VRIJE
UNIVERSITEIT
BRUSSEL

Master thesis submitted in partial fulfilment of the requirements for the degree of
Master of Science in de Ingenieurswetenschappen: Computerwetenschappen

EVALUATION AND EXTENSION OF MOTOR IMAGERY DECODING METHODS

A scoping review of brain-computer interfaces and adding LSTM to CNN classification methods

Lennert Bontinck

2021 - 2022

Promotors: Prof. Dr. Geraint Wiggins & Prof. Dr. Kevin De Pauw

Advisor: Arnau Dillen

Faculty of Sciences and Bioengineering Sciences

VRIJE
UNIVERSITEIT
BRUSSEL

Proefschrift ingediend met het oog op het behalen van de graad van Master of Science in de Ingenieurswetenschappen: Computerwetenschappen

EVALUATIE EN UITBREIDING VAN DECODERINGS-METHODEN VOOR INGEBEELDE MOTORISCHE HANDELINGEN

Inleiding tot hersen-computerkoppelingen en de effecten van toegevoegde LSTM lagen in CNN-classificatiemethoden

Lennert Bontinck

2021 - 2022

Promotors: Prof. Dr. Geraint Wiggins & Prof. Dr. Kevin De Pauw

Copromotor : Arnau Dillen

Faculteit Wetenschappen en Bio-ingenieurswetenschappen

Abstract

Brain-computer interfaces (BCIs) are systems that aim to translate brain signals into actions. A recent rise in both scientific and commercial interest in these systems has helped popularize them outside highly specialized labs. However, the highly interdisciplinary nature of BCI research among many other complex factors results in a steep learning curve which deters many individuals who wish to work with these kinds of systems.

This master thesis aims to facilitate this entry into the BCI field for computer scientists. Through an extensive but intuitive introduction of BCIs, a scoping review is performed to address recent advancements, current opportunities and open issues in the field together with some ethical concerns. This intuitive understanding is then grounded by providing the required theoretical and technical background on the biomedical signals (biosignals), brain signal acquisition systems and the general motor imagery (MI) electroencephalography (EEG) classification pipeline that enables these systems.

This theory is put to practice by discussing, implementing and evaluating seven unique offline MI EEG classification pipelines. In particular, two traditional two-step machine learning (ML) approaches based on common spatial pattern (CSP) feature extraction, three literature proposed state-of-the-art convolutional neural network (CNN) based approaches and two proposed extensions providing long short-term memory network (LSTM) functionalities to such CNN models are considered. The latter extensions were proposed to explore the question if added LSTM functionalities helps the state-of-the-art EEGNet model in optimally decoding MI EEG data. It was found that for 1.5-second windowed samples centred around a known event, one of the LSTM extension outperforms the CNN based state-of-the-art EEGNet model it is based on in an intersubject evaluation strategy for testing three-class MI EEG classification performance. This model obtained a mean test accuracy of 65.52 ± 0.89, over the 60.68 ± 1.64 obtained by the EEGNet model it is based on. Other confusion matrix (CM) derived metrics were also found favourable for the extension. It is discussed how this difference could increase further when using a different windowing technique.

Due to the considerable amount of information this master thesis wishes to provide, a multitude of new questions and potential future work arises. Some of these discussed future possibilities, such as moving from the offline MI EEG classification pipelines to a complete BCI system, benefit from pilot studies also presented in this master thesis. In this regard, a pilot study found that an intersubject test accuracy of over 90% can be reached for certain subjects of the open-source dataset used.

Acknowledgements

I'm extremely grateful to both my supervisor Professor Doctor Geraint Wiggins and my advisor Arnau Dillen for allowing me to write this thesis on a subject that is tailored to my interests. This thesis would not have been possible without their helpful advice, constructive criticism and insightful suggestions. I would like to thank them especially for bearing with me through a master thesis process that was filled with ups and downs. In this regard, I must also thank my girlfriend, parents and friends for their unparalleled support throughout this sometimes stressful process. I must also thank Professor Doctor Jef Vandemeulebroucke for teaching both the biomedical signals and imaging course as well as the clinical decision support systems course. The contents of these courses have contributed to a better understanding of some of the physical phenomena and used technologies addressed in this master thesis. I'd also like to recognize the statistical and writing expertise gained through the methods of scientific research course by Professor Doctor Bart de Boer.

Contents

I Understanding brain-computer interfaces **9**

1 Brain-computer interfaces **11**
1.1 Introduction . . . 11
1.2 Growing scientific and commercial interest in BCIs . . . 12
1.2.1 BCIs have gained big-tech interest and funds . . . 12
1.2.2 Continued improvement of brain-signal measuring facilities . . . 16
1.2.3 Availability of more powerful, affordable and portable equipment . . . 21
1.2.4 New and refined data processing techniques . . . 23
1.2.5 Summarizing the cycle of increasing popularity . . . 29
1.3 Common use cases for BCIs . . . 30
1.3.1 Using BCIs as an automated tool in all stages of medical phenomena . . . 31
1.3.2 Using BCIs for prosthesis and exoskeleton control . . . 34
1.3.3 Using BCIs to improve the working of existing systems . . . 36
1.3.4 Using BCIs as an alternative for eye tracking . . . 39
1.3.5 Summary on the use cases for BCIs . . . 42
1.4 Opportunities and obstacles for BCI research . . . 44
1.4.1 Seemingly small projects with huge impact . . . 44
1.4.2 Motivational aspects for this master thesis . . . 45
1.4.3 The potential of an AutoML variant for BCI pipelines . . . 52
1.4.4 A lack of standardized testing and reporting . . . 54
1.4.5 Challenges from the highly interdisciplinary nature of BCI systems . . . 56
1.4.6 Brain signals are a complex data source . . . 58
1.5 Ethical challenges for BCIs . . . 59
1.5.1 Painfully confronting users with their brain . . . 59
1.5.2 E-waste inside your skull . . . 60
1.5.3 Changing peoples personal identities . . . 60
1.5.4 A great risk of sloppy science . . . 61
1.5.5 Advertisements based on your thoughts . . . 62
1.5.6 Hacking BCI systems . . . 62
1.6 Chapter conclusions . . . 63

2 Origins and acquisition of biomedical signals **65**
2.1 Introduction . . . 65
2.2 Biosignals in the human body . . . 65
2.2.1 Origin of electrical activity inside the human body . . . 66
2.2.2 Other energy forms and their related biosignals . . . 68
2.3 Measuring biosignals from the brain . . . 69

2.3.1 Common measuring modalities in BCI systems . . . 69
2.3.2 Non-invasive EEG for BCIs . . . 73
2.3.3 Available EEG measuring equipment . . . 73
2.4 Working with EEG . . . 76
2.4.1 Standards and guidelines for EEG measurements . . . 77
2.4.2 Using the anatomy of the brain for electrode subsampling . . . 78
2.4.3 The five widely recognized brainwaves . . . 80
2.4.4 Common methods for inducing electrical brain activity . . . 81
2.4.5 Generalisation issues of brain activity . . . 82
2.4.6 Common EEG artefacts . . . 83
2.5 Chapter conclusions . . . 84

3 Decoding brain signals with machine learning 87
3.1 Introduction . . . 87
3.2 A general pipeline for classifying brain signals . . . 87
3.2.1 Data source . . . 89
3.2.2 Windowing . . . 89
3.2.3 Preprocessing . . . 90
3.2.4 Feature engineering . . . 93
3.2.5 Classification . . . 97
3.3 Machine learning and deep learning for classification . . . 98
3.3.1 Machine learning paradigms . . . 98
3.3.2 Traditional two-step ML vs one-step DL classifiers . . . 100
3.3.3 Common traditional two-step ML classifiers . . . 100
3.3.4 Common one-step DL classifier layers . . . 104
3.4 Evaluating the performance of the pipeline . . . 108
3.5 Chapter conclusions . . . 113

II Development of motor imagery EEG classifiers 115

4 EEG-based offline classification system for motor imagery tasks 117
4.1 Introduction . . . 117
4.2 CSP-based two-step ML approaches . . . 118
4.2.1 The idea behind common spatial pattern(s) (CSP) . . . 118
4.2.2 Traditional CSP with LDA, SVM and RF . . . 119
4.2.3 The issue with a traditional CSP approach . . . 121
4.2.4 Improving traditional CSP with FBCSP . . . 122
4.3 Convolutional-based one-step DL approaches . . . 123
4.3.1 EEGNet . . . 123
4.3.2 DeepConvNet . . . 125
4.3.3 ShallowConvNet . . . 126
4.3.4 Interpreting these black box models . . . 127
4.4 Adding memory to one-step DL approaches . . . 127
4.4.1 Adding an additional LSTM layer to EEGNet . . . 128
4.4.2 EEGNet with convolutional LSTM layer . . . 129
4.5 Chapter conclusions . . . 129

5 Moving towards an online BCI system **131**
5.1 Introduction . . . 131
5.2 BCI system try to solve another problem . . . 131
5.2.1 Using classification vs regression in a BCI system . . . 131
5.2.2 Complexer data and less powerful systems . . . 132
5.2.3 Limiting risky errors . . . 133
5.3 Common steps towards an online BCI system . . . 133
5.3.1 Working with continuous data . . . 133
5.3.2 Improving classification performance and adding more classes . . . 133
5.3.3 Calibrating the classification pipeline before use . . . 134
5.3.4 Spatial subsampling . . . 134
5.3.5 Configuring the high-level controller . . . 134
5.4 Chapter conclusions . . . 135

III Reflection on the results of this master thesis **137**

6 Evaluation of the proposed pipelines **139**
6.1 Introduction . . . 139
6.2 Three class MI EEG data source . . . 139
6.3 General remarks on experiment setups . . . 140
6.4 Considered evaluation metrics . . . 141
6.5 Intrasession evaluation . . . 142
6.6 Intersession evaluation . . . 144
6.7 Intersubject evaluation . . . 146
6.8 Longer window lengths . . . 146
6.9 Additional pilot studies . . . 148
6.9.1 Improving intersubject performance by providing more data . . . 149
6.9.2 Comparison of prediction times . . . 150
6.9.3 Pilot studies with less promising results . . . 150
6.10 Chapter conclusions . . . 151

7 Conclusions, future work and self-reflection **153**
7.1 Introduction . . . 153
7.2 Are brain-computer interfaces made easy? . . . 153
7.3 Is providing LSTM functionalities to a CNN model beneficial? . . . 154
7.4 Future work . . . 155
7.5 Final remarks and a personal endnote . . . 156

Extra figures **157**

References **161**

List of Figures

1.1 Funding of newer BCI related companies depicted in millions (USD). Figure after the results of research by Rao (2020). It is noted this data is limited to companies that were created after 2010 where funding information is made available. 16
1.2 The contrast between medical-grade EEG measuring equipment and a consumer-grade alternative. 18
1.3 The anatomy of the human head, specifically of the cerebrospinal system. Non-invasive EEG measuring equipment is placed on the scalp, causing signals from the brain to be blocked by the skull and cerebrospinal fluid (CSF) among other structures. Free to use figure by Blaus (2014), CC BY 3.0, via Wikimedia Commons. 19
1.4 The number of BCI-related papers over time. Figure after the results of research by Saha et al. (2021) who searched PubMed using the keyword "brain-computer interface". 24
1.5 General pipelines of a computer-aided diagnosis (CADx) system used for classification. A ML approach is called a two-step approach as it exists from feature extraction and classification. A deep learning (DL) approach is called a one-step approach as it combines both of these steps into a singular classification algorithm. 26
1.6 Summary of some of the potential reasons BCI research has seen its gain in popularity discussed in section 1.2. There is no clear chronological order for these events and they likely influence each other all at the same time. 30
1.7 An overview of some of the many use cases for BCIs. 43
1.8 The contrast between mappings of MI EEG for controlling a robotic arm to grasp an item. One approach allows for full control but requires four MI classes and many steps. One approach offers less control but succeeds in the task with three MI classes and very few steps. 53

2.1 Chart of the membrane potential during the action potential process in a neuron. Free to use figure by Synaptidude, GFDL 1.2, via Wikimedia Commons. 68
2.2 Illustration of the three most common montages for the differential amplifier in EEG. The annotations are the electrode names used in the international 10-20 system. 71
2.3 The original international 10-20 system for EEG electrode placement and its modified combinatorial nomenclature (the 10-10 system) which supports more electrodes. 78
2.4 Illustration of how the placement of electrodes in the 10-20 system is determined based on the four identifiable points of the skull around the nasion, inion and ears. Black crosses are used for identification of the blue interval only and no electrodes are placed here. Crosses with a coloured background are used for determining two intervals, the cross colour and background colour intervals. 79

2.5 Sideview of the brain with motor and sensory regions highlighted. The left side corresponds to the back of the head, with the visual cortex being just above the back of the human neck. Free to use figure by Blaus (2014), CC BY 3.0, via Wikimedia Commons. 80
2.6 Overview of the most common methods for inducing electrical brain activity and the signals they produce. 82

3.1 General training pipeline and prediction pipeline used for classifying brain signals. 88
3.2 Visualisation of the EEG recording from December 15 2015 overlayed with the provided markers for subject B in the open-source dataset by Kaya et al. (2018). The x-axis depicts the time in seconds, the y-axis depicts which channel of the recording is visualised and the colour overlay represents the active marker. 90
3.3 Different types of windowing techniques. 91
3.4 Different types of filtering techniques. 92
3.5 Power spectral density (PSD) for the EEG signal previously shown in Figure 3.2. Only the C3, Cz and C4 channels are shown. 95
3.6 A linearly separable problem with multiple possibilities for decision boundaries. . 102
3.7 An intuitive example of SVM's kernel trick where a 2D non-linearly separable problem is transformed into a 3D problem where it is linearly separable. Free to use Figure by Shiyu Ji, CC BY-SA 4.0, via Wikimedia Commons. 103
3.8 An example of a three-layer multilayer perceptron (MLP) with a multidimensional input layer and a one-dimensional output layer. Dots denote neurons and lines denote edges which have an associated weight that can be learned. 105
3.9 Visualisation of a convolutional layer. 107
3.10 Design of a binary and three-class confusion matrix. 110
3.11 ROC curve with some arbitrary examples. Free to use figure by cmglee, MartinThoma, CC BY-SA 4.0, via Wikimedia Commons. 112

4.1 Visualisation of the spatial transformation performed by CSP. 119
4.2 Visual overview of the pipeline used for MI EEG classification using the standard CSP method and multiple traditional two-step ML classifiers. 120
4.3 Visual overview of a proposed filter bank common spatial pattern (FBCSP) pipeline used for MI EEG using multiple traditional two-step ML classifiers. 122
4.4 Visual overview of a pipeline using a DL classifier for MI EEG classification. . . . 123
4.5 A conceptual overview of EEGNet after the overview by Lawhern et al. (2016). . 124
4.6 All layers of the EEGNet model. 125
4.7 All layers of the DeepConvNet model. 126
4.8 All layers of the ShallowConvNet model. 126
4.9 topographic maps (topomaps) created from multiple CSP components using subsampled channels from the international 10-20 system. 127
4.10 All layers of the CNN-LSTM model using all EEGNet layers for feature extraction and the addition of LSTM functionality for time series processing. 128
4.11 All layers of the CNN-LSTM model using a reduced version of EEGNet for feature extraction and the addition of convolutional LSTM functionality for time series processing. 129

5.1 Overview of a complete BCI pipeline. 132

6.1 Evolution of both validation accuracy and validation loss over time. Graphs shown are from the intrasession experiment of EEGNet for subjects B (dark blue), C (light blue) and E (green). The time axis denotes the epoch in which the result is obtained. 142
6.2 Intrasession results for all classification approaches. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches. 143
6.3 Intersession results for all classification approaches. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches. 145
6.4 Intersubject results for all classification approaches. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches. 147
6.5 Mean test accuracies and their standard deviation from the longer window experiments in an intersubject evaluation setting. 148
6.6 Intersubject results for some of the one-step DL classification approaches using more training data. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches. 149
6.7 Average prediction time in seconds for one 0.5 seconds long window with its standard deviation. 150

7.1 Intrasession results for all classification approaches using the longer window. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches. . 157
7.2 Intersession results for all classification approaches using the longer window. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches. . 158
7.3 Intersubject results for all classification approaches using the longer window. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches. . 159

List of Tables

2.1 Some of the different energy forms in living beings and the most important measurable biosignals they produce. Data from Semmlow (2018). 66
2.2 Overview of the most common neuroimaging modalities for use in BCI systems. *: Currently not available in a portable manner. Data based on Martini et al. (2020) and Ramadan and Vasilakos (2017). 70
2.3 Some of the available EEG measuring equipment with their most important specifications. Some equipment consists of multiple configurable components, these are annotated as "variable" with the default option provided in brackets. Prices are estimates based on buying a minimal total configuration of the product. Data was gathered from the original manufacturer's website and community blogs. More information on the provided properties is given in Section 2.3.3. 75
2.4 Overview of the five widely recognized brainwaves according to data from Priyanka et al. (2016), with further subdivision in low and high brainwaves by data from Kawala-Sterniuk et al. (2021). 81

List of abbreviations

Symbols

μV microvolts.

A

AC alternating current.

AEP auditory evoked potential.

AF atrial fibrillation.

AI artificial intelligence.

ALS amyotrophic lateral sclerosis.

ANN artificial neural network.

ApEn approximate entropy.

AR augmented reality.

AUC area under the (ROC) curve.

AutoML automated machine learning.

B

BAIID breath alcohol ignition interlock device.

BBCI Berlin Brain-Computer Interface research program.

BCI brain-computer interface.

bioelectrical signals electrical biomedical signals.

biosignal biomedical signal.

biosignal control biological signal control.

BMI brain-machine interface.

BPNN back propagation neural network.

BT Bluetooth.

C

CADe computer-aided detection.

CADx computer-aided diagnosis.

CM confusion matrix.

CNN convolutional neural network.

CNS central nervous system.

CPU central processing unit.

CSF cerebrospinal fluid.

CSP common spatial pattern.

CT computerised tomography.

CV cross-validation.

D

DEEG digital electroencephalography.

DFT discrete Fourier transform.

DL deep learning.

DMD duchenne muscular dystrophy.

DSP digitial signal processor.

E

ECG electrocardiography.

ECoG electrocorticography.

EDA electrodermal activity.

