0.624	**0.648**
EEGNet-4.8	0.510	0.663	0.758	0.743	0.548	0.607	0.645	0.644	0.615	0.568	0.567	0.701	0.683	0.673	0.675	0.740	0.491	0.701	0.875	0.479	0.808	0.720	**0.655**

Table B.40: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.538	0.760	0.692	0.712	0.558	0.635	0.663	0.683	0.625	0.692	0.577	0.577	0.587	0.731	0.750	0.567	0.644	0.519	0.913	0.500	0.740	0.548	**0.646**
MDM	0.462	0.663	0.625	0.721	0.587	0.587	0.644	0.673	0.596	0.625	0.625	0.577	0.606	0.740	0.712	0.567	0.644	0.510	0.846	0.413	0.702	0.529	**0.621**
TS+LR	0.558	0.702	0.731	0.721	0.596	0.596	0.625	0.731	0.587	0.683	0.606	0.587	0.587	0.673	0.740	0.635	0.663	0.510	0.885	0.510	0.731	0.567	**0.646**
D-ConvNet	0.540	0.655	0.702	0.800	0.518	0.643	0.566	0.527	0.570	0.625	0.557	0.645	0.653	0.568	0.529	0.605	0.518	0.530	0.828	0.555	0.734	0.693	**0.616**
S-ConvNet	0.587	0.691	0.699	0.712	0.608	0.519	0.608	0.499	0.486	0.545	0.565	0.655	0.641	0.634	0.683	0.684	0.548	0.596	0.855	0.553	0.731	0.712	**0.628**
EEGNet-2.32	0.529	0.731	0.788	0.733	0.627	0.626	0.606	0.522	0.579	0.635	0.549	0.711	0.663	0.665	0.694	0.699	0.558	0.735	0.865	0.508	0.788	0.691	**0.659**
EEGNet-4.8	0.443	0.692	0.798	0.743	0.585	0.595	0.617	0.683	0.635	0.617	0.550	0.701	0.672	0.702	0.655	0.655	0.594	0.714	0.875	0.478	0.778	0.666	**0.657**

Table B.41: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.625	0.462	0.490	0.519	0.481	0.577	0.538	0.529	0.558	0.712	0.519	0.510	0.577	0.577	0.490	0.538	0.596	0.558	0.558	0.462	0.529	0.567	**0.544**
MDM	0.558	0.452	0.462	0.519	0.490	0.577	0.606	0.577	0.625	0.692	0.615	0.548	0.519	0.529	0.548	0.558	0.538	0.548	0.567	0.481	0.452	0.462	**0.542**
TS+LR	0.625	0.471	0.452	0.490	0.481	0.596	0.567	0.587	0.577	0.702	0.606	0.548	0.548	0.596	0.500	0.548	0.673	0.577	0.567	0.510	0.519	0.519	**0.557**
D-ConvNet	0.422	0.422	0.396	0.520	0.508	0.521	0.465	0.546	0.483	0.616	0.586	0.570	0.547	0.562	0.528	0.623	0.460	0.521	0.545	0.558	0.478	0.453	**0.515**
S-ConvNet	0.527	0.483	0.663	0.531	0.598	0.459	0.516	0.479	0.576	0.694	0.442	0.600	0.525	0.590	0.727	0.643	0.499	0.589	0.528	0.517	0.588	0.598	**0.562**
EEGNet-2.32	0.567	0.442	0.553	0.519	0.521	0.452	0.511	0.496	0.556	0.566	0.404	0.560	0.527	0.579	0.595	0.585	0.607	0.675	0.493	0.499	0.540	0.561	**0.537**
EEGNet-4.8	0.577	0.464	0.536	0.509	0.531	0.480	0.519	0.556	0.541	0.579	0.525	0.581	0.538	0.589	0.603	0.718	0.514	0.517	0.502	0.509	0.398	0.606	**0.541**

Table B.42: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.606	0.471	0.471	0.510	0.500	0.596	0.558	0.538	0.558	0.683	0.529	0.519	0.606	0.510	0.490	0.567	0.635	0.558	0.538	0.442	0.490	0.548	**0.542**
MDM	0.567	0.433	0.481	0.500	0.510	0.548	0.587	0.577	0.606	0.644	0.615	0.548	0.538	0.510	0.500	0.548	0.558	0.558	0.567	0.490	0.423	0.462	**0.535**
TS+LR	0.606	0.462	0.471	0.519	0.529	0.587	0.567	0.587	0.577	0.683	0.606	0.558	0.567	0.538	0.490	0.558	0.635	0.548	0.567	0.481	0.471	0.490	**0.550**
D-ConvNet	0.482	0.488	0.471	0.488	0.462	0.518	0.567	0.518	0.566	0.624	0.530	0.579	0.575	0.490	0.455	0.575	0.501	0.570	0.471	0.445	0.499	0.569	**0.520**
S-ConvNet	0.545	0.518	0.635	0.531	0.558	0.440	0.506	0.510	0.577	0.694	0.469	0.608	0.515	0.550	0.633	0.645	0.451	0.553	0.545	0.508	0.568	0.607	**0.553**
EEGNet-2.32	0.605	0.501	0.515	0.499	0.483	0.482	0.501	0.527	0.578	0.568	0.434	0.599	0.555	0.609	0.596	0.623	0.491	0.620	0.480	0.517	0.529	0.532	**0.538**
EEGNet-4.8	0.551	0.473	0.525	0.518	0.606	0.481	0.479	0.547	0.537	0.580	0.564	0.582	0.509	0.616	0.595	0.642	0.534	0.569	0.541	0.527	0.455	0.585	**0.546**

Table B.43: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.558	0.471	0.529	0.500	0.529	0.510	0.606	0.625	0.587	0.740	0.606	0.606	0.567	0.606	0.452	0.596	0.558	0.538	0.529	0.490	0.510	0.529	**0.556**
MDM	0.596	0.413	0.567	0.490	0.519	0.635	0.577	0.567	0.625	0.644	0.644	0.606	0.567	0.538	0.481	0.587	0.500	0.548	0.538	0.538	0.538	0.538	**0.557**
TS+LR	0.558	0.413	0.538	0.471	0.519	0.577	0.596	0.606	0.615	0.750	0.635	0.615	0.587	0.548	0.452	0.577	0.577	0.538	0.558	0.490	0.558	0.577	**0.562**
D-ConvNet	0.519	0.460	0.536	0.508	0.540	0.459	0.470	0.540	0.539	0.560	0.568	0.480	0.507	0.510	0.481	0.537	0.500	0.619	0.516	0.491	0.556	0.530	**0.519**
S-ConvNet	0.423	0.490	0.615	0.469	0.567	0.451	0.535	0.536	0.607	0.656	0.522	0.584	0.595	0.505	0.536	0.641	0.531	0.598	0.613	0.479	0.457	0.606	**0.546**
EEGNet-2.32	0.596	0.434	0.520	0.491	0.529	0.421	0.491	0.558	0.588	0.643	0.490	0.568	0.587	0.505	0.577	0.633	0.532	0.604	0.481	0.459	0.565	0.565	**0.538**
EEGNet-4.8	0.462	0.500	0.555	0.466	0.673	0.490	0.484	0.567	0.515	0.606	0.519	0.618	0.538	0.544	0.562	0.654	0.520	0.579	0.541	0.509	0.472	0.566	**0.543**

Table B.44: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.567	0.442	0.529	0.481	0.490	0.510	0.606	0.615	0.587	0.769	0.587	0.577	0.577	0.529	0.481	0.567	0.558	0.548	0.548	0.538	0.558	0.567	**0.556**
MDM	0.587	0.423	0.548	0.481	0.500	0.635	0.577	0.567	0.625	0.654	0.644	0.615	0.577	0.548	0.490	0.596	0.500	0.538	0.529	0.548	0.519	0.558	**0.557**
TS+LR	0.615	0.413	0.538	0.471	0.500	0.577	0.587	0.615	0.596	0.779	0.635	0.635	0.587	0.519	0.462	0.587	0.587	0.538	0.548	0.529	0.519	0.577	**0.564**
D-ConvNet	0.513	0.411	0.478	0.480	0.482	0.490	0.452	0.528	0.529	0.639	0.548	0.493	0.536	0.549	0.595	0.651	0.478	0.598	0.429	0.548	0.452	0.521	**0.518**
S-ConvNet	0.490	0.468	0.602	0.499	0.486	0.461	0.568	0.519	0.591	0.636	0.550	0.562	0.574	0.512	0.585	0.668	0.539	0.581	0.567	0.544	0.451	0.625	**0.549**
EEGNet-2.32	0.567	0.423	0.527	0.451	0.415	0.509	0.481	0.548	0.607	0.626	0.501	0.538	0.566	0.543	0.525	0.624	0.512	0.675	0.549	0.545	0.597	0.555	**0.540**
EEGNet-4.8	0.596	0.463	0.472	0.499	0.559	0.487	0.503	0.536	0.495	0.560	0.499	0.579	0.539	0.501	0.519	0.671	0.522	0.546	0.509	0.558	0.470	0.538	**0.528**