EEG electroencephalography.

EGG electrogastrogram.

ELU exponential linear unit.

EMG electromyography.

EOG electrooculography.

EP evoked potential.

ERD event-related desynchronization.

ERG electroretinography.

ERN error-related negativity.

ERP event-related potential.

ERS event-related synchronization.

ESI scout EEG source imaging.

EU European Union.

F

FBCSP filter bank common spatial pattern.

FDA Food and Drug Administration.

FF feed-forward.

FFNN feed-forward neural networks.

FFT fast Fourier transform.

FIF functional image file format.

FIR finite impulse response.

FN false negative.

fNIRS functional near-infrared spectroscopy.

FP false positive.

FPGA field programmable gate arrays.

FT Fourier transform.

fTCD functional transcranial doppler.

FuzzEn fuzzy entropy.

G

GAN generative adversarial network.

GDPR general data protection regulation.

GPU graphics processing unit.

GSR galvanic skin response.

H

HMI human-machine interface.

HU Hounsfield unit.

Hz hertz.

I

ICA independent component analysis.

IDFT inverse discrete Fourier transform.

IFT inverse Fourier transform.

IIR infinite impulse response.

IWBW intensity weighted bandwidth.

IWMF intensity weighted mean frequency.

L

latex Is a markup language specially suited for scientific documents.

LDA linear discriminant analysis.

LIS locked-in syndrome.

LR logistic regression.

LSTM long short-term memory network.

M

MAV mean absolute value.

MEG magnetoencephalography.

MI motor imagery.

ML machine learning.

MLP multilayer perceptron.

MRCP movement-related cortical potential.

MRI magnetic resonance imaging.

mv millivolts.

N

NLP natural language processing.

NN neural network.

NPV negative predictive values.

P

PCA principal component analysis.

PE permutation entropy.

PET positron emission tomography.

PFBCSP penalized frequency band common spatial pattern.

POC proof of concept.

PPV positive predictive values.

PSD power spectral density.

PTFBCSP penalized time-frequency band common spatial pattern.

R

RBF radial basis function.

ReLU rectified linear unit.

RF random forest.

RL reinforcement learning.

RNN recurrent neural network.

ROC curve receiver operating characteristic curve.

S

sklearn scikit-learn.

SNR signal-to-noise ratio.

SPECT single-photon emission computerized tomography.

SSAEP steady-state auditory evoked potential.

SSC slope sign changes.

SSEP steady state evoked potential.

SSEP somatosensory evoked potential.

SSSEP steady-state somatosensory evoked potential.

SSVEP steady-state visual evoked potential.

STFT short-time Fourier transform.

SVC C-support vector classification.

SVM support vector machines.

T

TL transfer learning.

topomap topographic map.

ToS terms of service.

U

UI user interface.

UX user experience.

V

VEP visual evoked potential.

VR virtual reality.

W

WL waveform length.

WT wavelet transformation.

Z

ZC zero crossing.

Part I

Understanding brain-computer interfaces

Chapter 1

Brain-computer interfaces

1.1 Introduction

Brain-computer interfaces (BCIs) are systems, consisting of hardware and software, that aim to read or even stimulate a user's brain signals for a wide variety of applications. Whilst many of these applications for BCIs revolve around providing novel interaction methods for computer applications, they are capable of fulfilling more general tasks as well. Because of this, BCIs are also referred to as brain-machine interfaces (BMIs) and can be seen as a special type of the more general human-machine interfaces (HMIs) and biological signal control (biosignal control) systems. A well-known Professor in this field, Jonathan R. Wolpaw, was part of the group that provided the following first formal definition for BCIs:

> A brain-computer interface is a communication system that does not depend on the brain's normal output pathways of peripheral nerves and muscles.
>
> *J. Wolpaw et al. (2000)*

Considering Wolpaw's many contributions to the field, his opinion on what defines a *perfect medical BCI* is often strived for (Daly & Wolpaw, 2008; Shih et al., 2012; J. Wolpaw & Wolpaw, 2012; J. Wolpaw et al., 2000). Whilst being focused on medical applications, these properties would also be beneficial for commercial BCI systems. For these reasons, this first chapter relates many of the discussed topics to the definition below quite a lot. It is noted that the term communication in these definitions simply depicts the exchange of information rather than specific human communication such as speech.

> The perfect [medical] BCI is a safe and affordable system which works all the time, does not require the permanent assistance of a technician or a scientist, restores communication at "normal" speed, is aesthetically acceptable, is reliable and, for the same function, does not require more concentration for a patient user than what it does for an able-bodied person.
>
> *Peterson et al. (2020) and J. Wolpaw and Wolpaw (2012)*

This first chapter aims to provide all the required background information needed for understanding BCIs on an intuitive level. It is written in such a way that it is accessible to new people in the field that are otherwise confronted with the steep learning curve related to the BCI field. For those that already have experience in the field or are looking for a more technical

discussion of the field are invited to read diagonally through this first chapter, focusing primarly on the summaries provided in Section 1.2.5 and 1.3.5, Section 1.4 in its entirety and the chapter conclusions under Section 1.6.

For those interested in this chapter in its entirety, this chapter further introduces the main rationale behind BCIs research by discussing the rise in popularity of both medical and commercial BCIs, some practical examples of BCI systems that have been developed and some of the opportunities and obstacles in the field. The chapter ends with a note on some of the ethical challenges for BCI systems and a discussion of the proposed system for this thesis. The interested reader is also referred to the great introductory book on BCIs by J. Wolpaw and Wolpaw (2012) and the review article by Nicolas-Alonso and Gomez-Gil (2012) when more introductory insight is desired. Whilst these resources have dated a little and state-of-the-art has changed since then, the main ideas discussed in them remain unchanged. A more recent, systematic review article by Dillen et al. (2022) focuses on current deep learning (DL) techniques for use with BCIs among other biosignal control systems.

1.2 Growing scientific and commercial interest in BCIs

With brain signal measuring modalities such as electroencephalography (EEG) being over 100 years old, the idea of using those brain signals for a wide variety of use cases has been explored for many decades (Berger, 1929; Haas, 2003; Kübler, 2020). With feasibility studies of using BCIs already existing in the 1970s (for example by Vidal, 1973) showing that most of the ideas explored today are not new, a clear spike in both scientific and commercial interest can be seen after the 2000s. It is no coincidence that the first international meeting devoted to BCI research and development as part of the IEEE conference on Rehabilitation Engineering discussed in Section 1.1 was also from this period.

This rise in popularity can be explained by several events. Perhaps most importantly is the improvement of both brain signal measuring equipment and computational processing equipment in both efficiency, accuracy and portability. Recent improvements in machine learning (ML) and DL after some artificial intelligence (AI) winters between the 1970s and the 1990s are bound to also have played an important role. The interest of big tech companies such as Neuralink, Meta and Valve have also introduced unseen amounts of funds further accelerating BCI research.

This section focuses on discussing these most important contributing factors to the new rise of interest in BCI research. These factors are discussed in an arbitrary order, as most of them have influenced each other and it is hard to name a singular reason that explains this recent rise in interest. For a more in-depth overview of the rich history that BCI research has, the reader is referred to the work by Kübler (2020).

1.2.1 BCIs have gained big-tech interest and funds

Big tech has been catching on with the possibilities BCIs bring, and the amount of money they can earn from it. Although profitability is an important factor in most medical applications as well, the focus of medical applications lies on improving the life of a patient, whilst the focus of commercial applications can differ greatly. Since commercial BCIs are still in their early stages and the idea of constantly wearing a brain-signal recording headset has not been accepted by the wide public yet, many commercially oriented companies start with products that are a cross between medical and commercial applications.

Most noteworthy of these more commercially oriented companies is Neuralink, an Elon Musk company. Neuralink's initial white paper discusses its aim to create a scalable high-bandwidth

BCI system, focusing on its mechanical achievements (Musk & Neuralink, 2019). These mechanical achievements are rather impressive, with state-of-the-art robot surgery inserting ultra-thin sensors directly into the skull allowing for a sleek and visually pleasing package that is mostly hidden from the human eye. Comparing this to non-invasive methods of recording brain signals, which are methods that don't require inserting machinery into the human body, the signal quality is also expected to be far greater. However, an invasive approach currently introduces added health risks and more ethical challenges making non-invasive methods often more suited for general use (Dadia & Greenbaum, 2019; Dillen et al., 2022; Jawad, 2020). Since the publication of the Neuralink white paper, the company has held live demos of their BCI implanted directly into the skulls of animals such as pigs and monkeys. A video by Neuralink of a monkey playing pong using brain signals as input[1] has gathered over 6 million views on YouTube already. Combined with many news articles, the kind of exposure that Neuralink has gotten is unseen compared to the regular exposure of literature in the field. This can be questioned, as earlier work by Ifft et al. (2013) demonstrated monkeys taking control over two avatar arms simultaneously, a task that is arguably even harder to accomplish than simply playing pong. Adding to this, the experiment by Ifft et al. (2013) has an appropriate peer-reviewed paper backing it whilst Neuralink among other commercially oriented companies in the field often lack scientific backing for the claims they make. Thus, the scientific value of these more commercial demos and applications can be argued for, but the funds for research introduced by these companies and the exposure to the field have accelerated research in the field and helped popularize the field. Adding to this, the proposed system by Neuralink is one of the most aesthetically pleasing compared to alternative invasive or non-invasive systems on the market, which is one of the properties of Wolpaw's perfect BCI system given in Section 1.1.

Besides Neuralink, companies like Meta, Valve, Neurable, InteraXon and many more are exploring the commercial possibilities of BCIs as well. Some of the companies do this through direct internal research whilst others might provide funds for external projects (Alcaide et al., 2021; Cuthbertson, 2021; Moses et al., 2021; Stockman, 2020). There seem to be two main focuses of the technology in the commercial space. Either using the new interaction method to perform work more efficiently or using it for recreational purposes.

Using BCIs to boost work efficiency

Meta, formerly known as Facebook, has been playing with the idea of BCIs for quite a while but has been relatively quiet about it publicly. In 2021, Meta publicly announced it had provided funds for research on the use of a BCI-system to restore speech functionalities for people suffering from anarthria (Facebook, 2021; Moses et al., 2021). The system by Moses et al. (2021) achieved an average of 15 words per minute, decoded with a median error of 25%. Whilst this might not sound impressive, anarthria is a disease which causes patients to not be able to articulate speech at all due to lost control of the muscles required for making sounds. Adding to this, people suffering from anarthria often suffer from other lost muscle control as well, making alternatives such as keyboard typing or writing impossible. Taking this into account, these results should be seen as very impressive and such a system can be life-changing for certain patients. Whilst the system by Moses et al. (2021) was invasive just like Neuralink's system, it was far from visually pleasing. The patient was fixed in a chair and physically connected to a bulky processing unit in the form of a small server rack, which makes the system non-mobile and makes the user stand out if it were to be used in the real world where discreteness is often desired.

The system by Moses et al. (2021) is an example of one that is backed by the funds that big-tech companies have and which is mainly focused on medical applications whilst the final

[1] https://youtu.be/rsCul1sp4hQ

intention of the funding company is most likely of commercial nature. Indeed, it is not hard to imagine the commercial interest of Meta in developing a more general *virtual keyboard* to enable fast *thought to speech* or *thought to text* applications usable by the masses. In fact, during the F8 conference (Facebook, 2017) a couple of years before the paper by Moses et al. (2021), Meta stated the following:

> Specifically, we have a goal of creating a silent speech system capable of typing 100 words per minute straight from your brain - that's five times faster than you can type on a smartphone today.
>
> *Facebook (2017)*

Such a virtual keyboard could replace certain speech-to-text applications already broadly used for commercial purposes. In the same blog post by Facebook (2021) discussing the funding for the project by Moses et al. (2021), it is also mentioned that Meta has interest in using BCIs for high-bandwidth interactions in AR/VR. However, Meta has been subject to multiple privacy concerns lately (Fuller, 2019; Hu, 2020). The company's reputation has been damaged from this which doesn't help in selling the concept of them having a BCI which allows them to read the brain activity of the users. This could explain why they have recently started to shift their focus from BCIs towards muscle-based interfaces using electromyography (EMG) (Facebook, 2021).

A recent example of a BCI being used to boost work efficiency is covered by S. Chen (2022). He discusses how Chinese researchers have been working on a non-invasive and portable system that aims to detect if a user is watching pornographic content through brain signals. When presenting fifteen male participants aged between 20 and 25 with erotic content and regular content, an accuracy of over 80% was obtained for determining whether the user was watching erotic content or not. Such a system should aid in China's content regularisation which often bans such erotic content on their domestic social-media platforms. Current systems rely on manual evaluation by a reviewer for removing or keeping content that is flagged as inappropriate by either the community or an algorithm. Further automating this task through a brain-controlled system could boost the efficiency of this process significantly.

Using BCIs for recreational use

Whilst some promising results have been obtained when using BCIs in commercial settings to boost work efficiency, many of the systems still lack the desired performance to become truly viable. And thus, the most prominent type of commercial BCIs are those focusing on recreational use. In this regard, the headsets by InteraXon, produced under the Muse brand, are one of the earliest examples, with their first version being released in 2014. The first iteration of this product was advertised as a meditation aid. This headset relies on measuring Theta waves in the brain, which are lower frequency waves that suggest a user is meditating. Section 2.4.3 talks about these brainwave frequencies in more detail. The actual accuracy and usefulness of these types of systems are debated, as discussed by Stockman (2020). More recently, a newer version of the InteraXon headset came to market, named Muse 2, which also aids in sleep monitoring. This relies on detecting Delta waves among other patterns to determine the sleep quality of a user. Like before, this accuracy and usefulness can be argued for. A similar commercial product for sleep tracking is available from the company Dreem, under the name Dreem 2. Dreem has received more funding, as can be seen in Figure 1.1, but InteraXon, the company behind the Muse headset, has arguably contributed more to the field. Not only was it one of the first commercial BCIs that gained media attention, but the company also plays an important role in the commercialisation of BCIs as their headsets are cheap, non-invasive and visually pleasing whilst also being widely

available. Adding to this, these headsets have pretty good supporting libraries in Python among other programming languages that allow developers to use these headsets for other purposes as well. Besides InteraXon, some other companies that specialize in providing commercially usable brain-signal recording headsets exist, as will be further discussed in Section 2.3.3.

Perhaps the most promising short-term commercial use of BCIs is in combination with virtual reality (VR) and augmented reality (AR). Besides Meta's interest in this region, as discussed earlier in this section, Valve has also said it is actively researching how to use BCIs as a novel interaction system in VR games (Cuthbertson, 2021). Valve is the company behind Steam, one of the world's largest game marketplaces and they are specialized in creating games and gaming hardware as well. To achieve the goals of this project code-named *Galea*, Valve is working together with OpenBCI, a well-respected company in the BCI research field that has provided open-source hardware and software for use in BCI systems. Tobii, a company that specializes in eye-tracking software, is also working on project Galea. With a final goal of creating an open-source BCI that can be used in gaming, the anticipation for the headset has been high. However, just as with deadlines from other companies such as Meta and Neuralink, the project has been postponed multiple times. This is not surprising as the promises of what a BCI can do are near endless and initial trials often offer promising results but going to a final product has been proven to be incredibly hard due to several open issues (Dillen et al., 2022).

A final stream of money that is important to mention, is coming from militaries around the globe. The U.S. military among others is known to invest a lot of money in any form of innovation, especially related to devices and technologies that can give them a strategic edge when fighting in a war (with examples of funds for AI research in Center for Security and Emerging Technology et al., 2020; Hunter Christie, 2022). Whilst most of this information is classified, it is known that the U.S. Department of Defense and others have shown interest in a wide variety of applications using BCIs (Binnendijk et al., 2020). Whilst one can only guess what these government organisations are developing, it is likely that over time these applications might become public knowledge and aid the research field of BCIs in creating even better systems. Tullis (2019) discusses the U.S. militaries interest in BCIs further.

Summary of big-tech using BCIs for commercial applications

To summarize, there have been a lot of big companies showing interest in commercial BCI applications in the past few years. Some might contribute directly to the field by funding scientific research, which is often still focused on medical applications but whose results can show potential for certain commercial applications (Facebook, 2021; Moses et al., 2021; Musk & Neuralink, 2019). On the other hand, some companies are working on commercial products internally, mostly for improving work efficiency (S. Chen, 2022; Facebook, 2017) or for recreational use (Muse and InteraXon headband, Cuthbertson, 2021). These commercial products have yet to see truly successful examples, as they are either questionable in delivering what they promise (Stockman, 2020), experience delayed deadlines or are even cancelled in their entirety. Nonetheless, these companies focusing on commercial applications often have high amounts of funding and a focus on the user experience (UX) of BCIs which could help accelerate research in the field and make BCI systems more visually pleasing and accepted by the broader public. In this way, they also contribute to Wolpaw's vision of a perfect BCI system as discussed in Section 1.1.

Figure 1.1 shows the funding of BCI-related companies founded after 2010 as a rough indication of how much money is spent on start-up companies in the field. Interestingly, from the companies Neurable, Muse and Neuralink mentioned in this thesis, the funding amount is in proportion to the overall popularity of that company to the wider public, although this is by no means a proven relation. It also shows that whilst academic research on BCIs doesn't require

huge funding, with open-source datasets and relatively cheap hardware available as is further discussed in Section 1.2.4, 2.3.3, and 3.2.1, creating an effective commercial product can become an expensive affaire rather quickly.