Table B.45: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe from left side of the brain for MI vs. rest discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.567	0.433	0.490	0.577	0.577	0.500	0.558	0.606	0.529	0.606	0.587	0.500	0.510	0.394	0.529	0.510	0.567	0.462	0.587	0.481	0.529	0.452	**0.525**
MDM	0.567	0.462	0.490	0.490	0.606	0.500	0.596	0.577	0.567	0.635	0.625	0.548	0.529	0.500	0.490	0.577	0.567	0.510	0.587	0.433	0.510	0.385	**0.534**
TS+LR	0.500	0.462	0.462	0.538	0.596	0.481	0.587	0.615	0.529	0.625	0.596	0.538	0.577	0.404	0.500	0.548	0.606	0.500	0.606	0.433	0.519	0.423	**0.529**
D-ConvNet	0.463	0.462	0.491	0.547	0.445	0.471	0.508	0.498	0.523	0.625	0.501	0.625	0.566	0.520	0.528	0.623	0.470	0.531	0.445	0.432	0.513	0.522	**0.514**
S-ConvNet	0.463	0.495	0.525	0.529	0.512	0.385	0.541	0.606	0.536	0.653	0.528	0.572	0.545	0.590	0.663	0.673	0.578	0.608	0.605	0.528	0.565	0.508	**0.555**
EEGNet-2.32	0.485	0.405	0.488	0.539	0.590	0.492	0.425	0.546	0.493	0.724	0.502	0.500	0.452	0.597	0.615	0.653	0.474	0.560	0.563	0.556	0.508	0.481	**0.529**
EEGNet-4.8	0.511	0.487	0.459	0.521	0.575	0.489	0.480	0.577	0.548	0.671	0.570	0.539	0.472	0.647	0.517	0.757	0.492	0.561	0.514	0.421	0.550	0.500	**0.539**

Table B.46: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe from left side of the brain for MI vs. rest discriminating task. Signals filtered in (4-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.500	0.471	0.519	0.538	0.596	0.481	0.567	0.615	0.529	0.615	0.587	0.481	0.529	0.423	0.481	0.519	0.558	0.510	0.596	0.500	0.481	0.442	**0.524**
MDM	0.500	0.452	0.490	0.510	0.625	0.490	0.606	0.587	0.577	0.635	0.635	0.548	0.538	0.510	0.481	0.577	0.510	0.519	0.606	0.433	0.481	0.385	**0.531**
TS+LR	0.538	0.462	0.471	0.567	0.606	0.481	0.587	0.615	0.529	0.635	0.577	0.538	0.548	0.423	0.481	0.558	0.606	0.500	0.615	0.404	0.510	0.423	**0.531**
D-ConvNet	0.434	0.472	0.407	0.511	0.626	0.510	0.461	0.499	0.529	0.596	0.584	0.551	0.499	0.502	0.566	0.614	0.471	0.464	0.572	0.518	0.470	0.536	**0.518**
S-ConvNet	0.466	0.415	0.506	0.529	0.474	0.365	0.501	0.636	0.516	0.614	0.529	0.543	0.516	0.631	0.633	0.674	0.625	0.581	0.583	0.520	0.575	0.491	**0.542**
EEGNet-2.32	0.539	0.431	0.480	0.566	0.598	0.491	0.434	0.566	0.465	0.666	0.512	0.510	0.471	0.660	0.624	0.631	0.521	0.573	0.596	0.557	0.532	0.508	**0.542**
EEGNet-4.8	0.499	0.443	0.513	0.558	0.559	0.498	0.481	0.538	0.508	0.691	0.531	0.518	0.509	0.637	0.604	0.634	0.576	0.586	0.550	0.423	0.596	0.529	**0.545**

Table B.47: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe from left side of the brain for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.558	0.490	0.538	0.577	0.587	0.423	0.587	0.567	0.548	0.721	0.654	0.538	0.548	0.433	0.519	0.548	0.510	0.462	0.548	0.500	0.567	0.567	**0.545**
MDM	0.510	0.462	0.548	0.510	0.587	0.577	0.606	0.577	0.606	0.692	0.644	0.567	0.577	0.500	0.519	0.606	0.500	0.500	0.567	0.490	0.548	0.510	**0.555**
TS+LR	0.538	0.471	0.538	0.567	0.644	0.529	0.538	0.606	0.587	0.721	0.644	0.548	0.548	0.500	0.558	0.558	0.587	0.490	0.587	0.500	0.538	0.500	**0.559**
D-ConvNet	0.480	0.490	0.481	0.558	0.576	0.481	0.461	0.537	0.471	0.578	0.521	0.511	0.587	0.548	0.574	0.654	0.421	0.569	0.526	0.551	0.452	0.531	**0.525**
S-ConvNet	0.366	0.485	0.500	0.545	0.503	0.357	0.542	0.510	0.415	0.493	0.586	0.582	0.645	0.571	0.625	0.645	0.577	0.551	0.613	0.608	0.520	0.501	**0.534**
EEGNet-2.32	0.490	0.350	0.471	0.604	0.635	0.483	0.460	0.546	0.497	0.618	0.522	0.492	0.460	0.524	0.447	0.653	0.576	0.551	0.605	0.472	0.568	0.512	**0.524**
EEGNet-4.8	0.471	0.455	0.452	0.565	0.632	0.486	0.490	0.538	0.501	0.607	0.528	0.510	0.526	0.589	0.535	0.643	0.532	0.579	0.548	0.460	0.545	0.521	**0.532**

Table B.48: Classification accuracy obtained by different classifiers for 22 subjects using 3 channel layout over frontal lobe from left side of the brain for MI vs. rest discriminating task. Signals filtered in (8-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
CSP+LDA	0.548	0.452	0.577	0.510	0.615	0.452	0.596	0.577	0.529	0.712	0.644	0.558	0.587	0.500	0.606	0.558	0.529	0.471	0.558	0.462	0.538	0.567	**0.552**
MDM	0.538	0.471	0.548	0.510	0.587	0.606	0.606	0.567	0.606	0.692	0.644	0.577	0.577	0.500	0.538	0.606	0.500	0.500	0.587	0.500	0.538	0.519	**0.560**
TS+LR	0.548	0.452	0.596	0.577	0.625	0.490	0.567	0.615	0.587	0.683	0.644	0.548	0.567	0.481	0.567	0.596	0.567	0.481	0.606	0.519	0.529	0.538	**0.563**
D-ConvNet	0.482	0.457	0.452	0.541	0.549	0.469	0.470	0.555	0.541	0.578	0.558	0.547	0.624	0.588	0.556	0.661	0.501	0.529	0.557	0.541	0.463	0.547	**0.535**
S-ConvNet	0.423	0.525	0.462	0.566	0.493	0.445	0.465	0.548	0.451	0.560	0.548	0.593	0.566	0.602	0.679	0.654	0.548	0.610	0.595	0.539	0.535	0.540	**0.543**
EEGNet-2.32	0.491	0.389	0.537	0.538	0.550	0.475	0.471	0.537	0.514	0.724	0.504	0.483	0.478	0.530	0.515	0.635	0.490	0.551	0.495	0.497	0.635	0.522	**0.525**
EEGNet-4.8	0.510	0.595	0.530	0.518	0.566	0.448	0.499	0.547	0.491	0.646	0.512	0.500	0.567	0.552	0.565	0.673	0.549	0.530	0.522	0.442	0.500	0.578	**0.538**

Appendix C

Classification results for MNS vs. MI+MNS task using different EEGNet architectures

In this section, classification accuracy results obtained by 8 different EEGNet architectures (EEGNet-2.32, EEGNet-2.16, EEGNet-2.8, EEGNet-2.4, EEGNet-4.32, EEGNet-4.16, EEGNet-4.8, EEGNet-4.4) described in details in Section 3.6.4 for 22 subjects from dataset introduced in Section 4.3 are presented. Results are represented for 12 different channel layouts for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation.

Table C.1: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 128 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.923	0.954	0.961	0.961	0.825	0.587	1.000	0.971	1.000	0.804	0.923	0.951	0.991	0.962	0.865	0.990	0.915	0.990	0.916	0.901	0.857	0.886	**0.915**
EEGNet-2.16	0.922	0.991	0.942	0.981	0.809	0.568	0.990	0.982	1.000	0.829	0.931	0.950	0.981	0.962	0.970	0.990	0.877	0.990	0.982	0.961	0.898	0.972	**0.931**
EEGNet-2.8	0.943	0.991	0.921	0.970	0.874	0.567	1.000	0.981	1.000	0.877	0.962	0.941	0.991	0.972	0.943	0.991	0.935	0.990	0.981	0.971	0.898	0.981	**0.940**
EEGNet-2.4	0.970	0.991	0.940	0.951	0.884	0.545	1.000	0.973	1.000	0.923	0.971	0.950	0.991	0.981	0.972	1.000	0.953	0.990	0.954	0.952	0.897	0.963	**0.943**
EEGNet-4.32	0.873	0.962	0.863	0.932	0.767	0.681	0.954	0.963	0.960	0.770	0.961	0.923	0.981	0.972	0.914	0.960	0.878	0.970	0.925	0.822	0.869	0.720	**0.892**
EEGNet-4.16	0.933	0.982	0.971	0.951	0.867	0.557	1.000	0.963	1.000	0.878	0.961	0.933	0.990	0.952	0.981	0.961	0.895	0.980	0.963	0.951	0.879	0.952	**0.932**
EEGNet-4.8	0.931	0.991	0.970	0.980	0.843	0.548	0.990	0.972	1.000	0.945	0.960	0.950	0.990	0.972	1.000	1.000	0.954	0.990	0.925	0.970	0.926	0.972	**0.945**
EEGNet-4.4	0.941	0.981	0.930	0.962	0.885	0.586	1.000	0.991	1.000	0.934	0.970	0.942	0.991	0.981	0.970	0.990	0.955	0.990	0.954	0.951	0.962	0.972	**0.947**

Table C.2: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 47 channel layout for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.835	0.896	0.915	0.910	0.739	0.547	0.845	0.933	1.000	0.827	0.865	0.940	0.990	0.923	0.895	0.905	0.865	0.990	0.883	0.875	0.762	0.868	**0.873**
EEGNet-2.16	0.893	0.935	0.933	0.891	0.798	0.545	0.873	0.952	1.000	0.850	0.933	0.951	0.981	0.924	0.933	0.877	0.884	1.000	0.953	0.913	0.896	0.971	**0.904**
EEGNet-2.8	0.875	0.944	0.944	0.861	0.826	0.545	0.913	0.951	1.000	0.782	0.933	0.981	0.990	0.935	0.895	0.894	0.942	0.981	0.953	0.903	0.878	0.952	**0.903**
EEGNet-2.4	0.903	0.935	0.854	0.880	0.825	0.558	0.942	0.943	1.000	0.830	0.894	0.990	0.972	0.934	0.903	0.865	0.933	1.000	0.981	0.884	0.858	0.962	**0.902**
EEGNet-4.32	0.805	0.945	0.912	0.961	0.681	0.643	0.759	0.922	1.000	0.742	0.914	0.932	0.972	0.944	0.942	0.932	0.866	0.980	0.942	0.895	0.849	0.855	**0.881**
EEGNet-4.16	0.891	0.925	0.961	0.921	0.801	0.565	0.893	0.932	1.000	0.780	0.941	0.970	0.972	0.915	0.884	0.865	0.953	0.990	0.961	0.902	0.858	0.961	**0.902**
EEGNet-4.8	0.922	0.954	0.915	0.970	0.895	0.529	0.990	0.952	1.000	0.790	0.971	0.960	0.972	0.933	0.855	0.942	0.943	0.990	0.980	0.903	0.867	0.932	**0.916**
EEGNet-4.4	0.922	0.953	0.845	0.961	0.924	0.500	0.912	0.951	1.000	0.810	0.941	0.990	0.981	0.935	0.924	0.894	0.885	0.981	0.970	0.902	0.889	0.963	**0.910**

Table C.3: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 13 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.815	0.887	0.873	0.940	0.789	0.500	0.701	0.931	0.751	0.726	0.884	0.837	0.860	0.896	0.885	0.903	0.905	0.980	0.885	0.844	0.767	0.885	**0.838**
EEGNet-2.16	0.853	0.917	0.874	0.970	0.796	0.527	0.739	0.951	0.772	0.771	0.902	0.903	0.952	0.905	0.847	0.913	0.822	0.990	0.952	0.865	0.828	0.904	**0.861**
EEGNet-2.8	0.845	0.927	0.902	0.920	0.777	0.539	0.680	0.971	0.825	0.830	0.914	0.875	0.980	0.934	0.885	0.895	0.885	0.971	0.942	0.893	0.819	0.885	**0.868**
EEGNet-2.4	0.873	0.897	0.883	0.900	0.781	0.510	0.721	0.951	0.797	0.770	0.932	0.873	0.906	0.914	0.884	0.902	0.915	0.990	0.942	0.893	0.768	0.903	**0.859**
EEGNet-4.32	0.842	0.905	0.914	0.980	0.790	0.615	0.626	0.930	0.810	0.759	0.872	0.843	0.875	0.915	0.952	0.895	0.846	0.981	0.962	0.923	0.828	0.857	**0.860**
EEGNet-4.16	0.825	0.954	0.833	0.980	0.827	0.500	0.738	0.942	0.797	0.760	0.914	0.943	0.962	0.924	0.885	0.883	0.885	0.990	0.971	0.894	0.857	0.884	**0.870**
EEGNet-4.8	0.873	0.906	0.843	0.960	0.789	0.482	0.718	0.951	0.739	0.772	0.953	0.932	0.973	0.925	0.934	0.886	0.895	0.981	0.952	0.893	0.820	0.874	**0.866**
EEGNet-4.4	0.883	0.936	0.845	0.912	0.798	0.500	0.669	0.961	0.778	0.722	0.932	0.884	0.934	0.905	0.904	0.885	0.886	0.990	0.961	0.884	0.807	0.903	**0.858**

Table C.4: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 13 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.777	0.925	0.885	0.863	0.759	0.643	0.746	0.911	0.820	0.741	0.864	0.885	0.981	0.933	0.837	0.874	0.876	1.000	0.925	0.865	0.811	0.799	**0.851**
EEGNet-2.16	0.769	0.905	0.903	0.903	0.778	0.685	0.844	0.922	0.858	0.781	0.905	0.903	0.972	0.953	0.847	0.873	0.875	1.000	0.935	0.815	0.820	0.742	**0.863**
EEGNet-2.8	0.837	0.925	0.893	0.923	0.785	0.595	0.740	0.913	0.820	0.762	0.912	0.921	1.000	0.953	0.874	0.874	0.905	1.000	0.961	0.845	0.866	0.830	**0.870**
EEGNet-2.4	0.817	0.915	0.922	0.872	0.787	0.633	0.750	0.922	0.819	0.762	0.932	0.923	0.990	0.944	0.866	0.872	0.885	1.000	0.943	0.823	0.790	0.830	**0.863**
EEGNet-4.32	0.762	0.962	0.862	0.950	0.758	0.750	0.765	0.862	0.868	0.789	0.971	0.863	0.981	0.954	0.912	0.855	0.876	1.000	0.942	0.779	0.836	0.722	**0.864**
EEGNet-4.16	0.757	0.886	0.893	0.894	0.836	0.663	0.775	0.883	0.851	0.751	0.893	0.904	0.981	0.953	0.877	0.865	0.896	1.000	0.971	0.834	0.831	0.754	**0.861**
EEGNet-4.8	0.838	0.935	0.894	0.893	0.798	0.665	0.769	0.893	0.850	0.810	0.952	0.940	0.972	0.953	0.884	0.882	0.935	1.000	0.981	0.834	0.916	0.780	**0.880**
EEGNet-4.4	0.827	0.926	0.894	0.905	0.768	0.625	0.777	0.931	0.809	0.800	0.934	0.932	0.990	0.952	0.875	0.834	0.895	1.000	0.963	0.854	0.859	0.788	**0.870**