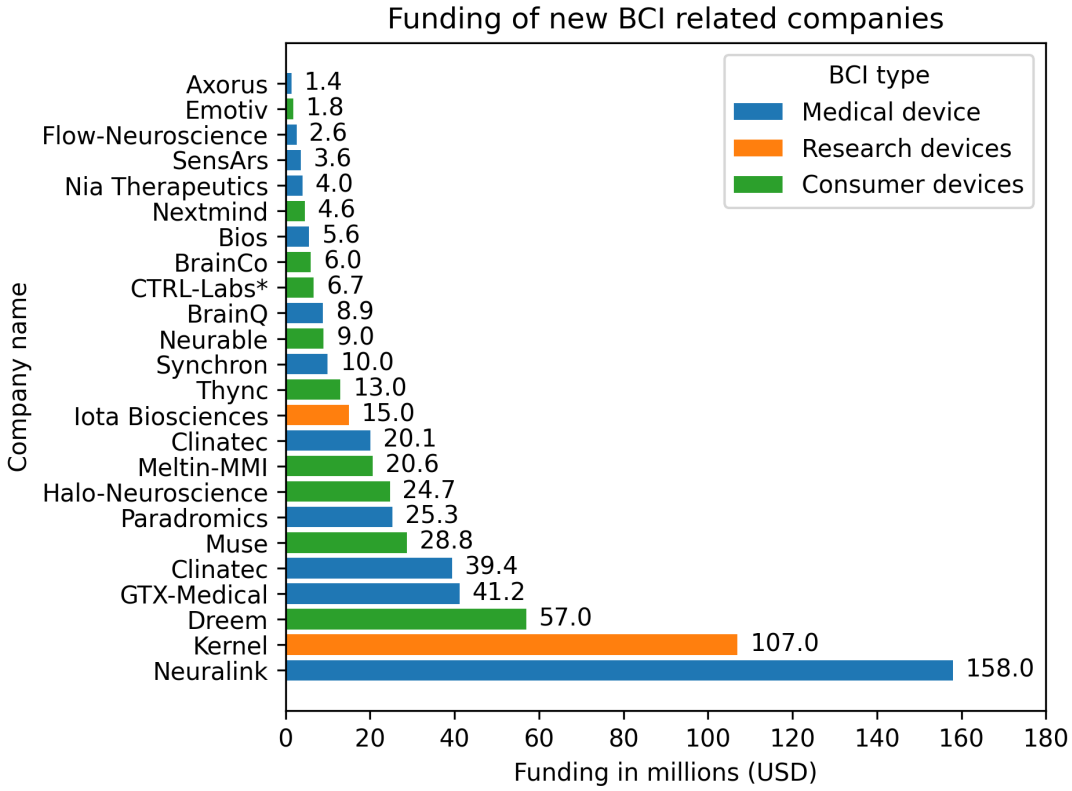


Figure 1.1: Funding of newer BCI related companies depicted in millions (USD). Figure after the results of research by Rao (2020). It is noted this data is limited to companies that were created after 2010 where funding information is made available.

1.2.2 Continued improvement of brain-signal measuring facilities

As brain-computer interfaces (BCIs) are a type of human-machine interface (HMI) relying solely on brain signals to operate, the measuring facilities for acquiring data of those brain signals have a direct impact on the capability of those systems.

Most BCIs rely on non-invasive measuring equipment that uses EEG as a source of data and this paper will focus mainly on such measuring equipment as well. Chapter 2 explains in greater detail what EEG and some of its alternatives are, the equipment used for acquiring brain-signal data and more. For this introduction, it suffices to know that non-invasive EEG measuring equipment measures the electrical potential difference, often in microvolts (μV), between electrodes placed on the scalp.

Following Wolpaw's definition for a perfect BCI given in Section 1.1, the recording hardware should ideally be aesthetically acceptable and shouldn't require the assistance of a professional

to install. In recent years, new developments in this hardware have made meeting these criteria more plausible, which are adressed in this section.

Hardware improvements in non-invasive EEG measuring equipment

Three major hardware distinction made between the electrodes used in non-invasive EEG measuring equipment is whether they are wet or dry electrodes, whether they are active or passive electrodes and whether communication to the processing unit happens wirelessly or not. When considering Wolpaw's definition of a perfect BCI described in Section 1.1, dry-electrodes with passive amplification that connect wirelessly to the processing unit would be ideal. However, when looking at data quality, a wired wet-electrode with active amplification is best. Luckily, recent advancements have made these differences in data quality more acceptable, as will shortly be discussed in what follows.

Wet EEG electrodes are electrodes which require an electrolytic gel to be applied between the electrode and the scalp. This gel functions as a conductor and, as discussed further in Section 2.3.3, currently allows wet electrodes to have better data quality compared to dry electrodes (Cruz-Garza et al., 2017; Mathewson et al., 2017; Tseghai et al., 2021). However, wet electrodes require the assistance of a professional to correctly apply the gel which leaves behind traces after use. Adding to this, the electrolytic gel could also cause allergic effects for the user. Due to the viscosity of the electrolytic gel changing over time, artefacts in measurements may also appear (Tseghai et al., 2021). These are unwanted properties and conflict with Wolpaw's vision of a perfect BCI. The medical-grade EEG equipment shown in Figure 1.2a uses wet electrodes with an electrode cap.

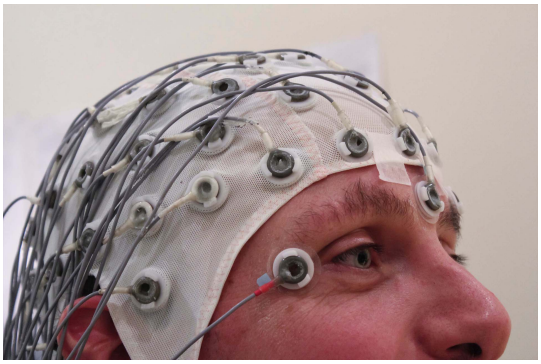
Advancements in dry electrodes are making the gap with wet electrodes smaller and smaller (Cruz-Garza et al., 2017; Mathewson et al., 2017; Tseghai et al., 2021). These dry electrodes don't require the use of an electrolytic gel and given the use of an appropriate headset can be installed on the scalp without the assistance of a professional. Both of these properties are in favour of Wolpaw's properties for a perfect BCI. The main reason dry-electrodes are becoming more viable to be used in real-life environments is due to improvements in active electrode technology (Mathewson et al., 2017). A relatively bulky setup of these dry electrodes with active amplification is shown in Figure 1.2b.

Active electrodes are electrodes which do more than just forwarding their measured voltage fluctuation to the main controller board whilst passive electrodes do just that. This is often necessary since the measured signal is of such low strength that even a short distance cable from the electrode to the main board can cause a lot of noise due to electromagnetic interference (J. Xu et al., 2017). To reduce this noise, a preamplifier is used which additionally amplifies the signal before transmission over the wire as opposed to only being amplified in the main controller board. This makes the final system less compact and more expensive but is often required in anything but lab environments, especially for wireless dry electrodes, as further discussed by Mathewson et al. (2017).

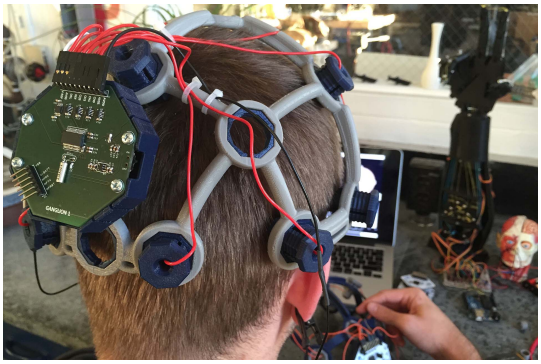
When talking about wireless electrodes, it is not the effective electrode itself that is wireless but rather the communication between the main controller board, a board to which all electrodes are connected by wire, and the processing unit such as a computer. Such a wireless setup is shown in Figure 1.2b. Whilst a wireless approach allows for the creation of an aesthetically more pleasing system where the measuring hardware and processing hardware are physically separated, a wired connection will always remain more efficient and reliable. However, as discussed by Tosi et al. (2017), Bluetooth (BT), an open standard for wireless communication, has seen extensions that are more reliable, power efficient and capable of higher transmission speeds. This has made wireless solutions more appealing in BCI systems but overall issues with wireless solutions, in

general, will prevail. Most important is the risk of connection loss and a higher latency resulting in a longer time between the point a signal is measured and it is received by the computational unit.

All of these advancements have enabled companies such as Muse, Dreem and OpenBCI to develop non-invasive, dry-electrode based EEG measuring equipment with active amplification in an affordable and often aesthetically acceptable manner. As BCIs become even more popular, a heavier focus on affordability and visuals with EEG measuring equipment is to be expected. These two properties were less important in previous medical settings where a patient would wear such equipment only when undergoing a test in the hospital. Figure 1.2 shows the contrast between a medical-grade EEG recording system and one that is consumer-grade.



(a) Medical-grade EEG measuring equipment that uses wet electrodes which are each connected to a separate main board via a long cable. Free to use image by Chris Hope, CC BY 2.0, via Wikimedia Commons.



(b) Commercial-grade EEG measuring device that uses dry electrodes with active amplification connected to an attached wireless controller board. Free to use image by Conorrussomanno, CC BY-SA 4.0, via Wikimedia Commons.

Figure 1.2: The contrast between medical-grade EEG measuring equipment and a consumer-grade alternative.

Algorithmic improvements for non-invasive EEG measuring equipment

Whilst hardware improvements has made the collection EEG data more affordable, reliable and accurate, one important issue still remains. Even with the best active wet electrodes, The contrast between spatial and temporal resolution is enormous. EEG is known to have a good temporal resolution but rather poor spatial resolution. A good spatial resolution would mean that the measurement from electrodes corresponds only to a small, known region of the brain, typically underneath that electrode. Such a correlation is helpful as it reduces noise and increases interpretability of the signal. It also allows for fewer electrodes to be used if only the activity of certain areas of the brain is of interest.

Thus, many attempts have been made at improving spatial resolution of EEG but it has been proven to be a challenging task (Ferree et al., 2001). Besides potential noise of the measurements, this is also caused by the anatomy of the human head. Remember that the electrodes used for non-invasive EEG measuring are placed on the scalp, the skin of the human head. As shown in Figure 1.3, besides the scalp, different structures such as the skull and cerebrospinal fluid (CSF) are in between the electrodes and the actual brain. These components *blur* and *disperse*

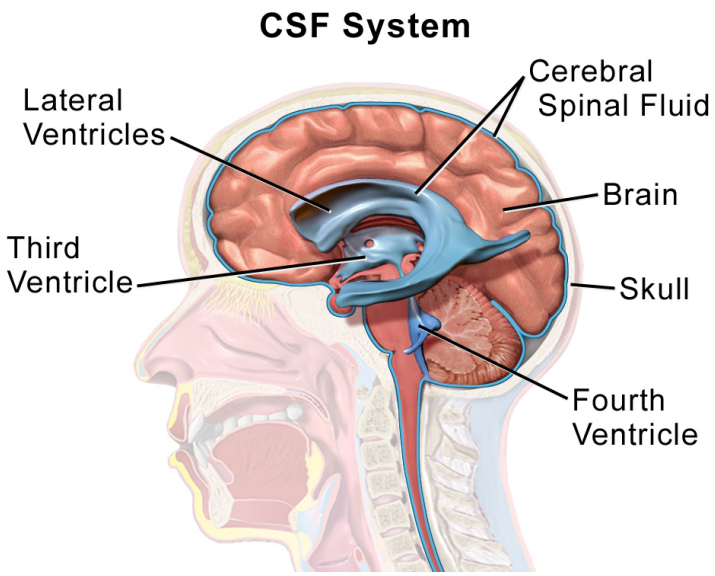


Figure 1.3: The anatomy of the human head, specifically of the cerebrospinal system. Non-invasive EEG measuring equipment is placed on the scalp, causing signals from the brain to be blocked by the skull and cerebrospinal fluid (CSF) among other structures. Free to use figure by Blaus (2014), CC BY 3.0, via Wikimedia Commons.

the perceived brain-signal, making it hard to track where the measured signal came from when looking at the electrical activity on the scalp.

Whilst increasing the number of electrodes placed on the skull physically limits the region under one single electrode, it doesn't guarantee an improve in spatial resolution. Indeed, clever processing algorithms are required to correct for overlapping signals between electrodes so that the overlapped signal is correct for and the effective spatial resolution is improved. Besides this, there is also the issue that decreasing the distance between electrodes introduces the need for placing more electrodes to cover the entire region of the brain. This increases cost, lowers user comfort and decreases the visual acceptance of the system. Another issue with increasing the number of electrodes and decreasing the spatial resolution means that the alignment of the electrodes on the skull is now even more prone to errors and change over time, e.g. due to movement of the user. This makes the need for a professional higher, which is also detrimental with respect to Wolpaw's criteria for a perfect BCI.

Ferree et al. (2001) has found 19-electrode EEG systems to have a highly varying spatial resolution in the 20 to 40 cm^3 range. Systems with 129 electrodes were found to have a spatial resolution of around 6 to 8 cm^3 (Ferree et al., 2001) when also using algorithmic tricks to further improve spatial resolution. However, according to Nunez and Cutillo (1995), around 10^7 parallel pyramidal neurons reside in each cm^3 of the brain cortex. This means the acquired data is still obtained from a incredibly large number of neurons even in the best spatial resolutions.

Whilst hardware improvements in both electrodes and headsets for better placement might improve the spatial resolution further, the spatial resolution improvements possible through hardware have been plateauing. As was the case for the comparison between few and many electrode systems by Ferree et al. (2001), appropriate algorithms have to be used to effectively

increase the spatial resolution. Recently, these techniques often rely on using Laplacians (Liu et al., 2020; Srinivasan, 1999; Srinivasan et al., 1996), although other approaches using for example convolutional neural networks (CNNs) have been proposed (Kwon et al., 2019). These techniques also have the added benefit of cleaning the time-varying signal as described by Yao et al. (2019).

Invasive BCIs try to combat the issues of EEG

The previous parts discussed how non-invasive EEG measurements have been improved. However, alternatives to EEG exist for measuring brain signals, some of which are further discussed in Section 2.3.1. Most notable in recent years is measuring modalities that rely on capturing brain signals by equipment directly inserted into the human body, making it an invasive approach. One such example of an invasive measuring modality is electrocorticography (ECoG) and the most popular invasive BCI at the moment is the one proposed by Musk and Neuralink (2019). The white paper by Musk and Neuralink (2019) has shown that invasive BCIs could greatly exceed the data quality and visual aesthetics of even the best non-invasive alternatives. As further discussed in 2.3.1, this invasive method places flexible electrodes directly inside the skull. These electrodes are invisible to the human eye with the only visual component being a rechargeable wireless transmitter that is magnetically attached to the skull. Neuralink's final aim is to make the brain-signal measuring equipment completely invisible to the human eye.

Musk and Neuralink (2019) has built robots to insert the electrodes inside the skull in a very precise location without the need for an open-skull operation or even anaesthesia. This allows for a magnitude more electrodes to be installed and is expected to suffer far less from noise resulting in a far greater temporal and spatial resolution compared to EEG. This suggests that invasive systems are superior to non-invasive alternatives, but the fact that they are more permanent, far more expensive and invasive gives rise to technical and ethical questions. These ethical questions are further discussed in Section 1.5. From a technical standpoint, maintaining and upgrading a non-invasive BCIs is far simpler and cheaper. The fact that you are inserting foreign objects into the brain also introduces far more health risks than non-invasive systems do. Convincing the user to put on a headset that can be removed will also be far easier than convincing the user to get a BCI permanently implemented in their skull.

It has also been shown that the theoretical more precise temporal and spatial resolution doesn't linearly correlate with improved BCI accuracy/control, rather it seems to plateau relatively quickly with current state-of-the-art signal processing and classification techniques (Aflalo et al., 2015; Lebedev, 2014). Some critics point to the dropping curve found by Aflalo et al. (2015) to conclude that the increased electrode amount and reachable neurons achieved by Musk and Neuralink (2019) don't have a direct impact on the usability of BCIs in real-world applications. Because of these aspects, the ease-of-use appeal and far cheaper price for non-invasive alternatives still outweigh the benefits offered by invasive methods for almost all but highly medical applications, at least in the opinion of the writer of this thesis. Nevertheless, future improvements in signal processing and classification techniques could prove invasive methods to be far superior for BCI applications and the mechanical achievements so far are not to be underestimated. An invasive system is also promising concerning Wolpaw's definition of a perfect BCI discussed in Section 1.1. Once installed, it would ideally require no more assistance from a professional, is aesthetically acceptable as it can be invisible to the human eye, has signs of being far more reliable than EEG and more.

Summarizing the improvement of measuring facilities

Since BCIs rely solely on brain signals to operate, the measuring facilities for acquiring data of those brain signals have a direct impact on the capability of those systems. As was discussed

in this section, the most commonly used modality for non-invasive data acquisition, EEG, has benefited from both hardware and software improvements. From a hardware point of view, the switch to dry electrodes using active amplification and wireless connection to a computational unit has made BCIs more favourable concerning Wolpaw's criteria for a perfect BCI (Mathewson et al., 2017; Tosi et al., 2017; J. Xu et al., 2017). From a software perspective, clever algorithms have enabled preprocessing of the signal to improve spatial resolution (Kwon et al., 2019; Liu et al., 2020; Srinivasan, 1999; Srinivasan et al., 1996). Improving the spatial resolution can also positively affect the temporal resolution due to inherent noise reduction as discussed by Yao et al. (2019). As is further discussed in Section 3.2.3, other prepossessing techniques have also been introduced and refined further aiding in improving the data quality.

1.2.3 Availability of more powerful, affordable and portable equipment

The improvements in brain signal measuring equipment have likely been influential in the gaining popularity of BCIs as it provides more precise data more affordably. However, having the possibility of obtaining clean data is only part of the way to a perfect BCI system. Other improvements concerning computational power, affordability and portability have also played an important role in BCI research, contributing to the rise of popularity in the process.

The emergence of faster and cheaper hardware

As chapter 3 will discuss in greater detail, working with EEG data, or other forms of brain signal data can require computationally very heavy operations to achieve desired processing results of that data. Luckily, together with the improvements in state-of-the-art measuring equipment, there is also an emerging supply of less accurate but far more affordable and portable EEG measuring equipment. Due to Moore's law (Schaller, 1997) and other advancements, central processing units and other computational hardware have also seen massive improvements in computational power. This has made algorithms previously requiring expensive specialized computational hardware possible on the average personal computer. All of these factors have made BCI applications, which were previously limited to lab environments with a high financial cost, accessible to a far broader public. The availability of open-source datasets for common tasks related to brain signals has also allowed computer scientists to experiment in the field without additional hardware cost (Kaya et al., 2018).