Table C.5: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 9 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.845	0.905	0.895	0.796	0.733	0.616	0.634	0.903	0.753	0.691	0.864	0.765	0.922	0.845	0.865	0.798	0.865	0.970	0.893	0.748	0.836	0.732	**0.812**
EEGNet-2.16	0.825	0.886	0.855	0.823	0.724	0.635	0.815	0.874	0.682	0.759	0.886	0.759	0.932	0.933	0.895	0.796	0.845	0.962	0.913	0.760	0.826	0.771	**0.825**
EEGNet-2.8	0.854	0.917	0.885	0.825	0.778	0.607	0.727	0.885	0.781	0.715	0.922	0.787	0.953	0.944	0.875	0.786	0.818	0.953	0.950	0.741	0.777	0.827	**0.832**
EEGNet-2.4	0.834	0.877	0.854	0.815	0.782	0.587	0.786	0.846	0.761	0.780	0.934	0.737	0.894	0.886	0.826	0.786	0.838	0.962	0.915	0.787	0.673	0.810	**0.817**
EEGNet-4.32	0.843	0.914	0.932	0.843	0.739	0.694	0.730	0.863	0.829	0.733	0.854	0.824	0.873	0.962	0.961	0.836	0.736	0.961	0.931	0.799	0.875	0.801	**0.842**
EEGNet-4.16	0.815	0.935	0.845	0.825	0.808	0.635	0.757	0.856	0.780	0.742	0.933	0.749	0.895	0.934	0.855	0.806	0.818	0.980	0.942	0.836	0.865	0.856	**0.839**
EEGNet-4.8	0.817	0.935	0.853	0.815	0.838	0.617	0.796	0.875	0.750	0.713	0.913	0.854	0.953	0.934	0.855	0.854	0.847	0.990	0.941	0.845	0.808	0.808	**0.846**
EEGNet-4.4	0.837	0.916	0.873	0.825	0.809	0.607	0.785	0.884	0.741	0.712	0.933	0.795	0.923	0.895	0.874	0.816	0.799	0.980	0.912	0.817	0.741	0.809	**0.831**

Table C.6: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 9 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.845	0.885	0.885	0.854	0.778	0.617	0.752	0.883	0.800	0.703	0.951	0.749	0.884	0.944	0.876	0.825	0.904	1.000	0.934	0.797	0.835	0.673	**0.835**
EEGNet-2.16	0.779	0.895	0.835	0.883	0.756	0.660	0.769	0.845	0.821	0.732	0.915	0.787	0.883	0.935	0.865	0.855	0.885	1.000	0.981	0.789	0.854	0.759	**0.840**
EEGNet-2.8	0.815	0.913	0.882	0.845	0.768	0.643	0.700	0.816	0.847	0.742	0.962	0.776	0.942	0.944	0.875	0.853	0.874	1.000	0.943	0.787	0.808	0.780	**0.842**
EEGNet-2.4	0.794	0.914	0.875	0.865	0.806	0.681	0.855	0.845	0.830	0.770	0.962	0.785	0.953	0.924	0.805	0.815	0.884	1.000	0.962	0.796	0.787	0.807	**0.851**
EEGNet-4.32	0.825	0.905	0.863	0.921	0.760	0.725	0.749	0.865	0.858	0.685	0.990	0.895	0.884	0.915	0.923	0.797	0.903	1.000	0.954	0.786	0.894	0.723	**0.855**
EEGNet-4.16	0.776	0.886	0.874	0.855	0.786	0.679	0.816	0.865	0.800	0.771	0.981	0.806	0.933	0.905	0.905	0.845	0.894	1.000	0.962	0.808	0.778	0.709	**0.847**
EEGNet-4.8	0.806	0.867	0.884	0.835	0.788	0.623	0.767	0.884	0.840	0.811	0.932	0.809	0.933	0.944	0.875	0.854	0.914	1.000	0.981	0.756	0.855	0.757	**0.851**
EEGNet-4.4	0.825	0.925	0.902	0.835	0.778	0.575	0.777	0.846	0.818	0.772	0.934	0.805	0.942	0.934	0.865	0.825	0.914	0.991	0.962	0.816	0.817	0.847	**0.850**

Table C.7: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 6 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.815	0.907	0.749	0.778	0.752	0.577	0.727	0.787	0.655	0.685	0.779	0.805	0.846	0.808	0.817	0.758	0.817	0.971	0.876	0.617	0.818	0.722	**0.776**
EEGNet-2.16	0.813	0.887	0.777	0.788	0.705	0.601	0.795	0.777	0.724	0.692	0.798	0.708	0.951	0.829	0.856	0.786	0.810	0.971	0.886	0.672	0.848	0.779	**0.793**
EEGNet-2.8	0.825	0.916	0.817	0.788	0.715	0.599	0.770	0.805	0.751	0.684	0.786	0.726	0.923	0.819	0.845	0.818	0.798	0.980	0.904	0.748	0.827	0.759	**0.800**
EEGNet-2.4	0.815	0.907	0.749	0.776	0.739	0.625	0.767	0.809	0.837	0.664	0.867	0.758	0.940	0.822	0.884	0.836	0.800	0.980	0.896	0.654	0.742	0.789	**0.803**
EEGNet-4.32	0.746	0.925	0.699	0.865	0.722	0.703	0.748	0.785	0.741	0.720	0.835	0.884	0.835	0.875	0.934	0.758	0.749	0.971	0.875	0.693	0.803	0.721	**0.799**
EEGNet-4.16	0.834	0.897	0.739	0.875	0.723	0.549	0.778	0.776	0.733	0.752	0.829	0.716	0.922	0.818	0.865	0.816	0.772	0.980	0.906	0.731	0.838	0.769	**0.801**
EEGNet-4.8	0.815	0.916	0.748	0.837	0.770	0.558	0.765	0.808	0.781	0.725	0.866	0.740	0.932	0.865	0.875	0.865	0.829	0.980	0.878	0.827	0.847	0.740	**0.817**
EEGNet-4.4	0.786	0.916	0.710	0.768	0.796	0.527	0.756	0.816	0.760	0.674	0.858	0.756	0.874	0.865	0.885	0.855	0.829	0.980	0.888	0.690	0.750	0.751	**0.795**

Table C.8: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 6 channel layout over motor cortex and frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.765	0.867	0.835	0.797	0.750	0.770	0.661	0.894	0.789	0.752	0.825	0.808	0.894	0.886	0.767	0.795	0.885	1.000	0.955	0.675	0.816	0.681	**0.812**
EEGNet-2.16	0.768	0.904	0.855	0.815	0.835	0.720	0.641	0.895	0.788	0.750	0.828	0.799	0.883	0.915	0.766	0.806	0.944	1.000	0.907	0.720	0.796	0.730	**0.821**
EEGNet-2.8	0.769	0.886	0.874	0.863	0.777	0.699	0.672	0.865	0.810	0.771	0.771	0.760	0.834	0.935	0.736	0.805	0.943	1.000	0.887	0.712	0.778	0.713	**0.812**
EEGNet-2.4	0.818	0.887	0.853	0.845	0.770	0.699	0.603	0.895	0.811	0.789	0.780	0.777	0.785	0.886	0.766	0.825	0.895	1.000	0.896	0.645	0.739	0.799	**0.807**
EEGNet-4.32	0.806	0.924	0.893	0.895	0.761	0.720	0.553	0.902	0.867	0.725	0.848	0.787	0.795	0.935	0.923	0.815	0.893	1.000	0.944	0.665	0.790	0.712	**0.825**
EEGNet-4.16	0.816	0.885	0.825	0.816	0.807	0.798	0.583	0.885	0.848	0.792	0.750	0.779	0.835	0.926	0.755	0.775	0.944	1.000	0.932	0.683	0.798	0.709	**0.816**
EEGNet-4.8	0.808	0.905	0.875	0.845	0.795	0.698	0.633	0.894	0.877	0.774	0.818	0.769	0.902	0.935	0.815	0.795	0.925	1.000	0.915	0.738	0.806	0.791	**0.832**
EEGNet-4.4	0.797	0.924	0.865	0.816	0.796	0.664	0.559	0.885	0.819	0.819	0.806	0.827	0.845	0.925	0.776	0.805	0.943	1.000	0.915	0.700	0.779	0.691	**0.816**