Splitting BCIs into multiple major components for portability and reusability

Early attempts at making BCIs more portable and affordable include those by Lin et al. (2008) and Shyu et al. (2010). In essence, these applications rely on separating the data acquisition process and data processing into two standalone systems connected over Bluetooth (BT). Remember from Section 1.2.2 that BT is an open standard for wireless communication that has seen improvement in the last couple of years. Dividing a BCI system in a data acquisition and data processing system allows for creating a lightweight measuring device to be placed on the user's head, with a heavier and bulkier computational unit to process the signals which ideally is still pocket-able. The latter was not a trivial task and introduced the need for custom hardware at the time. Lin et al. (2008) used a custom-made digitial signal processor (DSP) for the task whilst Shyu et al. (2010) opted for a more general field programmable gate arrays (FPGA) based DSP. Whilst these were great demonstrations of how the technology could be used outside the lab, the actual usage for a bci detecting driver's drowsiness (as proposed in the paper by Lin et al., 2008) and allowing multimedia control (as proposed in the paper by Shyu et al., 2010) was

rather limited. The idea of custom-made and possibly proprietary processing hardware which focuses on a single task is also very limiting, although it does have commercial benefits.

What did stick, was the idea of splitting the hardware into two standalone parts, a wireless EEG measuring device and a processing unit. As discussed in Section 1.2.2, a wireless connection between these two components is also favoured when taking into account Wolpaw's criteria on a perfect BCI. It also makes it possible for smaller research teams or even individuals with a certain specialisation to take part in the highly interdisciplinary field by not requiring knowledge of all components but just the one that is of interest. As an example, it enables computer scientists to purchase off-the-shelve affordable EEG measuring hardware and communicate with it through provided libraries for their favourite programming language. In most cases, the personal computer they already own is powerful enough for the experiments, especially for offline systems. This allows for reusing existing hardware which is great from a financial perspective. Section 2.3.3 discusses some of the EEG measuring equipment available on the market. It is noted that EEG measuring hardware is not strictly needed for a computer scientist as researchers such as Kaya et al. (2018) have made excellent free-to-use EEG datasets available.

With the introduction of the iPhone in 2007, it didn't take long for researchers to explore the idea of using a mobile phone as a processing unit for a BCIs. Y.-T. Wang et al. (2011) were one of the first to explore this idea, with a steady-state visual evoked potential (SSVEP)-based BCI. Section 2.4.4 will go into further detail on the types of measurable brain-signals. In essence, such a system relies on a category of brain signals that are often easy to detect but require a specific stimulation. This type of system can be used for a wide variety of applications. Imagine an audio-guided tour in a museum where visitors only need to stare at a screen next to an item of interest to start hearing the explanation of that item. This could be achieved with only a couple of dry electrodes placed on the skull in a headset that also provides the audio to the visitor. This headset could then be connected over Bluetooth (BT) to the visitor's phone running an app for the museum tour. The technology needed for such a system would lean close to that of so-called *P300 spellers*, which have already been heavily studied (Capati. et al., 2016; Hussein et al., 2020; Won et al., 2019). Such a system would also fit perfectly with Wolpaw's definition of a perfect BCI, albeit oriented to a commercial setting rather than a medical one.

Making BCIs a one-in-all device again for profitability

Whilst the advantages of using the computational power of devices a customer already owns are clear, it also imposes some disadvantages. For one, the varying type of computational devices is bound to give varying performance results, compatibility issues and overall limits the guarantee of a pleasing user experience (UX). Adding to this, the measuring equipment and processing equipment can't be connected from the factory resulting in an experience that is not plug-and-play. From a commercial perspective, it would be easier if the system was all-inclusive and possibly patentable.

Recent trends in computing hardware where manufacturers are shifting away from general all-purpose CPUs and them developing their own custom CPU architectures have shown that custom chips can outperform their general counterparts. Patenting the architecture of those chips is possible making it commercially interesting. Apple's mac M series processors announced in 2020 are one such recent example. These M series processors have a neural engine that is stated to accelerate the time needed for ML tasks[2]. Graphics processing units (GPUs) used for autonomous driving systems also differ from general-purpose GPUs.

Because of this, the author of this paper believes custom-made chips could create a future where the headset has a directly integrated processing unit once again. Whilst this would make

[2]https://nr.apple.com/dH8i4U3v2w

for a more attractive package for the customer and give commercial advantages to the manufacturer, it would be disadvantageous for research purposes. The manufacturer could limit the possibilities of using the BCI for different purposes, patent promising hardware and more. Another possible route the author of this paper sees is the use of cloud computing and fast 5G connections to also create a more simple user experience that doesn't require Bluetooth (BT) tethering to a close-by processing unit. This approach would still leave a separation between measuring hardware and processing hardware making changes to any of the two independently easier. Concerning Wolpaw's criteria of a perfect BCI, these approaches would also be acceptable. This belief of switching back to all-in-one devices or using a cloud service for processing the data is further endorsed by the findings of Dillen et al. (2022). In their systematic review of biosignal control systems, eight of the 46 studied papers used embedded hardware and one used cloud solutions.

Summarizing the improvements on computational power, affordability and portability

To summarize, due to Moore's law (Schaller, 1997) and other advancements, CPUs among other computational hardware have seen massive improvements in computational power. This increase in computational power has enabled more advanced processing of the data on more affordable and portable hardware. Early attempts at making BCIs more portable and affordable focused on splitting the brain signal measuring equipment from the data processing equipment (Lin et al., 2008; Shyu et al., 2010). The system by Y.-T. Wang et al. (2011) was one of the earliest examples of a true portable BCI-system that was affordable and relied on a smartphone as a processing unit. It showed how working with BCIs can be done using cheap and general-purpose hardware. The research was published at a turning point for BCIs where publication numbers on BCI-related papers started rising. This hints that the increased affordability and portability combined with more computational power played an important factor in the rise of interest in BCI. The rise of BCI-related papers is illustrated in Figure 1.4 based on data by Saha et al. (2021). Dillen et al. (2022) found that papers on biosignal control systems using DL have seen a steady increase over the last five years as well.

In the future, as BCIs see more commercial applications, this separation of a BCI in a measuring component and processing component might reverse to an all-inclusive device. This has potential downsides for scientific research but makes commercial sense. The replacement of physical computational units in close proximity to cloud solutions is another possible evolution.

1.2.4 New and refined data processing techniques

The previous sections 1.2.2 and 1.2.3 discussed how both measuring and computational hardware have seen recent improvements. Another important part of the puzzle is the algorithms that convert data from the now more user friendly measuring devices to useful actions using the now more powerful, affordable and more portable computational hardware. Most of these algorithms are data-driven classifiers that use ML techniques. In more recent years, deep learning (DL) techniques and alternative approaches have been incorporated in the BCI pipeline as well, which has been proven to be very successful. Chapter 3 discusses commonly used techniques in more detail and multiple ML and DL based BCI pipelines are discussed in later chapters of this thesis. This section gives a more high-level summary of recent developments in the AI field that have likely contributed to the rise in popularity of BCIs.

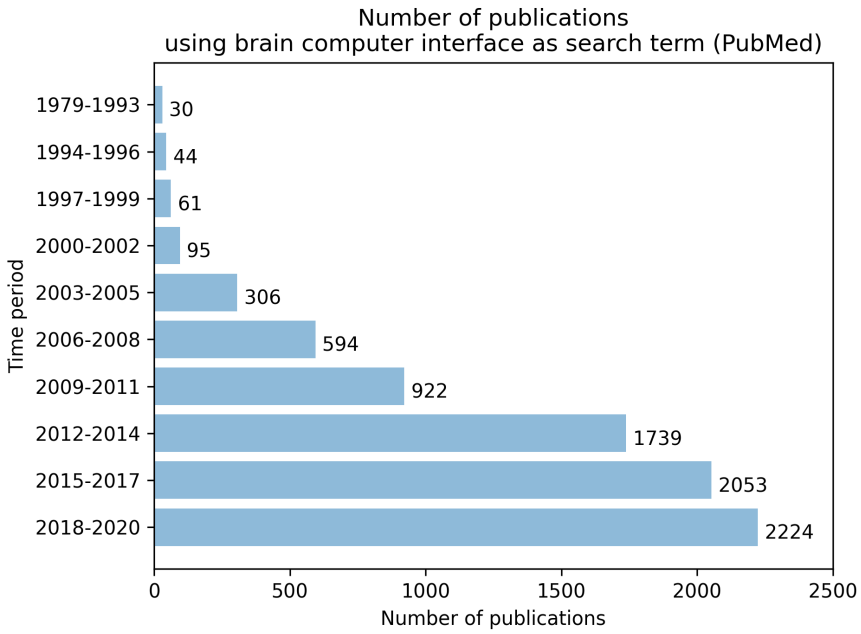


Figure 1.4: The number of BCI-related papers over time. Figure after the results of research by Saha et al. (2021) who searched PubMed using the keyword "brain-computer interface".

Postponing another AI winter

Machine learning (ML) and deep learning (DL) are techniques that fall under the AI umbrella. These techniques are being used as buzzwords in a whole suite of applications and it seems as if every week there is yet another big promise or threat related to AI discussed in major news outlets. Recent examples that have shown the world what new techniques in this field are capable of include the Go champion beating computer algorithm by Silver et al. (2016), the impressive text generation model GPT-3 by Brown et al. (2020) and the image generation model DALL-E by Ramesh et al. (2021). This abundance of new achievements and an overall high public interest in anything that mentions buzzwords from the AI umbrella has caused a long lasting AI summer since the last AI winter of the late 1980s and early 1990s. Such an AI summer means that there is incredible amount of funding available for improving ML and DL techniques among others. This in term causes further advancements in the field of ML and DL which results in more impressive achievements.

However, an AI summer also implies that an AI winter will inherently return. An AI winter is a period of time where the interest in the field is reduced and thus funding and research is limited. As discussed by Floridi (2020), such an AI winter may be relatively close. This is in part due to new regulations and public backlash on the more questionable but highly profitable applications DL is involved in. A recent example of this is the controversy surround Clearview AI. Here, state-of-the-art DL image recognition algorithms are used on billions of images collected from all over the internet, including social-media platforms, to recognize almost anyone with a public profile linked to them. As further discussed by Rezende (2020), this technology conflicts with many EU laws yet was used by multiple police departments. Adding to this, new regulatory changes

are being proposed to limit the use of algorithms which lack explainability and interpretability (European Commission. Joint Research Centre., 2020; The Royal Society, 2019).

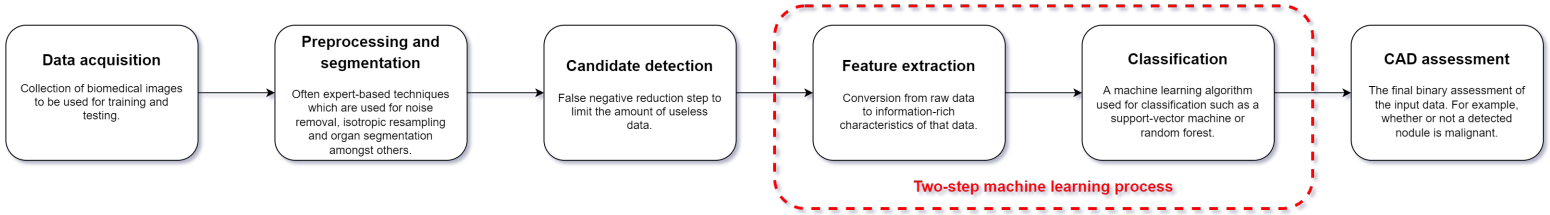
Nevertheless, there is still a high amount of resources being put into ML and DL research. Throughout history, these technologies have been linked with the biomedical setting a lot. As explained by Baldi (2018), DL and biomedical data have directly influenced each other's evolution's since the 1980s. Because of this, applications that process biomedical data have been an important factor at prolonging the current AI summer. Since BCIs use biomedical data as well, they have been one of the applications keeping interest in ML and DL research high. This is in part due to the science-fiction properties BCI systems have creating a lot of public interest as already discussed when talking about Elon Musk's Neuralink in Section 1.2.1. Thus, BCI systems, which rely heavily on ML and DL, are one of the research areas in these technologies that are so promising they help prolonging the current summer of AI.

Improved and new ML and DL concepts have enabled more capable BCI systems

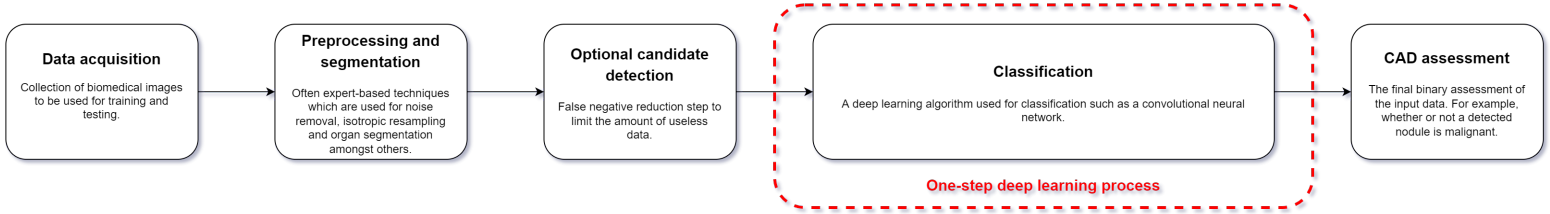
Most of the main concepts from both machine learning (ML) and deep learning (DL) are already multiple decades old. As DL is a subset of the ML techniques, pipelines for using these techniques are very similar. To illustrate this, a general pipeline of a computer-aided diagnosis (CADx) system used for classification is given in Figure 1.5 and commonly used techniques are discussed below. It is noted that besides classification tasks, some regression problems for CADx systems exist as well. However, such regression problems are far less common in BCI systems relying on EEG with the systematic review article of Dillen et al. (2022) only finding articles on classification problems for such systems. Because of this, this thesis focuses on classification problems. CADx systems are used extensively in hospitals for the interpretation of biomedical images and have been studied ever since computers were invented. The most common example of a CADx system is the classification of lung images as being either from a lung cancer patient or not, often also highlighting the nodules used for this classification. These pipelines are very similar to the ones used for BCIs, which is further discussed in Section 3.2.

Traditional ML is a commonly used name for the collection of ML techniques that are not DL. Such traditional ML rely on a *two-step* process where there is both a feature extraction step and a regular classification step as is visualised in Figure 1.5a. Feature extraction is the process of representing the often highly dimensional and unstructured raw data using characteristic properties. These representations are often derived from expert knowledge and are chosen by the designer of the system rather than learned from data. With respect to a CADx system for classifying if a patient has lung cancer, these features might be the size of nodules in the lung, a metric of how round these nodules are, the Hounsfield unit (HU) of certain parts of the lungs and more. These features are then used for learning by a traditional ML classifier. The true challenge in these two-step traditional ML systems lies in finding appropriate features, which can be a very time-consuming task that requires a lot of domain knowledge. It isn't uncommon for features in CADx systems to be refined in the span of multiple years (van Ginneken, 2017). Section 3.2.4 discusses the feature engineering step in more detail and Section 3.3.3 discusses traditional ML in more detail.

Alternatively, a *one-step* approach in CADx systems denotes the use of DL in the pipeline, such as shown in Figure 1.5b. DL generally differentiates itself from the previously discussed two-step traditional ML approach by working directly on the preprocessed data rather than a feature representation of the data. However, this isn't necessarily the case as feature extraction can still be beneficial for some DL methods whilst others might even work better on the raw, non-preprocessed, data. Section 3.3.2 discusses the difference between traditional ML and DL in more detail. Intuitively, DL models exist of multiple layers which can be seen as a graph-



(a) General pipeline for ML based CADx system.



(b) General pipeline for DL based CADx system.

Figure 1.5: General pipelines of a CADx system used for classification. A ML approach is called a two-step approach as it exists from feature extraction and classification. A DL approach is called a one-step approach as it combines both of these steps into a singular classification algorithm.

like structure which becomes deeper as the number of layers increases, hence the name deep learning (DL). Typically, multiple types of layers are present in a DL approach, with the earlier layers focusing on what could be seen as gradual feature extraction steps whilst the later layers are often more tailored towards the classification of those data-driven feature outputs from the previous layers. Section 3.3.4 discusses DL and the common variations in more detail. A DL approach is interesting as it can allow for skipping the feature extraction process where good features have to be found by the developer of the system. Finding such good features can not only be time-consuming but limited expert knowledge can also make creating features that carry enough differentiating information impossible, meaning that even the best classifiers can't learn from them. Adding to this, DL approaches in CADx systems have also been proven to outperform state-of-the-art ML approaches (van Ginneken, 2017). However, DL models are more challanging in terms of explainability and interpretability. The challenge in DL approaches is finding a suitable model and training it in a way such that it doesn't overfit. In a way, DL allows a person with limited domain knowledge of the data to obtain a better result as it bypasses the need for manual feature extraction. However, this is a double-edged sword as limited knowledge of the data increases the risk of bias, overfitting and other potential issues.

Relating this to BCIs, which have a very similar pipeline, and how the ML and DL techniques used often originate from older concepts, some of the most common BCI pipeline components are considered. A typical traditional ML BCI pipeline often relies on a form of the common spatial pattern (CSP) technique for feature extraction. This technique is quite old being introduced by Koles et al. (1990) around 30 years ago. Likewise, the traditional ML classification used is often a type of support vector machines (SVM). Once again, this technique was first introduced by Boser et al. (1992) around 30 years ago. Over the years, CSP has evolved and many extensions such as filter bank common spatial pattern (FBCSP) by Kai Keng Ang et al. (2008) have been introduced. Likewise, SVM has seen many extensions and improvements (Chervonenkis, 2013; V. Utkin, 2019). This has resulted in the combination of these two relatively old techniques, but with recent extensions, still performing among the state-of-the-art in BCI applications using traditional ML approaches (Rasheed, 2021; Tangermann et al., 2012). Likewise, when using a DL approach in the BCIs pipeline, CNNs are most commonly used (Dillen et al., 2022; Rasheed, 2021). This technique is again a rather old one, the foundations of which were first described by Fukushima (1980) over 40 years ago. Just as the CSP procedure, CNNs have seen multiple extensions and improvements over the year, just as other DL approaches. Most recent and noteworthy developments are works such as that by Schirrmeister et al. (2017) which focuses on proposing CNN architectures that can learn from very few samples and for which a described method exists to visualize the model's early layers. Other improvements, more general to DL, have also been made such as the development of new activation functions. D. Lee (2020) discus how changing the activation function from CReLU to ReLU6 offered a 35% performance increase while keeping other components fixed for certain experiments relying on a neural network (NN) in a reinforcement learning (RL) setting. Thus, these improved versions of older concepts have enabled far better performance making it possible to create more capable BCI systems.

This doesn't mean that all approaches used for processing the data in BCIs rely on decades-old techniques that have improved over the years. One interesting and relatively new approach is the use of transfer learning (TL) from drastically different domains. Previously, TL was mostly used in BCIs to train a model on data which may originate from different users performing similar but not necessarily identical tasks. This general model is then further refined on a specific patient and task, transferring the knowledge acquired from the previous data to the new data. When done correctly, this can provide far better performance compared to learning on the new data alone for problems where data is limited (Dillen et al., 2022). As available data specific to BCIs applications remains limited, some recent research has gone into transferring knowledge from

completely different domains to BCI specific data. G. Xu et al. (2019) used a model pretrained on images and transferred it to EEG data for a motor imagery (MI) task with promising results. Other attempts at transferring knowledge from other domains, such as natural language, have also been made (De Wulf, 2022).

More open-source datasets and BCI related libraries

Whilst more affordable and portable measuring hardware has enabled a low-cost solution for researchers to acquire their own data, as discussed in Section 1.2.3, the process of acquiring BCI related data remains significantly time-consuming and can impose multiple challenges. As addressed by Dillen et al. (2022) many publications don't properly report the data acquisition process, leaving a lot of ambiguity in both the meaning of data labels and how representative of the real-world the data is. Whilst two papers might discuss a MI task as being *imagined left-hand movement*, one might have collected the data in a lab-like environment from a trained user envisioning a single squeezing shut movement of the hand whilst another might have it correspond with any envisioned movement of the fingers on an untrained and non-focused user. On a legal aspect, as data on brain signals is biomedical data, heavy regulations are in place on how this data can be shared and used, next to the general data protection regulation (GDPR) (Malin et al., 2013; Vlahou et al., 2021).

Considering that many researchers in the field often focus on one specific component of the BCI pipeline, it is not feasible for them to go through the trouble of collecting data themselves. Adding to this, when a researcher wants to reuse a certain component of the BCI pipeline from another author's work, the source code of their project is often not present, not well documented, or not compatible with their programming environment (Dillen et al., 2022).

Luckily, BCI specific data and coding libraries have been made available in recent years. Some of the earliest BCI related datasets that are still commonly used to this day are from the BCI competitions organised by the Berlin Brain-Computer Interface research program (BBCI)[3]. Of the four different competitions, two were in part organized by Jonathan R. Wolpaw, from who this thesis took the definition of a BCI and a perfect BCI, further demonstrating his importance in the field (Blankertz et al., 2004, 2006). The fourth and final competition provides three EEG datasets labelled with MI tasks, making it a popular choice in literature. More recent datasets include those by Kaya et al. (2018), who have put a tremendous focus on discussing all necessary details of the data acquisition process, including testing the MI skills of each subject and providing the software and instructions given to the subjects. Not only do these publicly available datasets allow researchers to skip the data acquisition step, but it also improves the reproducibility of their work and makes comparing it to other work using the same dataset easier. However, there is still far from an abundance of data that can be used for training BCI pipelines compared to other fields of research. Indeed, collecting datasets of books for natural language processing (NLP) learning applications or cat images for computer vision applications is a far easier task than collecting EEG datasets for identical MI tasks with equal data acquisitions methods.

From a code perspective, many authors still fail to deliver a copy of their source code along with their article (Dillen et al., 2022). Some websites, such as Papers With Code[4] has been created to more easily find papers that do provide their code, but for BCI research this is still limited. Luckily, advanced libraries have been emerging which provide a multitude of common operations from the BCI pipeline. Perhaps the most famous of which is the Python MNE library by Gramfort (2013) which provides tools for organizing, visualizing and processing EEG data

[3]https://www.bbci.de/competition/
[4]https://paperswithcode.com/

such as windowing the EEG data, performing baseline correction and determining the CSP features. Other famous Python libraries include Braindecode by Schirrmeister et al. (2017) and EEG-DL by Hou, Jia, et al. (2020), Hou, Zhou, et al. (2020), Hou et al. (2022), and Jia et al. (2020) among others. Whilst Python is the most commonly used programming language for implementing BCI pipelines and has by far the most supporting libraries, libraries for MatLab (e.g. EEGLab by Delorme and Makeig, 2004), C++ (e.g. Brainaccess by Neurotechnology[5]) and other popular programming languages are starting to emerge as well. Besides these general libraries, some of the most famous articles whose source code isn't provided have also seen open-source implementation based on the original author's description. Most noteworthy is the Army Research Laboratory (ARL) EEGModels Project which provides implementations of the EEGNet model proposed by Lawhern et al. (2016) and the ShallowConvNet and DeepConvNet models proposed by Schirrmeister et al. (2017) written in Keras and Tensorflow (Python DL libraries by Chollet et al., 2015; Martín Abadi et al., 2015).

Summarizing the emergence of specialized data processing techniques

Besides improvements on a hardware level, both for the measuring equipment and the processing equipment, significant software improvements have also been made. Most importantly are the improvements made to two-step traditional ML approaches and single-step DL approaches. This includes improvements of the CSP feature extraction algorithm such as FBCSP by Kai Keng Ang et al. (2008), improvements to the SVM classifier (Chervonenkis, 2013; V. Utkin, 2019) and general DL improvements (Dillen et al., 2022; D. Lee, 2020; G. Xu et al., 2019). The introduction of more open-source datasets and libraries also aids in far faster development of new BCI pipelines and allows researchers to focus on a specific component of the BCI pipeline to improve. The fact that BCIs still provide a science fiction feeling to the general public has also made it one of the technologies that aid in postponing another AI winter.

Relating this to Wolpaw's definition of a perfect BCI, it becomes apparent that these software improvements play an important role in achieving the goal of a perfect BCI. Better pipelines should allow a BCI system to function more reliably and in more challenging environments. Improved techniques could also improve the number of classifications a system can handle in a certain period. This can enable the BCI system to match or even surpass the conventional interaction method it wishes to replace. Whilst these are very important aspects of a perfect BCI, it should be remembered the user of a BCI will most likely not appreciate this evolution to the same degree a computer scientist will. At the end of the day, an ideal processing pipeline should be one the user never notices is there, whilst a wrong classification that results in the wrong command being executed or other issues with the pipeline will result in unpleasant experiences the user will remember.

1.2.5 Summarizing the cycle of increasing popularity

When looking at the number of papers published on PubMed with the keyword *brain-computer interface* visualised in Figure 1.4, a clear upwards trend is visible over the years. This upward trend seems to have started in the 2000s and the jump was most significant around 2012. Whilst proving which elements are responsible for this upward trend in both scientific and commercial interest isn't directly possible, this section has highlighted multiple potential reasons. It is unlikely any one of these potential reasons is the sole reason for the increased popularity of BCI systems. Rather, all of the discussed reasons likely influence each other as portrayed in Figure 1.6.

[5]https://www.neurotechnology.com/brainaccess-documentation/C++Api

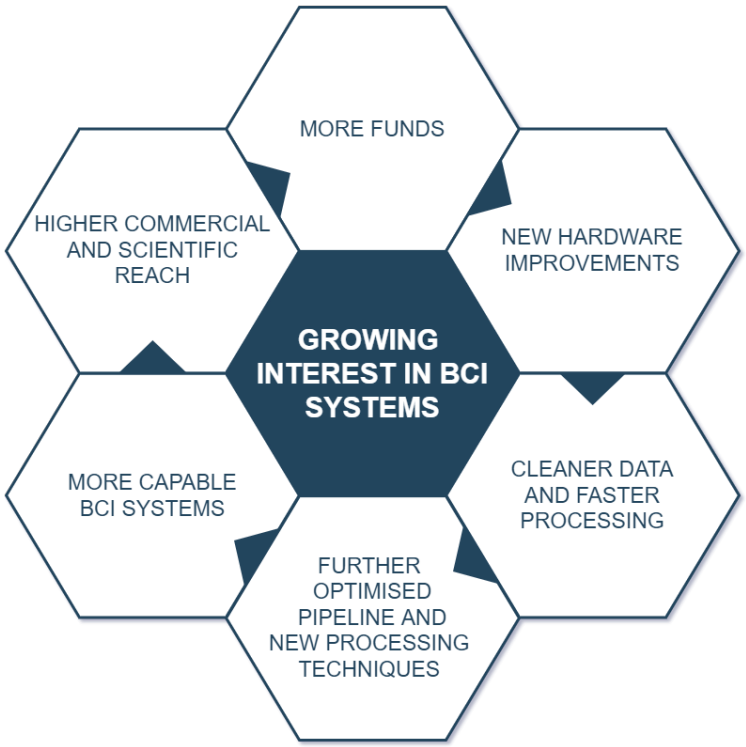


Figure 1.6: Summary of some of the potential reasons BCI research has seen its gain in popularity discussed in section 1.2. There is no clear chronological order for these events and they likely influence each other all at the same time.

The potential reasons that were discussed can be separated into four major categories. First, an increased interest by big-tech companies causes big funds to get into the field. Secondly, improvements in measuring equipment allow for higher quality data and a more pleasant user experience (UX). Third, processing equipment has seen increased affordability whilst also improving the computational power and portability allowing systems to be more affordable and capable. Finally, more optimised software allows for better classification results and faster development of new systems. It is highly unlikely all of these reasons have equal influence or that they are the only reasons for the improved popularity, but they should give the reader a better understanding of the context BCIs are currently in without becoming too technical. The next section focuses on effective use cases for BCIs to let the reader understand the different capabilities of these systems even more.

1.3 Common use cases for BCIs

The previous section aimed to address potential reasons for the increase in popularity in BCI research. This already discussed multiple BCI applications, showing some of the use-cases for BCIs. When looking at Wolpaw's definition of a BCI, a BCI is nearly limitless in what it can be used for. Whilst it should be a *communication device*, this terminology can be seen rather broadly.

Most obvious is perhaps the use of a BCI as a control system, which then communicates with an external device so that it can be controlled. Already discussed examples of this type include the multimedia control system proposed by Shyu et al. (2010) from Section 1.2.3. However, when using BCIs to detect common patterns of a disease, these patterns could also be communicated to a doctor and hence it could also be considered a communication device. It should also be noted that the given definition of both a BCI and perfect BCI from section 1.1 are around 20 and 10 years old respectively. During the time that has passed since then, many new use cases for BCIs have emerged, especially in a commercial setting. Thus, these definitions are by no mean the only definition for what a BCI system is and it is not uncommon to find varying opinions on what can be called a BCI system and what can't be called one. This section aims to familiarize the reader with a broad idea of possible use cases of BCIs and to give some concrete examples whilst also addressing some of the shortcomings found in the articles describing them. As this thesis focuses on non-invasive EEG based BCIs, this section will also focus on non-invasive EEG related articles.

For the interested reader, Dillen et al. (2022) provides more examples of biosignal control systems, of which BCIs are a subset, that use DL. The focus of Dillen et al. (2022) is on articles that provide at least a completely working proof of concept (POC), from live data acquisition to external device control. Rasheed (2021) reviews several BCI specific papers and focuses on the type of ML task that is performed and which ML and DL techniques are used for this. Only a few articles discussed by Rasheed (2021) have a working proof of concept (POC) on live data with many simply looking at the evaluation metrics of the ML and DL strategies. Whilst the articles discussed by Dillen et al. (2022) are thus more representative of real-world performance and use-cases, finding such complete works is rather challenging. Because this section is only meant to demonstrate potential use cases of BCIs and common shortcomings in the papers discussing them, there is no enforcement of requiring the discussed papers to have a fully working POC. This allows for offline systems and articles focusing on only certain components of the BCI pipeline to be discussed as well. It also allows for discussing systems which may not strictly be BCIs but which do use almost identical pipelines, further demonstrating what is possible when working with brain signals as data. In this regard, it is similar to the work of Abdulkader et al. (2015) who give an overview of the most common use cases for BCIs and the challenges they are faced with.

1.3.1 Using BCIs as an automated tool in all stages of medical phenomena

As EEG is relatively affordable and portable, it is an often used modality in the hospital. The visualisation of the measurements can be read by a specialist to diagnose a whole suite of diseases. Especially seizure-related disorders such as epilepsy are often diagnosed through EEG readouts (Rosenow et al., 2015). Other neurological disorders such as the locked-in syndrome (LIS) can also be correctly identified by EEG readouts (Markand, 1976). Even the psychological field uses EEG as a physiological measuring tool to aid in the diagnosis and psychological treatment of patients, although psychiatrists should be aware of the limitations EEG has for them as further discussed by Badrakalimuthu et al. (2011). Since EEG is used for a wide variety of diagnoses and by far the most common source of data for non-invasive BCI (as discussed by Dillen et al., 2022), much research has gone into how BCIs can be used as an automated aid for medical diseases throughout all stages of a medical phenomena. In general, three important distinctions are made for the stage of such phenomena and thus for the medical application of the BCI: prevention, diagnosis/monitoring and rehabilitation/restoration.

Using BCIs for the prevention of medical phenomena

One popular example of prevention relates to the use of BCI systems for continuously monitoring signs of decreased attention (Abdulkader et al., 2015). Such a decrease in attention might show an exaggerated usage of alcohol and other drugs, which allows the user to be alerted before drinking *just one more*. Whilst it is unlikely that alcohol addicts will wear a BCI to the pub, such devices can significantly help in the reduction of traffic accidents by becoming mandatory during driving. According to World Health Organization (2018), traffic accidents were the number 1 cause of death for children and young adults. Whilst many types of equipment are already in place to prevent traffic accidents from occurring, such as speed cameras and lane assist technology, it has been studied how BCIs can be used as a prevention measure as well.

Section 1.2.3 discussed research by Lin et al. (2008) as one of the first attempts to make a fully portable BCI. As it measured drivers' drowsiness levels, the proposed system alerts or even enforces drivers to take a break from driving when drowsiness levels are too high. Research such as that by Fan et al. (2012) has discovered that emergency situations in vehicles can be classified faster from EEG data than the user's response, which could have a significant impact on reducing the number of traffic accidents. Fan et al. (2012) used a driving simulator to conduct a real-world experiment and they found that their proposed system has an accuracy of around 70%. One potential issue with this paper is the fact that their focus lies on this overall accuracy metric rather than providing the *sensitivity* and *specificity*. Sensitivity could be seen as the accuracy of the model on positive cases (i.e. identifying a dangerous scenario as dangerous) whilst specificity denotes the accuracy on negative cases (i.e. identifying a non-dangerous scenario as non-dangerous). However, Fan et al. (2012) do provide the receiver operating characteristic curve (ROC curve) and provide reasoning for choosing an optimal classification threshold that resulted in this 70% accuracy. This threshold specifies how certain the classification prediction should be before acting on its classification. This shows that it is important to look at more than just accuracy, as for this application being good at taking action in case a dangerous scenario is present is ideal, but taking action when there is no dangerous scenario could create a dangerous scenario itself. Section 6.4 discusses how to evaluate systems using different metrics and reasoning in more detail. Many other examples of using BCIs for the detection of symptoms that can lead to traffic accidents exist. Chin-Teng Lin et al. (2013) developed a system that could estimate motion sickness levels, which in turn could again be used to alert drivers to take a break.

The three discussed examples show how BCIs can be used in a variety of ways for the prevention of traffic accidents. It also highlighted the importance of correct model evaluation and how a net benefit of a system is important. For example, if the system proposed by Fan et al. (2012) would use the emergency brake each time it detects a dangerous scenario but has a poor specificity, it would heavily break in circumstances where it is not needed. This in term creates risk if nearby drivers don't pay attention and cause a collision due to the abrupt emergency braking. Thus, the question should be asked if a net benefit would be reached by those systems if the amount of prevented accidents outweighs the added risk of new accidents. Likewise, it has to be considered if a BCI system is the most viable option, even if it has promising results. Using a BCI to monitor if a driver is drunk and prohibiting him from driving if so might be more expensive and less efficient than breath alcohol ignition interlock device (BAIID) which already exist. Nonetheless, many applications exist for the prevention of traffic accidents and the prevention of other medical phenomena using BCIs, even if it is used as supplemental validation over existing devices and tests. Abdulkader et al. (2015) provides some extra examples of this type.

Using BCIs for the diagnosis and monitoring of medical phenomena

Section 1.2.4 already introduced the computer-aided diagnosis (CADx) pipeline. CADx systems, and the very similar computer-aided detection (CADe) systems, using medical imagery are widely used for diagnosis and treatment monitoring in oncology and other fields. Comparable systems exist using EEG and other brain-signal measuring modalities for diagnosis medical phenomena and monitoring of the treatment. The commercial Muse headset for sleep tracking that was discussed in Section 1.2.1 is one example of a system that can be used for monitoring and diagnosing sleep patterns.

Whilst the Muse headset is a commercial product, medical BCIs exist which are made for detecting sleep patterns. One very popular example is using EEG for detecting sleep apnea, a medical sleep disorder. Taran and Bajaj (2020) proposed an EEG based BCI system for automating the detection of sleep apnea. Taran and Bajaj (2020) named their used ML approach *Artificial Bee Colony Optimize Hermite Basis Functions*. This paper is an example of a system that is highly specialized for a specific problem, namely detecting sleep apnea, requiring a lot of expert knowledge since it relies on a complex feature engineering step tailored towards detecting sleep apnea. Whilst this makes it unlikely the same pipeline can be used for the detection of other disorders, it does provide outstanding performance with a discussed sensitivity, specificity and accuracy of over 99%. The article by Taran and Bajaj (2020) focuses on the ML aspect of the proposed system, by re-using an existing dataset and describing only metrics related to the classification performance. Since this system doesn't take specific control over an external device, it isn't a control system. However, it could be argued it still communicates the detection of sleep apnea patterns and it can be deployed in a live system to alert a nurse when these patterns occur during a trial so that the nurse can pay additional attention during that period. In this regard, naming the system proposed by Taran and Bajaj (2020) a BCI is controversial, as is the case for most EEG based systems that diagnose and monitor medical phenomena. Whilst Taran and Bajaj (2020) don't mention the term brain-computer interface (BCI) in their article anywhere, in an article of a similar system, Poorvitha et al. (2020) do call their EEG based detection of sleep apnea a BCI.