Table C.9: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 6 channel layout over motor cortex and frontal lobe from left side of the brain for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.721	0.846	0.883	0.813	0.751	0.719	0.761	0.894	0.887	0.779	0.914	0.816	0.625	0.933	0.875	0.834	0.885	1.000	0.963	0.788	0.806	0.719	**0.828**
EEGNet-2.16	0.876	0.846	0.892	0.803	0.712	0.729	0.722	0.865	0.868	0.760	0.922	0.815	0.677	0.915	0.905	0.862	0.895	1.000	0.953	0.824	0.835	0.680	**0.834**
EEGNet-2.8	0.847	0.895	0.912	0.811	0.702	0.651	0.675	0.845	0.897	0.741	0.922	0.867	0.690	0.915	0.905	0.882	0.876	1.000	0.953	0.826	0.826	0.716	**0.834**
EEGNet-2.4	0.809	0.885	0.834	0.827	0.768	0.706	0.676	0.809	0.876	0.779	0.953	0.816	0.707	0.905	0.876	0.873	0.829	1.000	0.944	0.835	0.797	0.712	**0.828**
EEGNet-4.32	0.848	0.904	0.843	0.864	0.779	0.788	0.647	0.912	0.867	0.777	0.952	0.845	0.722	0.925	0.961	0.835	0.876	0.991	0.962	0.808	0.894	0.694	**0.850**
EEGNet-4.16	0.825	0.847	0.854	0.841	0.761	0.710	0.693	0.844	0.888	0.778	0.904	0.844	0.699	0.934	0.885	0.855	0.914	1.000	0.971	0.844	0.865	0.703	**0.839**
EEGNet-4.8	0.836	0.885	0.901	0.870	0.771	0.717	0.655	0.845	0.905	0.730	0.902	0.855	0.726	0.944	0.905	0.844	0.895	1.000	0.953	0.795	0.847	0.702	**0.840**
EEGNet-4.4	0.864	0.875	0.844	0.840	0.740	0.777	0.675	0.835	0.915	0.820	0.952	0.893	0.775	0.934	0.876	0.873	0.876	1.000	0.972	0.816	0.816	0.713	**0.849**

Table C.10: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 3 channel layout over motor cortex for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.775	0.867	0.815	0.806	0.644	0.548	0.634	0.845	0.684	0.665	0.744	0.721	0.843	0.835	0.709	0.795	0.808	0.953	0.882	0.639	0.696	0.635	**0.752**
EEGNet-2.16	0.752	0.878	0.826	0.788	0.672	0.579	0.670	0.836	0.684	0.635	0.721	0.798	0.873	0.815	0.727	0.766	0.809	0.952	0.863	0.536	0.665	0.684	**0.751**
EEGNet-2.8	0.739	0.916	0.885	0.749	0.682	0.635	0.615	0.884	0.702	0.625	0.712	0.720	0.832	0.845	0.698	0.757	0.810	0.923	0.863	0.658	0.645	0.705	**0.755**
EEGNet-2.4	0.731	0.877	0.873	0.722	0.654	0.635	0.673	0.875	0.721	0.617	0.715	0.749	0.834	0.845	0.686	0.740	0.838	0.953	0.855	0.578	0.626	0.701	**0.750**
EEGNet-4.32	0.788	0.916	0.894	0.835	0.597	0.605	0.701	0.864	0.724	0.693	0.743	0.779	0.825	0.895	0.853	0.767	0.778	0.962	0.835	0.545	0.752	0.708	**0.775**
EEGNet-4.16	0.741	0.915	0.875	0.768	0.702	0.576	0.603	0.875	0.732	0.654	0.739	0.769	0.834	0.787	0.747	0.796	0.857	0.952	0.882	0.661	0.722	0.713	**0.768**
EEGNet-4.8	0.750	0.916	0.866	0.804	0.701	0.624	0.652	0.856	0.683	0.685	0.712	0.749	0.863	0.827	0.670	0.767	0.818	0.962	0.893	0.670	0.664	0.665	**0.764**
EEGNet-4.4	0.741	0.859	0.882	0.777	0.671	0.623	0.670	0.855	0.664	0.635	0.714	0.779	0.793	0.847	0.669	0.797	0.847	0.942	0.845	0.575	0.625	0.732	**0.752**

Table C.11: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 3 channel layout over frontal lobe for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.567	0.624	0.579	0.546	0.732	0.688	0.623	0.605	0.664	0.734	0.525	0.490	0.537	0.817	0.788	0.716	0.645	1.000	0.768	0.613	0.672	0.693	**0.665**
EEGNet-2.16	0.597	0.664	0.579	0.500	0.760	0.642	0.592	0.555	0.671	0.714	0.585	0.501	0.470	0.825	0.845	0.737	0.645	0.991	0.759	0.451	0.684	0.645	**0.655**
EEGNet-2.8	0.587	0.576	0.565	0.548	0.752	0.723	0.531	0.625	0.702	0.683	0.615	0.518	0.507	0.856	0.778	0.756	0.502	0.991	0.758	0.461	0.644	0.673	**0.652**
EEGNet-2.4	0.601	0.568	0.490	0.557	0.750	0.633	0.538	0.615	0.675	0.675	0.595	0.521	0.452	0.847	0.759	0.786	0.559	0.991	0.758	0.547	0.682	0.674	**0.649**
EEGNet-4.32	0.644	0.626	0.510	0.695	0.721	0.706	0.564	0.722	0.608	0.685	0.487	0.500	0.517	0.788	0.778	0.687	0.585	0.991	0.756	0.587	0.710	0.654	**0.660**
EEGNet-4.16	0.589	0.663	0.548	0.517	0.764	0.694	0.625	0.645	0.713	0.752	0.575	0.491	0.480	0.846	0.799	0.745	0.636	1.000	0.768	0.489	0.713	0.615	**0.667**
EEGNet-4.8	0.577	0.618	0.508	0.548	0.732	0.736	0.587	0.645	0.684	0.732	0.641	0.587	0.462	0.865	0.778	0.728	0.548	1.000	0.758	0.559	0.645	0.614	**0.662**
EEGNet-4.4	0.579	0.567	0.527	0.586	0.750	0.650	0.584	0.644	0.684	0.691	0.663	0.492	0.509	0.837	0.762	0.785	0.521	0.991	0.768	0.510	0.682	0.672	**0.657**

Table C.12: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 3 channel layout over frontal lobe from left side of the brain for MNS vs. MNS+MI discriminating task. Signals filtered in (4-38 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.635	0.560	0.499	0.754	0.752	0.612	0.655	0.596	0.869	0.738	0.612	0.441	0.534	0.825	0.653	0.785	0.723	1.000	0.758	0.462	0.663	0.583	**0.669**
EEGNet-2.16	0.743	0.711	0.596	0.714	0.749	0.562	0.638	0.581	0.790	0.741	0.634	0.546	0.495	0.815	0.683	0.775	0.703	1.000	0.777	0.551	0.645	0.595	**0.684**
EEGNet-2.8	0.695	0.701	0.614	0.733	0.760	0.574	0.616	0.617	0.780	0.735	0.605	0.664	0.557	0.865	0.621	0.825	0.608	1.000	0.767	0.605	0.684	0.635	**0.694**
EEGNet-2.4	0.655	0.652	0.536	0.804	0.750	0.592	0.597	0.617	0.780	0.762	0.545	0.633	0.575	0.835	0.586	0.862	0.521	1.000	0.777	0.636	0.703	0.662	**0.685**
EEGNet-4.32	0.567	0.606	0.536	0.802	0.751	0.664	0.580	0.703	0.810	0.712	0.693	0.507	0.462	0.827	0.722	0.845	0.654	1.000	0.766	0.540	0.694	0.605	**0.684**
EEGNet-4.16	0.706	0.750	0.549	0.755	0.780	0.612	0.635	0.578	0.838	0.759	0.594	0.545	0.537	0.825	0.653	0.825	0.710	1.000	0.785	0.527	0.683	0.693	**0.697**
EEGNet-4.8	0.685	0.680	0.595	0.744	0.742	0.612	0.619	0.543	0.819	0.781	0.574	0.606	0.557	0.826	0.651	0.834	0.605	1.000	0.776	0.633	0.721	0.653	**0.693**
EEGNet-4.4	0.637	0.640	0.575	0.764	0.761	0.585	0.630	0.537	0.799	0.789	0.661	0.559	0.555	0.815	0.615	0.793	0.577	1.000	0.777	0.635	0.665	0.644	**0.682**

Appendix D

Classification results for MI vs. rest task using different EEGNet architectures

In this section, classification accuracy results obtained by 8 different EEGNet architectures (EEGNet-2.32, EEGNet-2.16, EEGNet-2.8, EEGNet-2.4, EEGNet-4.32, EEGNet-4.16, EEGNet-4.8, EEGNet-4.4) described in details in Section 3.6.4 for 22 subjects from dataset introduced in Section 4.3 are presented. Results are represented for 12 different channel layouts for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation.