Many other types of diagnosis and monitoring can be done by non-invasive EEG based BCIs. The review article by Abdulkader et al. (2015) addresses the detection of brain tumours (Selvam & Shenbagadevi, 2011; Silipo et al., 2013) and breast cancer (Poulos et al., 2012). The review article by (Fadzal et al., 2011) discusses the use of BCIs for the detection of dyslexia. It should become apparent that many detection systems for medical phenomena can be made from non-invasive EEG data. Even if these medical phenomena seem unrelated to the brain, such as for the detection of breast cancer (Poulos et al., 2012). However, as these systems lack a true communication factor besides informing doctors or nurses, it can be argued if these systems are indeed BCIs w.r.t. Wolpaw's definition of a BCI given in Section 1.1.

Using BCIs for the rehabilitation and restoration after medical phenomena

Restoration of lost mobility and rehabilitation is the third and final stage of most medical phenomena. In this stage, a BCI can be of aid to both the medical staff and patient. Out of the three discussed stages it is also the one that has seen the most BCI related research. One reason for this is that many techniques for the restoration of lost mobility and rehabilitation rely on EMG, a modality used for measuring muscle-based biomedical signals (biosignals). Famous examples of such EMG based systems include prostheses which are connected to the remaining muscles of an amputated body part or an exoskeleton which is placed over body parts to strengthen their movement or re-enable them. Whilst in general EMG based systems far outperform EEG based systems for prosthesis and exoskeletons, the EMG modality does not apply to patients suffering

from an impaired neuromuscular system. Thus, for these patients an EEG based solution may offer them a possibility of regaining mobility again. Since EEG based BCIs for prosthesis and exoskeleton control are very common research in the BCI field, it will be discussed further in it's own section, Section 1.3.2.

BCIs have found their usage in other aspects of the rehabilitation and mobility restoration process as well. In neurological rehabilitation, BCIs can function as a tool for guiding both patients and medical staff in what abnormalities are present in the brain and how they can be circumvented. Intuitively, these BCI systems aim to improve the brain's ability to adapt itself based on experience. This self-adapting property of the brain is known as brain plasticity or neuroplasticity. A well-developed BCI system could show patients and medical staff which regions of the brain are currently being used and which types of brain signals and abnormalities are being detected. This information could then be relayed to the patient through various means so that it induces neuroplasticity. Such systems are still in development and require sophisticated neurological expertise that falls outside the scope of this master thesis. However, they are very promising and the interested reader is referred to the overview provided by Daly and Wolpaw (2008). It is noted once more this paper was co-authored by Professor Jonathan R. Wolpaw, which has been discussed in this paper multiple time for his definition of a BCI and a perfect BCI.

Another area that is often researched in the rehabilitation and restoration process, is using motor imagery (MI) tasks to improve recovery of control over certain body parts. The main idea behind such systems is to let the patient perform a MI task, which doesn't require effective movement of the body part with reduced control. This MI task can then be classified by the BCI system based on the EEG data and an exoskeleton is used to perform that motion as passive feedback. Most commonly, a pedalling MI task is used to control a motorized pedal, which evokes event-related desynchronization (ERD) patterns resulting in a greater potential for lower-limb recovery. The works by Cardoso et al. (2021) and Cardoso et al. (2022) discuss such systems in greater detail. Likewise, stroke rehabilitation and a reduction of the effects from the Parkinson's disease can be achieved with BCI guidance as well, as further discussed by Adama and Bogdan (2021).

It is noted that many more BCI applications exist for preventing, monitoring and controlling diseases than those discussed here. The work by Shih et al. (2012) highlights some use-cases of BCI in medicine for example.

1.3.2 Using BCIs for prosthesis and exoskeleton control

Arguments were given why the medical applications of EEG processing systems discussed in Section 1.3.1 can or can't be considered BCI systems. The main discussion revolves around the kind of communication that the system does. To some authors, a system that alerts or visually informs medical staff based on processed data from EEG can be considered a BCI system whilst others specifically expect an external device to perform a specific action based on the brain signals to name a system a BCI system. The latter would be a type of control system, as the BCI controls an external device. In a review article by Dillen et al. (2022), it is found that among EEG based BCI control systems using DL, the control of robotic arms is most popular in research. This section will discuss the use of BCIs as a control system for prostheses and exoskeletons.

Why BCIs are used for prosthesis and exoskeleton control

Most of the current robotic prostheses and exoskeletons rely on muscular activity in the body. Just like brain activity can be measured by non-invasive electroencephalography (EEG), muscular

activity can be measured by non-invasive electromyography (EMG). For example, patients who have had a partial loss of a body part often still have muscular activity in the remaining body part. In the case of surgical amputation, the amputation is often done such that as much muscle remains to ensure as much possible muscular activity also remains. These patients can then often still generate the muscular activity that would have been required for moving the missing body part. This muscular activity can then easily be measured with the EMG modality, which is in general of higher signal quality than EEG (Bakshi et al., 2018). This EMG data can then be processed to classify the wanted movements and control robotic prostheses. Sudarsan and Sekaran (2012) discuss the design and development of such a system based on EMG in greater detail. Since the overall pipeline of such EMG based systems is quite comparable to the pipeline of EEG based BCIs, many of the processing techniques are similar to those discussed in this thesis. Alternatively, when the body part is still fully intact but the control over that part is lost or requires extra support, an exoskeleton may be used similarly. Just like robotic prostheses, most exoskeletons rely on EMG. A thesis by the German Fleischer (2007) highlights the fundamentals of EMG based exoskeletons.

As was already touched upon in the rehabilitation and restoration part of Section 1.3.1, EMG measurements are not applicable for all patients. In particular, people who have neurological diseases, limiting the production of the required muscular-based biosignals, fall outside the scope for these solutions. However, the viability of robotic prostheses and exoskeletons for these patients has been steadily on the rise as EEG based BCIs have been proposed for this purpose. AL-Quraishi et al. (2018) give an in-depth systematic review of upper and lower limb exoskeletons and robotic prostheses controlled by EEG-based BCI. Just as was the case for rehabilitation and restoration discussed in Section 1.3.1, these systems often make use of brain signals measurable after imagined movement. This process of thinking of movement but not doing the movement is known as motor imagery (MI) and further discussed in Section 2.4.4.

The risk of using BCIs for prosthesis and exoskeleton control

Whilst it is clear to see why EEG based BCIs are promising for prosthesis and exoskeleton control from the discussion above, it should be noted that there is also a great risk in doing so. AL-Quraishi et al. (2018) address the high risk associated with failed instructions for robotic prostheses and exoskeletons. This means that the risks that can follow from misinterpreted instructions of exoskeletons and robotic prostheses are of such a degree that even high accuracy systems might not be good enough to guarantee a net benefit for the user. This could be the reason that the review article by Dillen et al. (2022) didn't find any proof of concept (POC) applications for exoskeleton or prosthesis control using EEG based BCIs. However, Dillen et al. (2022) did find many papers controlling an external robotic arm which is not connected to the body and thus not seen as a prosthesis.

AL-Quraishi et al. (2018) also highlight that whilst multi-label classification of EEG is possible with considerable accuracy in an offline lab setting, the number of detectable classes is limited in a real-time and real-life environment. Because of this, EEG-based systems in these applications still have some challenges to overcome to match the precision and reliability of EMG counterparts. Whilst improvements regarding these aspects have been made since the work of AL-Quraishi et al. (2018) was published, the main challenges remain to this day, especially when using affordable systems. Because of this, widespread adoption of EEG-based exoskeletons and robotic prostheses is still very limited.

Promising steps towards BCI controlled prosthesis and exoskeleton control

Whilst the discussed risk has caused widespread adoption of EEG-based exoskeletons and robotic prostheses to still be very limited, promising steps are being made to bring them to market. These systems mainly focus on robotic arms, which are not strictly a prosthesis but they pave the road to BCI controlled arm prosthesis in the future. These proposed systems often rely on incorporating supplementary information such that risky movements are cancelled, even if the BCI requests them. B. Xu et al. (2022) proposed such a system to control a robotic arm not only through a EEG based BCI working with MI tasks but also by using obstacle avoidance algorithms to reduce the risk of harmful contact, computer vision for object detection to get a better idea on the possibly wanted interaction and eye-tracking to gather extra information surrounding the user's intention. Other successful research on making more reliable BCI systems for controlling robotic arms is done by Kuhner et al. (2019), Sahaya et al. (2020), Shim et al. (2019), and Tayeb et al. (2019), these articles are also included in the review article on DL based biosignal control systems by Dillen et al. (2022).

Whilst limb prostheses, currently being pilot run through robotic arm research, are one of the most common and researched types of prostheses, they are only a fraction of all prostheses in existence. Everything from dentures to artificial breasts can also be labelled as a type of prostheses. For example, visual prostheses such as bionic eyes are another type of prostheses that has active research in the BCI field. Not only can BCI systems improve visual prostheses, many of the existing visual prostheses could be seen as a special type of BCI system as a whole. Both the works by Ptito et al. (2021) and Niketeghad and Pouratian (2019) give an overview on the progress in visual prostheses in the BCI field. These BCIs are often invasive and opposed to only reading brain activity, they can also stimulate the brain and other parts of the body. Through this stimulation or by other means, the user can regain some form of vision from these BCI systems. Second Sight is one of few companies that has commercially made visual prostheses with approval from the Food and Drug Administration (FDA). Their system is discussed in the overview on BCI-related vision restoration systems by Niketeghad and Pouratian (2019). The international trial by Humayun et al. (2012) on the products of Second Sight shows promising results, although it is noted the study is performed by Second Sight employees and not by an independent research team. The exact working of visual prostheses or, more specifically, Second Sight products is not of interest for this work, but the recent decisions of Second Sight company reveal one of the largest risks of invasive BCIs and BCIs in general. Due to the discontinuation of some of the Second Sight products, hundreds of users are left without product support for a system that shaped their everyday life. Besides this, the now non-functioning product is still present inside their body. The issues and ethical questions this brings to the table are discussed further in Section 1.5.2.

In summary, hybrid systems like the one by B. Xu et al. (2022) are very promising as they can greatly reduce the risk involved in many BCI systems, such as prostheses-related applications, whilst also increasing the overall accuracy of the system. Section 1.3.3 discusses more of these hybrid systems but where BCIs are used as supplementary component to an existing system.

1.3.3 Using BCIs to improve the working of existing systems

The previous sections discussed the use of BCIs as a control system for prostheses and exoskeletons. Whilst this approach was promising for certain patients, the high risk associated with such systems meant that developing reliable systems relying solely on EEG based BCIs was not yet possible (AL-Quraishi et al., 2018). One area that did see success was the use of *hybrid BCIs* as a control system for robotic arms. Such hybrid systems combines BCI techniques with other

techniques to create an overall system that is more accurate, safe and reliable and thus complies more with Wolpaw's definition of a perfect BCI.

The discussed hybrid systems from Section 1.3.2 considered the BCI as the main component with the other techniques, such as computer vision, added as support and improvement to it. There are also multiple examples of where a BCI system is added as an *extra component* to an existing system and thus not considered the main component. This allows for BCIs to find use cases in non-trivial domains. To demonstrate this, an extension to classical hearing aids using an EEG based non-invasive BCI is discussed in this section. This section also addresses the use of BCIs to make *smart homes* more accessible for people with a handicap and the elderly. It also addresses some other examples of hybrid BCIs.

Extending hearing aid systems with BCIs

According to Seol et al. (2020) over 450 million people suffer from disabling hearing loss. Most solutions to hearing loss rely on a microphone to capture environment audio which is then amplified and played through a speaker that is placed in or near the ear. Most commonly, the microphones used for amplification are integrated inside the speakers that are placed in the ear to form a stereo setup that mimics regular hearing. This is not always ideal when there is a lot of ambient noise. Sometimes using an external directional microphone placed closer to the audio source of interest can form a solution. For example, placing a microphone on the desk of a professor teaching in a noisy room of students. However, this solution is not applicable in all situations. Thus, most hearing aids include some noise suppression on the microphones directly to filter out ambient noise and amplify noise coming from human speech. Wong et al. (2018) evaluated such noise suppression for Mandarin-speaking users and found the results to be good but not ideal, as there is a lot of variation in human speech tone making it hard to detect what is and what isn't ambient noise.

Even when such a noise suppressing filter would work optimally, people with hearing aids often still have trouble understanding people when multiple speakers are close-by at once. Recent research by da Silva Souto et al. (2016) has shown that using a non-invasive EEG based BCI can improve traditional hearing aids in solving this problem. Da Silva Souto et al. (2016) does this by determining which speaker a user is listening to by analysing directional queues from the measured brain-signals. If this information would be communicated to the hearing aids, it can allow them to optimize the microphones to pick up speech from that area only, filtering out other speakers. Whilst great in theory, da Silva Souto et al. (2016) discuss how a long waiting time to determine the area of interest challenges the practical usability of their system as of now. Nonetheless, it shows one of many non-trivial ways a BCI could be used as an extension of existing systems to improve them.

Using BCIs to improve accessibility of existing applications

There are still a lot of applications which lack support for people with disabilities or people with less technical skills. Take for example the growing ecosystem of smart home applications, since it aims to automate many tasks such people might find difficult or impossible to perform, it can be very beneficial to these people. However, conventional interaction methods are app-like through touch screens mounted all around the house. Such interaction methods are not fit for people with limited movement capabilities or technical knowledge.

Since EEG based BCIs are known to work well for conscious people that can learn how to effectively do motor imagery (MI) tasks, it is often seen as a solution to provide novel interaction method with existing applications to improve accessibility. W. T. Lee et al. (2013) were one of the first to explore the idea of using a BCI for smart home control as they believed it could

greatly improve the quality of life for the elderly and people with a handicap. They used the Emotiv EPOC headset as a non-invasive source for the EEG using dry electrodes and achieved high accuracy results through a primitive binary selection interface. This means that it seems to comply with most properties of Wolpaw's definition for a perfect BCI except for one of the most important ones, it is not at all as fast as the regular interaction method. Adding to this, their system doesn't make use of actual brain signals but rather relies on muscular activity, such as smirks and eyebrow movement. This movement of the facial muscles is known to provide an easily detectable artefact in the EEG and is often used as a supplementary input for the BCI system. As section 2.4.6 will discuss in greater detail, such an approach has many negatives. Finally, the system of W. T. Lee et al. (2013) was tested on four subjects but the article fails to deliver exact details on this experiment and it is assumed the participants were neither elderly nor people with a handicap.

Whilst this means the system by W. T. Lee et al. (2013) isn't a true brain-computer interface (BCI) due to the use of EEG artefacts and the paper has questionable scientific value, it was one of the first to explore the idea of using a BCI for smart home control. Since then, more articles have studied this idea using effective brain signals. Kosmyna et al. (2016) discuss this in their feasibility study about using BCIs for smart home control. The article by Kosmyna et al. (2016) does include clear details on the experiment setup which includes two people with a handicap among 12 other healthy subjects. It was found that the people with a handicap had an 81% average accuracy whilst the healthy subjects only had a 77% average accuracy. Due to the low sample size, this difference in results is likely not statistically significant but these accuracy numbers are usable when taking into account the small risk of failed actions in basic smart home control systems.

Other examples of hybrid BCIs

Section 1.3.2 already discussed the use of a hybrid BCI system where computer vision, eye tracking and other technologies were used in combination with a BCI system to create a more reliable and accurate final system for robotic arm control (as proposed by B. Xu et al., 2022). With existing algorithms such as the real-time grasp detection algorithm, GraspNet, by Asif et al. (2018) for low-powered devices, it is very easy to envision how they can be used to improve a BCI related system. Articles which propose a fully working POC, such as the ones reviewed by Dillen et al. (2022), often require these additional algorithms to ensure safe working in real-world environments.

One very interesting hybrid solution is the use of a non-invasive EEG based BCI system with measurements from non-invasive EMG. Bakshi et al. (2018) proposed such a hybrid EEG-EMG system for upper limb prosthesis control. Since EMG is known to be a more accurate modality for prosthesis control, the system by Bakshi et al. (2018) uses EMG for movement of the prosthesis where possible and supplements it with EEG data for determining the other wanted movements. In particular, the proposed prosthesis by Bakshi et al. (2018) uses EMG present in the remaining upper arm muscles for elbow motion and EEG for determining the desired wrist, grip and finger motions. The EMG part, which also imposes the greatest risk, had an accuracy of over 90% whilst the EEG part has an accuracy of over 65%.

The combination of EEG and EMG can also be used for other purposes. Dillen (2018) proposed the use of both EMG and EEG data for training a classifier which then relies solely on EEG input for making final classification predictions. Since the number of combinations to create hybrid BCI systems are nearly endless, it is left to the reader's imagination to come up with other potential use cases.

1.3.4 Using BCIs as an alternative for eye tracking

BCI systems relying on P300 signals where already briefly mentioned in Section 1.2.3. As further discussed in section 2.4.4, a P300 signal is a positive bio-electrical wave measurable with EEG around 300ms after a specific stimulus occurred. This specific stimulus consists of a rare and contrasting stimulus when the user was focusing on what is otherwise a relatively static object with a frequent stimulus (Halder et al., 2018). The most famous example is the use of a computer screen showing elements in a grid-like pattern as the static object with a frequent stimulus and flashing one specific element as the rare and contrasting stimulus. If the element the user was focused on flashes, the P300 signal will be present and relatively easily detectable in non-invasive EEG measures. A clever design of the interface can enable a wide variety of BCI applications. This section will highlight a few of these applications. It is noted that other types of stimulus can be used, such as auditory ones, to evoke a P300 signal. However, this section focuses on the discussed, visual, grid-based methodology.

With such visual P300 based BCI relying on the before-mentioned method of focusing on a part of the screen and using flashing patterns to recognise which part the user is focused on, a viable alternative would be eye-tracking technology. This makes the effective use of these types of BCI systems debatable, an issue that is also further discussed in this section.

Why a low learning curve makes P300-based BCI systems attractive

In general, BCIs using the P300 signal are often used as they have a low learning curve and there is a relatively low variation in performance between users compared to other types of BCIs (Hussein et al., 2020; Won et al., 2019). The reason these systems have a low learning curve is due to their simple user interface and a combination of a system that generalises well and that has been studied thoroughly for the use of transfer learning (TL). Early examples of using TL for P300 related BCIs include those by Kindermans et al. (2012). Adding to this, P300 signals can be detected with non-invasive EEG using dry electrodes, which are available in multiple EEG measuring headsets at affordable prices in a comfortable and visually pleasing package, as further discussed in Section 2.3.3. Whilst these factors hint that such BCI systems comply with Wolpaw's definition for a perfect BCI addressed in Section 1.1, it will become clear in this section that the communication rate of these systems is still a limiting factor.