Table D.1: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 128 channel layout for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.501	0.658	0.746	0.510	0.749	0.490	0.713	0.598	0.635	0.784	0.637	0.749	0.655	0.731	0.714	0.789	0.585	0.684	0.932	0.706	0.777	0.542	**0.677**
EEGNet-2.16	0.550	0.675	0.789	0.481	0.723	0.513	0.721	0.680	0.722	0.779	0.709	0.737	0.654	0.660	0.761	0.756	0.558	0.615	0.942	0.815	0.865	0.644	**0.698**
EEGNet-2.8	0.615	0.645	0.770	0.520	0.685	0.578	0.712	0.692	0.685	0.706	0.633	0.655	0.625	0.705	0.724	0.795	0.666	0.636	0.952	0.768	0.796	0.574	**0.688**
EEGNet-2.4	0.498	0.754	0.763	0.500	0.751	0.558	0.685	0.708	0.605	0.790	0.672	0.702	0.645	0.766	0.724	0.805	0.684	0.596	0.951	0.787	0.810	0.517	**0.694**
EEGNet-4.32	0.558	0.740	0.615	0.530	0.616	0.702	0.635	0.595	0.568	0.713	0.587	0.634	0.685	0.702	0.561	0.651	0.598	0.653	0.912	0.663	0.662	0.512	**0.641**
EEGNet-4.16	0.550	0.771	0.782	0.578	0.705	0.541	0.759	0.641	0.654	0.751	0.711	0.767	0.633	0.714	0.740	0.767	0.609	0.729	0.942	0.707	0.787	0.598	**0.702**
EEGNet-4.8	0.461	0.700	0.789	0.547	0.683	0.609	0.730	0.710	0.640	0.743	0.740	0.730	0.664	0.778	0.790	0.815	0.645	0.675	0.942	0.698	0.777	0.502	**0.699**
EEGNet-4.4	0.553	0.751	0.791	0.490	0.701	0.548	0.713	0.711	0.548	0.753	0.738	0.787	0.651	0.826	0.735	0.855	0.713	0.625	0.921	0.691	0.796	0.539	**0.702**

Table D.2: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 47 channel layout for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.565	0.712	0.751	0.490	0.750	0.549	0.760	0.703	0.625	0.829	0.682	0.759	0.664	0.751	0.750	0.795	0.634	0.597	0.931	0.816	0.887	0.606	**0.709**
EEGNet-2.16	0.566	0.715	0.741	0.509	0.693	0.507	0.730	0.703	0.625	0.749	0.692	0.748	0.653	0.760	0.751	0.815	0.620	0.772	0.941	0.864	0.868	0.635	**0.712**
EEGNet-2.8	0.617	0.788	0.759	0.489	0.682	0.499	0.723	0.695	0.596	0.713	0.645	0.671	0.664	0.771	0.715	0.785	0.646	0.722	0.952	0.835	0.838	0.645	**0.702**
EEGNet-2.4	0.535	0.751	0.752	0.519	0.635	0.511	0.780	0.701	0.567	0.847	0.703	0.721	0.732	0.798	0.695	0.805	0.647	0.657	0.942	0.785	0.858	0.703	**0.711**
EEGNet-4.32	0.510	0.701	0.732	0.520	0.558	0.691	0.605	0.693	0.596	0.723	0.618	0.671	0.658	0.692	0.751	0.777	0.665	0.565	0.972	0.747	0.779	0.507	**0.670**
EEGNet-4.16	0.625	0.789	0.782	0.499	0.690	0.567	0.733	0.780	0.626	0.725	0.672	0.711	0.644	0.730	0.725	0.817	0.694	0.751	0.960	0.815	0.867	0.578	**0.717**
EEGNet-4.8	0.603	0.778	0.770	0.502	0.644	0.558	0.770	0.711	0.567	0.753	0.656	0.700	0.692	0.731	0.733	0.816	0.611	0.704	0.941	0.815	0.827	0.590	**0.703**
EEGNet-4.4	0.633	0.760	0.781	0.500	0.711	0.586	0.751	0.771	0.575	0.793	0.633	0.721	0.662	0.736	0.667	0.816	0.627	0.674	0.931	0.825	0.848	0.655	**0.712**

Table D.3: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 13 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.614	0.714	0.791	0.500	0.751	0.494	0.797	0.684	0.658	0.768	0.594	0.751	0.682	0.701	0.734	0.767	0.577	0.732	0.932	0.825	0.848	0.614	**0.706**
EEGNet-2.16	0.520	0.771	0.845	0.509	0.730	0.520	0.789	0.771	0.731	0.768	0.691	0.733	0.692	0.720	0.752	0.836	0.588	0.722	0.951	0.845	0.856	0.585	**0.724**
EEGNet-2.8	0.576	0.781	0.855	0.529	0.683	0.443	0.816	0.625	0.598	0.741	0.664	0.731	0.701	0.759	0.705	0.815	0.547	0.790	0.941	0.865	0.818	0.629	**0.710**
EEGNet-2.4	0.554	0.773	0.847	0.500	0.684	0.498	0.769	0.732	0.664	0.780	0.585	0.742	0.741	0.748	0.647	0.788	0.605	0.673	0.943	0.806	0.809	0.664	**0.707**
EEGNet-4.32	0.492	0.800	0.753	0.490	0.606	0.674	0.686	0.674	0.589	0.724	0.663	0.713	0.704	0.725	0.675	0.796	0.655	0.691	0.942	0.720	0.848	0.464	**0.686**
EEGNet-4.16	0.606	0.753	0.875	0.480	0.705	0.558	0.800	0.702	0.636	0.759	0.750	0.761	0.645	0.800	0.734	0.798	0.635	0.751	0.941	0.765	0.828	0.596	**0.722**
EEGNet-4.8	0.606	0.780	0.876	0.482	0.674	0.575	0.770	0.635	0.649	0.724	0.622	0.788	0.691	0.739	0.675	0.817	0.597	0.722	0.941	0.767	0.865	0.701	**0.713**
EEGNet-4.4	0.540	0.809	0.876	0.491	0.662	0.540	0.820	0.673	0.651	0.753	0.615	0.722	0.692	0.768	0.617	0.788	0.565	0.684	0.932	0.787	0.810	0.693	**0.704**

Table D.4: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 13 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.520	0.735	0.759	0.819	0.712	0.550	0.797	0.695	0.642	0.722	0.633	0.664	0.664	0.741	0.664	0.789	0.541	0.759	0.942	0.776	0.770	0.645	**0.706**
EEGNet-2.16	0.568	0.741	0.789	0.822	0.683	0.625	0.808	0.675	0.653	0.753	0.632	0.647	0.654	0.686	0.734	0.806	0.681	0.801	0.951	0.745	0.694	0.607	**0.716**
EEGNet-2.8	0.565	0.771	0.797	0.849	0.585	0.556	0.808	0.741	0.589	0.781	0.682	0.644	0.720	0.758	0.715	0.757	0.673	0.733	0.942	0.806	0.743	0.607	**0.719**
EEGNet-2.4	0.576	0.750	0.808	0.828	0.548	0.557	0.818	0.674	0.706	0.763	0.663	0.628	0.702	0.681	0.684	0.797	0.654	0.683	0.943	0.786	0.705	0.634	**0.709**
EEGNet-4.32	0.489	0.752	0.740	0.714	0.625	0.662	0.761	0.647	0.606	0.767	0.666	0.577	0.653	0.738	0.615	0.748	0.595	0.733	0.941	0.750	0.782	0.591	**0.689**
EEGNet-4.16	0.531	0.742	0.799	0.770	0.714	0.557	0.790	0.715	0.685	0.750	0.685	0.677	0.625	0.758	0.685	0.788	0.645	0.840	0.932	0.777	0.807	0.625	**0.723**
EEGNet-4.8	0.568	0.742	0.798	0.848	0.625	0.585	0.799	0.704	0.683	0.817	0.702	0.735	0.729	0.731	0.714	0.775	0.655	0.811	0.961	0.745	0.791	0.713	**0.738**
EEGNet-4.4	0.580	0.742	0.798	0.840	0.606	0.540	0.807	0.733	0.624	0.801	0.681	0.720	0.721	0.680	0.647	0.719	0.683	0.635	0.951	0.766	0.742	0.715	**0.715**