However, these aspects do make it possible for P300-based BCIs to be rapidly configured and used by a new user. All of this aids in creating a pleasant user experience (UX) even if the communication rate is low, as shown by user studies such as the one by Utsumi et al. (2018). Other types of BCI systems, such as those relying on Motor imagery (MI), can have a tedious training procedure in advance. This can cause psychological burden and other side effects for the user. This gives rise to multiple ethical challenges, some of which are discussed in Section 1.5. Considering these things, P300 based BCIs are an interesting choice as a *first BCI* to introduce the user to the possibilities of BCI systems. This might make it easier for the user to move to a more capable and sophisticated system that has a steeper learning curve, higher cost and more demanding training.

Research on the UX is often overlooked by initial BCI system proposals, where the focus is often on objective numerical measures such as speed, accuracy, sensitivity and specificity. However, as discussed by Dillen (2018), user studies are essential to even consider using a proposed BCI system in the real world.

Using BCIs for P300 spellers

By far the most common usage of the P300 signal is to create *p300 spellers*, an idea first described by Farwell and Donchin (1988). In its simplest form, P300 spellers simply show the alphabet as grid elements and let the user spell words letter-by-letter. More complex forms can use auto-correction to correct faulty classifications, reducing the effective error of the system. Advanced text prediction can also be used to show complete words or sentences for selection, increasing the communication rate as opposed to letter-by-letter input. Other techniques can also further increase the reliability and communication rate of the system.

As such, P300 spellers are examples of novel interaction methods that aim to replace keyboards for those who don't have the required capabilities to operate them. Especially patients with severely limited motor skills and communication capabilities such as people suffering from serious cases of amyotrophic lateral sclerosis (ALS) or locked-in syndrome (LIS) can benefit from these systems by regaining basic communication skills (Hussein et al., 2020; Won et al., 2019). Whilst wrong classifications in a speller application could result in unpleasant situations, it is clear that the risk involved is far smaller than with BCI system used for controlling prostheses and exoskeletons such as discussed in Section 1.3.2. This makes P300 spellers one of the few types of BCI systems that are actively being used in the real world.

Guy et al. (2018) performed a usability study of P300 spellers with 20 ALS patients in a real-life environment. According to Guy et al. (2018), most participants achieved over 70% accuracy, which is in line with the findings of Hussein et al. (2020) and Utsumi et al. (2018) in similar studies amongst other types of patients. More interestingly, even though the accuracy wasn't extremely high, all participants of the experiment by Guy et al. (2018) succeeded in the given tasks. This is in part due to our ability as a human to understand typo's in words and sentences relatively easily. Another important factor is that the studied system makes use of the earlier discussed auto-correction and text-prediction techniques. Whilst far from advanced variants of these techniques were used, it almost doubled the communication rate of the P300 speller already with the mean number of correct symbols per minute going from 3.6 to just over 5.

Whilst 5 symbols per minute is significantly slower compared to traditional communication skills, it enables useful communication for those who can't communicate through regular means. It should also be considered that these results are for people suffering from ALS. These people can have difficulties with controlled eye movement or rapid eye movement which results in a slower operation speed compared to able-bodied persons. However, since it is so slow compared to the regular interaction method of able-bodied persons, it is far from a perfect BCI when taking into account Wolpaw's definition of a perfect BCI. A similar study by Utsumi et al. (2018) focuses more on the UX for duchenne muscular dystrophy (DMD) patients using P300 spellers. This study has shown that P300 spellers can be used by DMD patients with satisfactory results in a pleasant manner for the user, even if the communication rate is slow.

Using the P300 signal for other BCI applications

When taking a broader look at P300-based BCI systems using a visual grid system, they are just a mechanism for recognising which portion of the screen a user is looking at. These grids are often of a limited size, such as 6 by 6, for the best trade-off between accuracy and communication rate. The effective content of the grid and the actions taken upon classification can vary widely. This has enabled the exact technology of P300 spellers to be easily adapted for other applications.

One such example is the *Facebrain* application by Warren and Randolph (2019). Facebrain provides a non-invasive EEG based BCI for interacting with the social media platform Facebook. In essence, it's a regular P300 speller with the first screen(s) representing possible actions to take on the platform. When text input is required, a regular P300 speller user interface is presented.

This allows a user to operate almost all of Facebook's functionalities with only a P300-based BCI. The application by Warren and Randolph (2019) is one of many that shows the same strategy and classification algorithms as P300 spellers can be used for a wide variety of applications by changing the meaning and functionalities of the shown grid elements. Rather than showing symbols, a complete image could be shown and separated into grid elements as well. Kapgate (2022) used this idea by showing a live video stream from a quad-copter and moving it towards where the user is looking based on the P300 signal after flashing different portions of the image. However, as the video live stream is not static, the contrast between the flash and regular screen content shrinks making the P300 signal less noticeable and thus the classification task harder.

Why eye tracking may still be preferred

Whilst the discussed P300 applications have shown success in achieving their goals, the effective real-world use of such systems can still be debated. The communication rate is slow and for able-bodied users, traditional eye-tracking can far outperform these systems (Halder et al., 2018). Adding to this, as P300 signals are a type of event-related potential (ERP), there are some known limitations and issues with techniques relying on them as well. The BCI handbook by Nam et al. (2018, Chapter 26) discusses the crowding effect, adjacency problem, repetition blindness and user discomfort amongst other issues ERPs have. Most of these problems arise from the often limited space for sending visual stimuli without overlap and the changing behaviour of both the brain and participant's experience after a prolonged session where many stimuli have been applied.

Multiple articles have been written on when and why a P300-based BCI might be preferred over a regular eye-tracking system (Halder et al., 2018; Pasqualotto et al., 2015). The experiments performed by Halder et al. (2018) compared the accuracy and communication rate of visual P300 based BCIs against traditional eye-tracking solutions. Halder et al. (2018) found the accuracy of both systems to be comparable but the communication rate of eye tracking solutions was around 50% faster. However, when questioning the ten participants of the study, it was found that there is a higher preference for the BCI system than the eye-tracker. Whilst favourable for the BCI alternative, the results of the study by Halder et al. (2018) should be interpreted with caution. The participants of their study were all able-bodied persons and the eye-tracking experiments were always performed last. This and other aspects of their methodology are bound to introduce bias in the results.

A study by Käthner et al. (2015) compared the use of eye-tracking software and auditory P300 based BCIs for a patient suffering form the locked-in syndrome (LIS). Whilst the patient was able to work with both systems autonomously, there was a strong preference for the traditional communication strategy which relied on a human to determine the wanted communication of the patient. This preference was purely based on the UX rather than the communication rate. Whilst only being an experiment based on a singular participant, it does show how automated systems for regaining basic communication might cause a negative psychological impact on the user. Thus, the results found by Halder et al. (2018) on UX are in direct contrast with the results from the experiment by Käthner et al. (2015). However, Käthner et al. (2015) lacked a study of visual P300 based BCIs, so it could be that such a system would be preferred.

Similar conflicting results are present on other aspects of visual P300 based BCIs as well. One debate that has been active for years is the dependence of P300-based BCIs on the capability of a user to look directly to a single target for a prolonged time. As discussed by Brunner et al. (2010), it has been argued that this dependence is low as peripheral vision allows the user to focus on the desired element even if the eyes aren't directly rotated towards the element. This would be favourable for P300 solutions compared to eye-tracking solutions as some users might not have

the required muscle control to point their eyes to a specific point for prolonged periods, a task known as eye gazing. However, the study by Brunner et al. (2010) does show a dependence of P300-based systems on eye gaze and they argue the effective uses of visual P300 systems should be reconsidered. Yet, the review article by Riccio et al. (2012) addresses multiple articles which concluded through real-world tests that being capable of eye gazing is not necessary for effective use of P300-based BCIs.

These contradicting results on the use of P300-based BCI applications make it hard to determine which kind of system would be usable in the real world, providing a net benefit for the user. It also shows that promising objective results such as a high accuracy and communication rate are not guaranteed to provide a system which is enjoyed by the end user. Thus, many proposed systems require follow-up articles focusing on performing in-depth real-world experiments with the target population to conclude if they are worthwhile. However, doing such a study in the same article proposing the new novel system would mean that one singular article takes an incredible amount of time. This is contradictory to the expectations of many researchers in the field to publish a certain amount of articles yearly. However, the latter is an ethical dilemma that entails a higher risk of *sloppy science*.

1.3.5 Summary on the use cases for BCIs

Since BCIs can be seen as a novel interaction method for controlling external devices, the possible use-cases are endless. Even for the discussed P300 based BCI systems alone, a huge variety of applications can be created as addressed in Section 1.3.4. This has caused review articles to focus solely on these types of BCI systems or even subset of those types of systems (Alrumiah, 2020; Fazel-Rezai et al., 2012; Rezeika et al., 2018).

Thus, addressing all possible use cases for BCIs is far outside the scope of this thesis. For this section, the goal was for the reader to obtain a more general insight into the most studied uses for BCIs. In this regard, it succeeded in addressing all of the most popular use cases of BCIs as biosignal control systems in the real world as found by Dillen et al. (2022) and the most interesting use cases of BCIs addressed in the review article by Abdulkader et al. (2015). It also highlighted that current real-world applications still rely on other technologies to limit risk and improve accuracy as standalone non-invasive EEG based BCIs are still not accurate and reliable enough for most real-world usage. Some of the main positive and negative points of the discussed articles were also addressed.

Most notable from this section is the fact that objective measures are only a small part of making a system that can be used in the real world. Take for example the discussed experiment by Guy et al. (2018) which studied the use of P300 spellers by patients suffering from ALS. With a classification accuracy of 70% and only 5 symbols per minute, the objective measures from this system look terrible compared to able-bodied alternatives such as regular keyboard typing. However, Guy et al. (2018) found the users to have enjoyed the use of the P300 spellers and when taking into account the options available to the target users rather than able-bodied users, the objective measures are far more impressive.

However, even the articles which perform experiments that take into account the user experience (UX) can sometimes cause conflicting results. Take for example the discussed differences between the found user experience of P300 based BCIs between the experiment by Halder et al. (2018) and Käthner et al. (2015). Able-bodied persons from the experiment by Halder et al. (2018) mentioned a pleasant UX. A more representative real-world study by Käthner et al. (2015), who looked at a singular patient suffering from LIS showed that this person had no interest in using such systems as a replacement for the existing human-based communication system which was already present. Even-though speed and autonomous use were far better and

the UX according to the able-bodied persons was great, the UX for this person suffering from LIS was not good enough. Thus, the UX is something that can only truly be measured when performing the experiments for an extended period with the target audience of the system, not through a single experiment on able-bodied participants. It is hard for an able-bodied person to understand that the confrontational aspect of BCIs on the limitation a user has can outweigh the benefits it seemingly offers. This and other more ethical questions are also addressed in Section 1.5.

To summarize, as a BCI performs actions based on brain signals, a very fitting analogy could be made: *a BCI could do anything you can think of.* That is, of course, only true in theory. Due to limited knowledge of the brain (Hodson, 2019) and limitations in what can be measured, a BCI can only truly do what you can think of in a *measurable way.* Still, this is a whole lot of applications and an overview diagram of the most important ones is given in Figure 1.7. This diagram is not only based on the use cases discussed in this thesis but also those discussed in the articles by Dillen et al. (2022) and Abdulkader et al. (2015) among others (Kübler, 2020; Nam et al., 2018; Nicolas-Alonso & Gomez-Gil, 2012; Panoulas et al., 2010; Shih et al., 2012; Sonam, 2018). From Figure 1.7, it should become apparent that most short-term goals of BCIs still lie in improving the quality of life for people with disabilities. However, the rise in popularity of BCIs in the gaming industry and amongst other big tech companies, as discussed in Section 1.2.1, shows there is a potential future where BCIs find more real-life use-cases in other fields as well.



Figure 1.7: An overview of some of the many use cases for BCIs.

1.4 Opportunities and obstacles for BCI research

The most likely factors of why BCIs are gaining popularity and what the most common use cases are for BCIs were already discussed in Section 1.2 and 1.3 respectively. From these sections, an interested reader could have already spotted some opportunities and potential obstacles for doing research in the field themselves. This section discusses some opportunities and obstacles present according to the author of this paper. The aspects that are covered in this section are oriented towards the interests of the author from this master thesis and aim to explain why this master thesis came to fruition. They also lay the foundation to the proposed BCI biosignal control system of this master thesis, which will be further discussed in Section 1.6.

1.4.1 Seemingly small projects with huge impact

As BCIs and other technologies become more commercially oriented, the main focus is often shifted to providing products that can be used by the masses to provide the highest possible revenue. Whilst BCI research sees its foundation from the medical world and most of the current users are people with a certain medical condition, a similar shift is likely, especially for the larger projects funded by big-tech companies. This isn't necessarily a bad thing for medical applications, since this doesn't mean academic research focusing on those people with a certain medical condition disappears, rather it becomes a smaller portion of all research in the field. The new technology and knowledge commercial products deliver can directly contribute to academic research focusing on medical applications. This may include cheaper and more capable measuring equipment, a more socially acceptable picture of BCIs aiding the aesthetic aspect of Wolpaw's definition for a perfect BCI and better performing classification algorithms. It is also possible that a commercial BCI application which is meant as a novelty for most, is life-changing to some.

Take the evolution of live captions as a recent example from a different domain, it has found its way directly into the Android smartphone operating system. Whilst initially implemented for users to enjoy video content in situations where they can't listen to the audio, people with hearing difficulties now have a direct way to enjoy more content too. The push for autonomous cars will allow those that are traditionally unable to drive a car to finally enjoy the freedom a car can offer as well without being dependent on others to do so. Object and scene recognition algorithms made for optimizing smartphone cameras can also be used as a way of describing what is on a photograph for those who have limited vision.

Likewise, as more people are exposed to complex medical and commercial applications, many individuals start to envision seemingly small applications that could have a major impact on them or people they know. This has been the case with smartphone applications for a long time. A colour picker app that uses the camera to describe the colour of the central pixel is something that can be implemented with very limited resources. Yet, such an application has already proven to significantly aid people who have colour blindness. For example, they may use this application to determine whether a banana is ripe or not, something they can't visually determine but is easy with a textual description of the colour. For such small applications to be feasible, an abundance of open-source material to re-use existing code for operating the camera, getting a textual colour description of an RGB colour code and more is required.

This presence of open-source code has started to emerge for the BCI field as well, as discussed in Section 1.2.3. Combined with publicly available EEG dataset such as the one by Kaya et al. (2018), it has become possible for developers to work on a pipeline without needing the monetary investment in EEG measuring equipment. If desired, the affordable headsets by OpenBCI and others, further compared in Section 2.3.3, make it possible to have EEG measuring equipment with good open-source libraries for well under 1000 euros. This has made individuals such as

the author of this paper excited to create a wide variety of smaller applications that could have impactful meaning to some very specific people. Especially the possibility of regaining even the simplest form of mobility through some very basic controls derived from motor imagery (MI) is what motivated the author of this master thesis to study the field.

1.4.2 Motivational aspects for this master thesis

The discussed use cases from Section 1.3 were motivating examples for this master thesis in themselves. However, there are some specific articles and aspects from the field that have motivated the creation of this master thesis which will be discussed in detail here. These are mainly related to motor imagery (MI) based BCIs working with affordable consumer-grade and non-invasive EEG measuring equipment. Whilst many articles have demonstrated the potential of such BCI systems, the lack of widespread adoption suggests there are still hurdles to overcome for them to become a reality. When discussing these motivating examples, some of these hurdles will already be discussed, the main obstacles for BCI research are further discussed in later subsections.

Consumer-grade EEG measuring equipment has promosing potential

Section 1.2.2 and 1.2.3 have already discussed the evolution of EEG measuring equipment as a potential reason for the increased popularity of BCI research. Multiple studies comparing different aspects of cheaper consumer-grade systems to the more traditional medical-grade BCI systems have already been done (Cruz-Garza et al., 2017; Frey, 2016; Hinrichs et al., 2020; Rashid et al., 2018; Tseghai et al., 2021). In general, cheaper systems work with dry electrodes at a lower electrode count and have a lower signal quality compared to medical systems. This means the cheaper systems provide poorer data quality, resulting in a more difficult classification problem.

This raises the question if useful applications can be made using data from these cheaper systems. One of the best-known providers of EEG equipment is OpenBCI and their products are further discussed in Section 2.3.3. This OpenBCI headset makes use of the Texas Instrument ADS1299 chip to convert the analogue signal of EEG electrodes to a digital one. Whilst affordable and widely available, it has been shown that this chip is comparable to far more expensive alternatives that have been used in the laboratory for years (Frey, 2016; Rashid et al., 2018). However, a cheap consumer-grade system differs from medical systems in more areas than the general chip responsible for EEG amplification and digitalisation.

Studies comparing the use of cheaper dry electors against medical wet electrodes have already been discussed in Section 1.2.2. Whilst the obtained results from these studies show the gap between both is shrinking but still present (Cruz-Garza et al., 2017; Mathewson et al., 2017; Tseghai et al., 2021). However, the experiments used for obtaining these results come from very controlled experiments in lab-like environments. This is to be expected, as most comparisons aim to eliminate as many random factors as possible. However, this means that these experiments don't give great insight into the real-life usability and applicability of these cheaper BCI systems. Taking into account that none of the academic research discussed in the systematic review article by Dillen et al. (2022) makes use of such cheap systems hints that real-world usage of these systems remains hard.

Trainable motor imagery as input thoughts for the BCI system

Section 1.3.4 discussed P300 based BCI systems. Whilst the P300 signal is relatively easy to detect and evoke, the discussed visual stimuli used, have many limiting factors. Perhaps most limiting is that the user can not evoke the signal without requiring external stimuli. This is

undesired in several ways. This external stimulus requires an additional system to be present, such as the flashing grid commonly used for visual P300 BCI, which adds to the cost and complexity of the system and limits portability. A system where a person sits behind a flashing screen is also hard to call aesthetically acceptable, making it unsuitable for Wolpaw's definition of a perfect BCI.