Table D.5: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 9 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.481	0.790	0.809	0.741	0.661	0.741	0.732	0.685	0.667	0.780	0.557	0.695	0.714	0.709	0.675	0.818	0.567	0.655	0.962	0.825	0.855	0.674	**0.718**
EEGNet-2.16	0.562	0.771	0.837	0.771	0.660	0.623	0.732	0.635	0.636	0.779	0.652	0.677	0.742	0.749	0.706	0.826	0.664	0.720	0.942	0.815	0.819	0.665	**0.727**
EEGNet-2.8	0.483	0.760	0.856	0.781	0.612	0.622	0.715	0.664	0.635	0.768	0.610	0.745	0.713	0.769	0.645	0.795	0.638	0.733	0.952	0.806	0.837	0.713	**0.721**
EEGNet-2.4	0.539	0.759	0.828	0.760	0.691	0.673	0.751	0.694	0.618	0.769	0.603	0.693	0.721	0.721	0.676	0.797	0.595	0.692	0.951	0.796	0.829	0.644	**0.718**
EEGNet-4.32	0.465	0.787	0.808	0.713	0.643	0.752	0.741	0.665	0.560	0.702	0.739	0.685	0.725	0.693	0.686	0.816	0.550	0.705	0.933	0.799	0.897	0.557	**0.710**
EEGNet-4.16	0.596	0.759	0.847	0.742	0.663	0.645	0.759	0.693	0.615	0.760	0.605	0.712	0.645	0.785	0.665	0.826	0.574	0.779	0.952	0.845	0.846	0.635	**0.725**
EEGNet-4.8	0.484	0.780	0.847	0.772	0.712	0.682	0.733	0.684	0.643	0.712	0.650	0.763	0.751	0.787	0.656	0.805	0.615	0.732	0.951	0.786	0.855	0.670	**0.730**
EEGNet-4.4	0.492	0.720	0.839	0.742	0.691	0.604	0.788	0.723	0.556	0.752	0.605	0.723	0.741	0.712	0.618	0.778	0.567	0.715	0.933	0.797	0.790	0.625	**0.705**

Table D.6: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 9 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.527	0.725	0.768	0.770	0.634	0.615	0.761	0.734	0.645	0.769	0.625	0.540	0.662	0.641	0.656	0.740	0.582	0.622	0.961	0.817	0.801	0.596	**0.690**
EEGNet-2.16	0.558	0.797	0.771	0.809	0.695	0.596	0.762	0.706	0.565	0.778	0.690	0.657	0.762	0.692	0.705	0.698	0.674	0.627	0.941	0.822	0.840	0.637	**0.717**
EEGNet-2.8	0.615	0.740	0.768	0.772	0.661	0.605	0.789	0.647	0.605	0.797	0.645	0.657	0.733	0.692	0.683	0.826	0.665	0.695	0.952	0.824	0.789	0.605	**0.717**
EEGNet-2.4	0.560	0.749	0.769	0.820	0.560	0.517	0.781	0.665	0.613	0.644	0.682	0.647	0.626	0.684	0.638	0.769	0.732	0.644	0.942	0.795	0.744	0.644	**0.692**
EEGNet-4.32	0.474	0.758	0.816	0.760	0.606	0.711	0.762	0.644	0.556	0.705	0.701	0.558	0.635	0.731	0.619	0.748	0.521	0.556	0.942	0.786	0.735	0.463	**0.672**
EEGNet-4.16	0.628	0.732	0.750	0.789	0.730	0.558	0.790	0.735	0.615	0.714	0.653	0.662	0.627	0.673	0.665	0.758	0.636	0.652	0.961	0.815	0.772	0.585	**0.705**
EEGNet-4.8	0.588	0.750	0.779	0.810	0.625	0.598	0.768	0.704	0.633	0.684	0.634	0.626	0.750	0.760	0.684	0.747	0.591	0.656	0.952	0.824	0.831	0.643	**0.711**
EEGNet-4.4	0.579	0.732	0.722	0.819	0.676	0.634	0.798	0.754	0.606	0.751	0.634	0.576	0.662	0.684	0.659	0.739	0.580	0.655	0.943	0.795	0.802	0.693	**0.704**

Table D.7: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 6 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.564	0.817	0.771	0.725	0.730	0.615	0.730	0.735	0.633	0.691	0.692	0.607	0.634	0.721	0.618	0.778	0.489	0.694	0.923	0.788	0.837	0.542	**0.697**
EEGNet-2.16	0.557	0.796	0.774	0.790	0.702	0.565	0.721	0.686	0.617	0.731	0.641	0.605	0.585	0.729	0.685	0.786	0.595	0.753	0.932	0.827	0.821	0.605	**0.705**
EEGNet-2.8	0.568	0.737	0.789	0.782	0.664	0.582	0.750	0.665	0.635	0.704	0.650	0.598	0.663	0.699	0.655	0.757	0.547	0.724	0.942	0.728	0.739	0.532	**0.687**
EEGNet-2.4	0.565	0.681	0.722	0.781	0.673	0.613	0.731	0.694	0.626	0.681	0.680	0.617	0.605	0.720	0.683	0.787	0.596	0.635	0.942	0.786	0.752	0.563	**0.688**
EEGNet-4.32	0.548	0.809	0.761	0.694	0.565	0.625	0.675	0.761	0.606	0.664	0.731	0.589	0.673	0.770	0.722	0.777	0.461	0.625	0.924	0.670	0.800	0.419	**0.676**
EEGNet-4.16	0.521	0.788	0.792	0.770	0.625	0.623	0.713	0.733	0.616	0.730	0.690	0.704	0.652	0.684	0.635	0.749	0.538	0.743	0.942	0.756	0.800	0.521	**0.697**
EEGNet-4.8	0.559	0.787	0.723	0.781	0.722	0.650	0.741	0.694	0.655	0.684	0.634	0.683	0.673	0.712	0.666	0.749	0.552	0.735	0.962	0.729	0.770	0.542	**0.700**
EEGNet-4.4	0.499	0.787	0.790	0.771	0.653	0.623	0.789	0.724	0.576	0.673	0.565	0.695	0.635	0.730	0.666	0.779	0.595	0.640	0.942	0.758	0.744	0.513	**0.688**

Table D.8: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 6 channel layout over motor cortex and frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.557	0.685	0.720	0.762	0.617	0.575	0.661	0.674	0.545	0.807	0.633	0.691	0.672	0.701	0.703	0.767	0.723	0.712	0.905	0.702	0.778	0.730	**0.696**
EEGNet-2.16	0.566	0.734	0.767	0.790	0.624	0.546	0.637	0.663	0.566	0.743	0.608	0.741	0.711	0.768	0.754	0.777	0.644	0.702	0.933	0.739	0.818	0.693	**0.706**
EEGNet-2.8	0.617	0.693	0.777	0.771	0.521	0.623	0.682	0.605	0.613	0.759	0.663	0.740	0.671	0.750	0.714	0.745	0.635	0.753	0.875	0.653	0.739	0.715	**0.696**
EEGNet-2.4	0.556	0.684	0.748	0.759	0.499	0.588	0.704	0.587	0.594	0.733	0.670	0.634	0.710	0.664	0.617	0.765	0.634	0.731	0.875	0.652	0.729	0.693	**0.674**
EEGNet-4.32	0.469	0.675	0.699	0.764	0.530	0.622	0.589	0.645	0.613	0.761	0.662	0.713	0.743	0.722	0.608	0.737	0.540	0.779	0.933	0.586	0.751	0.605	**0.670**
EEGNet-4.16	0.636	0.684	0.751	0.733	0.675	0.529	0.634	0.608	0.640	0.742	0.635	0.725	0.692	0.749	0.705	0.759	0.722	0.757	0.934	0.692	0.721	0.667	**0.700**
EEGNet-4.8	0.550	0.655	0.751	0.772	0.606	0.616	0.633	0.547	0.568	0.820	0.635	0.751	0.690	0.701	0.705	0.795	0.627	0.674	0.924	0.719	0.749	0.674	**0.689**
EEGNet-4.4	0.596	0.675	0.722	0.798	0.605	0.576	0.664	0.576	0.613	0.773	0.596	0.731	0.681	0.671	0.617	0.785	0.645	0.654	0.885	0.670	0.739	0.695	**0.680**

Table D.9: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 6 channel layout over motor cortex and frontal lobe from left side of the brain for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.463	0.700	0.703	0.821	0.587	0.531	0.722	0.577	0.576	0.740	0.645	0.606	0.598	0.694	0.674	0.798	0.585	0.771	0.962	0.817	0.741	0.571	**0.676**
EEGNet-2.16	0.415	0.731	0.721	0.820	0.580	0.528	0.701	0.550	0.587	0.692	0.681	0.636	0.558	0.695	0.635	0.799	0.645	0.809	0.923	0.787	0.705	0.571	**0.671**
EEGNet-2.8	0.508	0.730	0.750	0.866	0.564	0.546	0.710	0.555	0.588	0.683	0.731	0.608	0.637	0.694	0.703	0.786	0.518	0.721	0.944	0.787	0.803	0.548	**0.681**
EEGNet-2.4	0.511	0.730	0.749	0.790	0.608	0.535	0.740	0.517	0.608	0.674	0.662	0.590	0.616	0.674	0.724	0.796	0.579	0.631	0.933	0.756	0.762	0.566	**0.671**
EEGNet-4.32	0.500	0.615	0.662	0.829	0.590	0.644	0.721	0.552	0.459	0.739	0.685	0.508	0.683	0.710	0.711	0.740	0.597	0.769	0.943	0.777	0.753	0.518	**0.668**
EEGNet-4.16	0.473	0.661	0.739	0.885	0.588	0.501	0.761	0.473	0.642	0.663	0.713	0.646	0.578	0.751	0.692	0.800	0.694	0.838	0.924	0.787	0.810	0.544	**0.689**
EEGNet-4.8	0.535	0.681	0.671	0.811	0.548	0.510	0.788	0.527	0.559	0.682	0.684	0.600	0.637	0.781	0.759	0.815	0.635	0.753	0.934	0.756	0.811	0.615	**0.686**
EEGNet-4.4	0.470	0.737	0.691	0.790	0.638	0.557	0.740	0.595	0.607	0.701	0.605	0.619	0.665	0.722	0.695	0.825	0.706	0.645	0.933	0.779	0.781	0.662	**0.689**

Table D.10: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 3 channel layout over motor cortex for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.499	0.742	0.778	0.773	0.516	0.566	0.597	0.625	0.579	0.721	0.625	0.749	0.711	0.654	0.703	0.720	0.585	0.702	0.865	0.555	0.788	0.665	**0.669**
EEGNet-2.16	0.413	0.685	0.769	0.761	0.550	0.651	0.692	0.625	0.608	0.655	0.662	0.722	0.681	0.671	0.675	0.682	0.605	0.725	0.845	0.487	0.787	0.701	**0.666**
EEGNet-2.8	0.492	0.733	0.785	0.760	0.559	0.644	0.667	0.587	0.520	0.665	0.605	0.655	0.739	0.692	0.704	0.690	0.576	0.693	0.827	0.535	0.729	0.702	**0.662**
EEGNet-2.4	0.500	0.704	0.768	0.750	0.492	0.652	0.664	0.557	0.533	0.666	0.548	0.665	0.711	0.634	0.695	0.712	0.633	0.625	0.846	0.526	0.712	0.712	**0.650**
EEGNet-4.32	0.452	0.782	0.777	0.780	0.567	0.701	0.615	0.647	0.509	0.616	0.578	0.684	0.759	0.633	0.685	0.671	0.528	0.547	0.884	0.524	0.780	0.580	**0.650**
EEGNet-4.16	0.585	0.752	0.796	0.720	0.509	0.663	0.694	0.675	0.580	0.627	0.612	0.711	0.691	0.662	0.714	0.672	0.626	0.714	0.854	0.480	0.786	0.743	**0.676**
EEGNet-4.8	0.490	0.665	0.767	0.741	0.586	0.681	0.695	0.590	0.567	0.629	0.557	0.741	0.652	0.679	0.674	0.690	0.526	0.707	0.836	0.470	0.797	0.724	**0.657**
EEGNet-4.4	0.524	0.693	0.767	0.751	0.587	0.662	0.713	0.630	0.541	0.587	0.556	0.704	0.661	0.625	0.655	0.708	0.595	0.595	0.826	0.500	0.760	0.673	**0.650**

Table D.11: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 3 channel layout over frontal lobe for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.567	0.442	0.553	0.519	0.521	0.452	0.511	0.496	0.556	0.566	0.404	0.560	0.527	0.579	0.595	0.585	0.607	0.675	0.493	0.499	0.540	0.561	**0.537**
EEGNet-2.16	0.578	0.386	0.517	0.469	0.419	0.516	0.484	0.496	0.501	0.644	0.430	0.678	0.501	0.675	0.623	0.673	0.524	0.616	0.498	0.454	0.522	0.510	**0.532**
EEGNet-2.8	0.579	0.482	0.493	0.489	0.539	0.486	0.515	0.558	0.463	0.572	0.410	0.580	0.470	0.471	0.556	0.738	0.544	0.567	0.480	0.566	0.502	0.608	**0.530**
EEGNet-2.4	0.576	0.435	0.586	0.537	0.504	0.520	0.539	0.499	0.450	0.529	0.441	0.588	0.500	0.549	0.525	0.662	0.510	0.549	0.489	0.481	0.529	0.595	**0.527**
EEGNet-4.32	0.453	0.521	0.508	0.424	0.442	0.395	0.507	0.472	0.434	0.636	0.511	0.530	0.507	0.555	0.595	0.584	0.523	0.635	0.557	0.461	0.512	0.500	**0.512**
EEGNet-4.16	0.586	0.482	0.603	0.479	0.502	0.487	0.430	0.537	0.557	0.645	0.431	0.570	0.545	0.618	0.652	0.642	0.607	0.548	0.425	0.528	0.415	0.539	**0.538**
EEGNet-4.8	0.577	0.464	0.536	0.509	0.531	0.480	0.519	0.556	0.541	0.579	0.525	0.581	0.538	0.589	0.603	0.718	0.514	0.517	0.502	0.509	0.398	0.606	**0.541**
EEGNet-4.4	0.578	0.511	0.508	0.508	0.527	0.448	0.578	0.579	0.454	0.598	0.515	0.627	0.519	0.580	0.518	0.670	0.485	0.538	0.519	0.469	0.406	0.559	**0.532**

Table D.12: Classification accuracy obtained by different variations of EEGNet for 22 subjects using 3 channel layout over frontal lobe from left side of the brain for MI vs. rest discriminating task. Signals filtered in (8-30 Hz) frequency band, results averaged across 10 folds of cross-validation

Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	AVG
EEGNet-2.32	0.485	0.405	0.488	0.539	0.590	0.492	0.425	0.546	0.493	0.724	0.502	0.500	0.452	0.597	0.615	0.653	0.474	0.560	0.563	0.556	0.508	0.481	**0.529**
EEGNet-2.16	0.494	0.508	0.569	0.489	0.552	0.526	0.517	0.530	0.464	0.703	0.511	0.616	0.444	0.606	0.597	0.672	0.565	0.589	0.579	0.462	0.568	0.471	**0.547**
EEGNet-2.8	0.424	0.547	0.497	0.565	0.570	0.539	0.503	0.625	0.502	0.645	0.469	0.518	0.535	0.562	0.525	0.718	0.501	0.538	0.484	0.470	0.547	0.518	**0.536**
EEGNet-2.4	0.530	0.508	0.434	0.539	0.615	0.488	0.482	0.567	0.511	0.683	0.500	0.559	0.462	0.588	0.521	0.718	0.491	0.482	0.536	0.422	0.529	0.471	**0.529**
EEGNet-4.32	0.502	0.471	0.507	0.493	0.527	0.433	0.513	0.577	0.512	0.663	0.501	0.512	0.614	0.520	0.578	0.662	0.520	0.626	0.587	0.525	0.585	0.455	**0.540**
EEGNet-4.16	0.568	0.489	0.510	0.548	0.606	0.481	0.504	0.497	0.529	0.674	0.511	0.555	0.461	0.635	0.585	0.759	0.655	0.568	0.542	0.478	0.565	0.481	**0.555**
EEGNet-4.8	0.511	0.487	0.459	0.521	0.575	0.489	0.480	0.577	0.548	0.671	0.570	0.539	0.472	0.647	0.517	0.757	0.492	0.561	0.514	0.421	0.550	0.500	**0.539**
EEGNet-4.4	0.513	0.482	0.519	0.511	0.623	0.453	0.505	0.557	0.577	0.651	0.509	0.567	0.574	0.589	0.547	0.709	0.532	0.403	0.550	0.404	0.508	0.464	**0.534**