Detectable signals that can be evoked voluntarily by the user are more interesting for most BCI applications. Likewise, the process to evoke this detectable signal should be one that doesn't require physical movement so that it is applicable for users suffering from limited muscle control. It should ideally also be a process that is invisible to those around the user, making it aesthetically ideal for Wolpaw's definition of a perfect BCI. Motor imagery (MI) is such a cognitive process and is often used for controlling BCIs and in other fields such as sports, psychology, music, medicine and education (Schuster et al., 2011). Mulder (2007) defines the MI process as follows:

> Motor imagery is a cognitive process in which a subject imagines that (s)he performs a movement without actually performing the movement and without even tensing the muscles. It is a dynamic state during which the representation of a specific motor action is internally activated without any motor output. In other words, motor imagery requires the conscious activation of brain regions that are also involved in movement preparation and execution, accompanied by a voluntary inhibition of the actual movement.
>
> *Mulder (2007)*

MI does have some drawbacks as well, mostly following from MI being a cognitive process with no visual clues. Firstly, explaining how to *do MI correctly* is a difficult task. Whilst it can be trained, it has a far steeper learning curve to obtain pleasing results than the earlier mentioned P300 systems relying on ERP-based signals for example. Alimardani et al. (2018) describes this problem and tips for teaching MI in greater detail. The most general procedure of explaining the MI task to a user consists of verbally explaining the MI task. This can be further supported by a physical example of the task that should be envisioned. People with the capabilities of performing the envisioned task physically can also be asked to perform the task whilst simultaneously thinking about it as a first step too. After this, a training procedure of the user performing the MI task without physical movement starts.

This introduces the second problem, evaluating the users *capability of doing MI*. Such an evaluation exists of two parts, a survey taken beforehand and feedback during or after training. Multiple types of surveys have been proposed to determine beforehand if a person will be good in MI tasks (Gregg et al., 2010; Malouin et al., 2007; McAvinue & Robertson, 2008; Vuckovic, 2010; Vuckovic & Osuagwu, 2013). Most of these surveys are empirically created and based on the found correlations between the answers given by participants and their performance on a MI BCI for those participants. However, correlation does not mean causation, and it has been the case that these questionnaires do poorly at predicting a survey respondent's capability of doing MI in a detectable manner. For example, Peterson et al. (2020) found no statistically significant correlation between the KVIQ-10 score of participants and the found classification accuracy. Thus, determining beforehand if a participant will have pleasant accuracy results beforehand through the KVIQ-10 questionnaire by Malouin et al. (2007) wasn't reliable for the experiment by Peterson et al. (2020). This is an issue, as knowing this information beforehand can give a potential buyer a better indication if the system would be fit for them or not. Other surveys have been proven more successful at predicting MI capability of a user but further research in finding a survey that is reliable at predicting a user's MI skills is still required (McAvinue & Robertson, 2008). Besides this survey that is taken beforehand, evaluating the MI tasks performed by a user is also not an easy task. Most of the time, a model is trained or calibrated after obtaining

training and test samples of the user performing the MI tasks. The accuracy of this trained model can then be used as an indication of how good the user is at the MI task. Alternatively, the user might be exposed to live feedback during data collection in the form of a visual or physical stimulus that indicates the BCI is detecting a specific MI task. Including live feedback has been proven highly beneficial in training MI, although it makes the training procedure even longer (Alimardani et al., 2018). As evaluation in this way is mainly empirical based on the obtained classification result, it is not uncommon to see varying ways of doing the MI task between users. Some might perform the MI tasks by envisioning the task from a third-person perspective whilst others opt for a first-person perspective. This introduces many variables in the data, causing a high variability between users and even between sessions of the same user. Visualisation of the brain signals and their decoding might aid in guiding users to perform a more equal MI task, reducing the variability of the data. However, since brain signals are non-stationary and the visualisation techniques are limited, variability will always remain an issue. Forcing a user to perform the task in a specific manner also has downsides. Their personal accuracy for the BCI system may be poorer when forced to do the MI task in a very specific way compared to finding the optimal method for them. The psychological burden of the long MI training process will also be higher if the task is very strict, requiring more focus. Finally, the data may be so strictly obtained that it is far from realistic and the system performs well in the real world.

This highlights the third and final important issue with MI, the issue of generalisation. This issue is present in two different forms. First, there is the general issue not strictly limited to MI that training and testing data is often obtained in strict manners and thus lab-like. When used in the real world, more noise is present in the signal resulting in far poorer results. The trained system does not generalize well to new, unseen data from the real world. Second, as discussed the brain signals produced during MI tasks can vary greatly between participants and sessions. This variability means that creating a general model is far less successful, the trained model does not generalize well to other users. This makes transfer learning (TL) far more difficult. This issue cascades to making the training procedure longer and having higher inter-patient variability in terms of accuracy results (Alimardani et al., 2018; Peterson et al., 2020). The generalisation issue is an important one in many ML applications.

These issues with MI result in datasets that seem comparable from a high-level description of the performed MI task but are very different in the used training procedure and data acquisition process. The general MI capabilities of the participants can also differ greatly between datasets, with some performing a survey beforehand to only include participants with a high possibility of being good at MI. This means that choosing datasets from *well capable MI performers* will result in higher accuracy scores as the data is easier to learn from. Even when open-source databases are re-used between articles, it is not uncommon to find articles where the authors left out certain data in the training and testing procedure, arguing these users had poor MI capability. This makes a comparison of results between articles far more difficult. This is an issue of not having standardized testing, which will be discussed in Section 1.4.4.

Detailed literature on MI pipelines

Section 1.2.4 discussed the emergence of more open-source datasets and code-providing libraries for BCI research. However, many articles still fail to provide their source code and lack the required detail to fully recreate the used pipeline. This makes articles describing their pipeline and design decisions in great detail stand out. In general, these articles don't aim to improve the state-of-the-art in any component of the pipeline. Rather, they contribute to the field by providing a detailed look at the working of pipelines based on available state-of-the-art solutions and a correct evaluation of them.

The article by Peterson et al. (2020) discussing the feasibility of a complete low-cost consumer-grade BCI system is one that focuses on these aspects. Peterson et al. (2020) performed their feasibility study by discussing the steps required to make an offline binary MI classification system using common low-cost consumer-grade hardware. The BCI distinguished two cases, the MI task of a grasping movement with the participants' dominant hand and a rest condition. They compare three traditional ML approaches for this classification task in growing complexity. These approaches differ in the feature extraction used, namely the use of traditional CSP and that of two extension, penalized frequency band common spatial pattern (PFBCSP) and penalized time-frequency band common spatial pattern (PTFBCSP). The work by Peterson et al. (2020) has five interesting aspects worth highlighting here.

First, whilst they did use the consumer-grade OpenBCI Cyton and Daisy board they did not use the 3D printable Ultracortex Mark IV headset from OpenBCI. They argued that this is due to the Ultracortex Mark IV headset becoming uncomfortable quickly due to the use of dry electrodes combined with limited adjustability. This complaint on user comfort for the Ultracortex Mark IV headset is recurring with other authors including the one from this master thesis. Because of this, they opted for wet EEG electrodes attached to a very flexible and far more comfortable Electro-Cap. Peterson et al. (2020) opting for wet electrodes is slightly odd, as it is not really consumer-grade nor does it fit well with Wolpaw's definition of a perfect BCI.

Secondly, for the data gathering of their system, Peterson et al. (2020) used a *common office room*, rather than a lab-like environment. Except for a 3D printed holder for the OpenBCI board, there was no specialized shielding in place to protect the electrodes or OpenBCI boards from unwanted interference. They argue this makes the data more realistic. Whilst this is true to a certain extent, it is important to note there is still far less stochasticity than there would be in real life. The office room and hardware were identical for each of the participants in the data collection stage. The room was free of external stimuli with the participant *left alone* to focus on the task and the task only during the entire trial. All participants were right-handed and thus the dominant hand used for the MI task was always the right hand. Adding to all of this, the dominant hand of the participant was placed inside a box to not allow them to see their hand. This makes the data acquisition procedure used by Peterson et al. (2020) far from comparable to a real-world use case. Still, Peterson et al. (2020) found that the non-shielded regular office already caused significant noise compared to full lab settings. In one trial the electromagnetic noise amplitude was four times higher than the meaningful EEG data.

Third, during the data collection, an EMG system was also in place. This EMG system was used to filter out samples of the collected EEG data where the envisioned movement was also physically performed, meaning it wasn't true MI. Whilst this forms an interesting and automated approach to data filtering when collecting training samples, the supplemental hardware that is only used during the training phase is probably a tough sell in a commercial application.

Fourth, they used the KVIQ-10 questionnaire by Malouin et al. (2007) to determine how good a participant would be in MI, a task that is proven to be harder for some individuals. As discussed earlier, the results of these surveys are not reliable and only give a rough indication with many exceptions and surprising results possible.

Finally, even though the data acquisition happened under guidance in the work of Peterson et al. (2020), there were still issues with the recordings. Out of the 12 participants, there were multiple moments where connection loss with the OpenBCI main board occurred and for one participant a defect rendered the data useless. These issues are probably related to the quality of the consumer-grade hardware. Part of the reason medical-grade hardware is about 5 to 10 times as expensive for a similar experiment is due to the strict certification medical-grade BCI should comply too. This certification guarantees some form of quality and reliability from which it is expected connection issues and defects wouldn't appear as frequently. Besides this, for one

participant there were EMG detected movements of the hand for more than half of the MI tasks rendering the trial of that patient useless as well. Whilst this is an issue independent of the hardware used, it does indicate that learning how to do MI requires training which takes time and effort.

To conclude, the work by Peterson et al. (2020) discusses the creation of a complete low-cost consumer-grade BCI system. This system consists of the OpenBCI measuring equipment where the dry electrodes on the 3D printed Ultracortex Mark IV are replaced with electrodes in a more comfortable Electro-Cap. The effective classification of the system is a binary motor imagery (MI) classification on whether or not the participant imagines a grasping movement of the right, dominant hand or not. Peterson et al. (2020) achieved an average accuracy between 70% and 85%, scaling with the complexity of the used CSP variation. It is important to note that the evaluated models are on a patient-per-patient basis. This means that each patient has their own uniquely trained model and that data from the same patient is used in the evaluation process. Whilst the binary nature of the system makes it hard to find viable real-life applications, the performance reached is almost identical to those of medical-grade systems and follows from a slightly less lab-like environment than is typically the case. The system proposed by Peterson et al. (2020) is of less importance in their work, rather the steps and pitfalls highlighted are of value.

Detailed literature on complex classification pipelines

The above discussed article by Peterson et al. (2020) provides great insight on the steps required to develop an EEG-based consumer-grade BCI which uses MI related signals. Since Peterson et al. (2020) uses a binary classification model, there are only two possible outputs of the classifier, which is too limited for most applications. However, when working with BCIs, a lack of training samples combined with noisy and often high-dimensional data makes multi-class classification considerably harder than binary. That being said, spatial filters such as the CSP approach and its extensions used by Peterson et al. (2020) have been extended to support multi-class feature extraction. The articles by Abdeltawab and Ahmad (2020) and Olivas-Padilla and Chacon-Murguia (2019) have also studied the use of spatial pattern techniques for feature extraction. This is done in combination with traditional ML classifiers and DL ones, both with promising results.

Many other multi-class classification pipelines have been proposed in literature that work well with MI related EEG data (Abdeltawab & Ahmad, 2020; Z. Chen et al., 2021; Hou, Zhou, et al., 2020; Kai Keng Ang et al., 2008; Lawhern et al., 2016; Mane et al., 2021; Mussi et al., 2019; Olivas-Padilla & Chacon-Murguia, 2019; Schirrmeister et al., 2017). These proposed pipelines generally work on both consumer-grade and medical-grade systems, although consumer-grade systems can often benefit more from pipelines with specific noise-reduction steps. Some of the proposed pipelines also focus on providing a general model which has been trained on data from multiple users and has usable performance for unseen users. Whilst such models have poorer performance compared to one trained for a specific user, they can be used as an initial model to allow the user to explore the possibilities of the BCI without having to undergo the often tedious training data collection process. Such a general model can also be used as a base model for calibration, a process based on the idea of transfer learning (TL) further discussed in Section 5.3.3.

A complete in-depth review of all of the different approaches that have been proposed for EEG classification falls outside the scope of this research paper. Guerrero et al. (2021) compared logistic regression (LR), artificial neural network (ANN), support vector machines (SVM) and convolutional neural network (CNN) for a binary classification task of either being epileptic EEG

data or not. Whilst this is a binary classification task that is more tailored towards computer-aided diagnosis (CADx), the techniques used in the experiments are often used in the multi-class classification of EEG data for common BCI purposes. Guerrero et al. (2021) found that artificial neural networks (ANNs) performed best for their classification task. In general, ANNs and other DL models such as CNNs have proven to be more successful at EEG data related tasks compared to traditional ML approaches.

Because of this, many of the current state-of-the-art models for EEG classification rely on DL models. Especially classification pipelines that include CNNs have proven to be successful for EEG classification (Hou, Zhou, et al., 2020; Lawhern et al., 2016; Mane et al., 2021; Mussi et al., 2019; Olivas-Padilla & Chacon-Murguia, 2019; Schirrmeister et al., 2017).

The CNN-based approach by Schirrmeister et al. (2017) is commonly regarded as current state-of-the-art for MI classification. The article by Schirrmeister et al. (2017) includes two different models, a deep CNN named DeepConvNet and a shallower one named ShallowConvNet. In Chapter 4, both of these models will be implemented and discussed further. Schirrmeister et al. (2017) also describe a method of extracting a visualisation of the used brain signals by the model, which can aid in the explainability and interpretability of the model.

One issue with more complex CNN-based pipelines such as the DeepConvNet variant by Schirrmeister et al. (2017), is the time and computational resources it takes to train the model and do predictions with it. The latter is an issue for real-time classification, something that is needed for an online BCI system which often works with relatively low-powered computational units. Pipelines such as the one by Lawhern et al. (2016) have been developed to use CNNs in such a way that real-time classification is possible. The model proposed by Lawhern et al. (2016) also has promising results and will also be implemented in Chapter 4 where it is discussed in greater detail as well.

Whilst CNN-based pipelines are among the most popular for the classification of MI EEG data, they fail to use a fundamental aspect of EEG data. CNNs are not designed to perceive the input data as sequential data, which EEG data is. As such, they do not explicitely use this property of the data for learning. Recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) in particular are a type of DL networks that do explicitely use internal memory to explicitely use the sequential property for learning, as Section 3.3.4 will discuss in greater detail. These approaches are most popular in speech processing as it is the sequence in which words are spoken that gives meaning to a sentence rather than individual words. P. Wang et al. (2018) used a traditional LSTM pipeline for MI EEG data classification with satisfactory results whilst Garcia-Moreno et al. (2020) combined both ideas from CNNs and LSTMs into a singular network that rivals CNN-based state-of-the-art for MI EEG classification.

Some more noteworthy MI related EEG classification approaches include the one by Z. Chen et al. (2021) and the one by Hou, Zhou, et al. (2020). Z. Chen et al. (2021) took an interesting approach by first visualizing EEG data as an image and using techniques from image processing for classification. This yields decent results but doesn't reach the same level as the discussed state-of-the-art models. However, the approach by Hou, Zhou, et al. (2020) which incorporates the technique of scout EEG source imaging (ESI) has shown to be as good or even better than state-of-the-art in specific experiments.

Connecting the classification model to physical devices

As was discussed in Section 1.1, there is no exact definition of a BCI. In general, a complete BCI system is seen as a combination of three different processes: a data collection process, a data processing step and a step where effective actions are taken by the BCI based on the processed data. It is this last process that forms some ambiguity in what can be considered a BCI and

what can't be, as was already discussed in Section 1.3. BCIs that functions as biosignal control systems, which explicitely control an external device in real-time, are included in each definition of a BCI. However, how such an external device is controlled can vary greatly and requires special attention which may influence all other components of the BCI as well.

The labels provided by the classifier can be seen as an incoming input stream for the process responsible for mapping those classifications to an action on the external device. When using MI EEG, it is intuitive to link the envisioned task with a similar action on the external device. For example, an envisioned left-hand squeeze corresponds to a left movement whilst an envisioned right-hand squeeze corresponds to a right movement. However, nothing is stopping the implementer from doing this mapping differently and in some scenarios, a less intuitive mapping might provide a more reliable and capable system that is easier to use.

This follows from some of the challenges with MI EEG classification, in particular a limited amount of output classes and the fact that it is easier to distinguish between major MI tasks such as left-hand and right-hand movement rather than individual finger movement for example. Since classification accuracy should be sufficient to operate the external device safely and reliably, a limited number of different and easy to distinguish MI tasks are often used.

To illustrate how such limited classes can be used for effective biosignal control systems, imagine the movement of a robotic arm to pick up objects using MI EEG data through a classifier that can distinguish 4 MI tasks: left-hand squeeze, right-hand squeeze, foot movement and an idle state. Intuitively, one might want to obtain complete control over the robotic arm but the limited MI tasks render a direct mapping between the classification label and all possible movements of the arm impossible. The most intuitive solution is the creation of a menu for movement options as illustrated in Figure 1.8a. Using the foot MI task, the user switches between 3 operation modes of the robot arm: horizontal movement, vertical movement and grasping movement of the hand. Dependent on the active menu, the left-hand squeeze and right-hand squeeze MI tasks can be used to move the robot arm left and right, up and down or to open and shut the robot hand. This provides the user with all available control whilst remaining relatively intuitive. However, this assumes all commands are interpreted correctly. When taking into account that state-of-the-art struggles to reach an average of 80% accuracy for classifying four MI tasks in favourable conditions, it is not unreasonable to think that best case there will still be a misclassification once every five steps on average (Khan et al., 2020; Mishra et al., 2018; Rahman et al., 2019; Zhang et al., 2019). Taking into account that for picking up the desired object, the robot arm needs to be perfectly aligned both vertically and horizontally and the user needs to switch to the menu for grasping through envisioned foot movement and then chose the desired grasping movement through envisioned hand movement, it is highly unlikely this sequence of events will happen without misclassification. This long sequence will not only make the system inefficient to use, but it will also expose one user's intention to more points of failure as it will require many steps.

Stepping away from the idea of wanting to control every movement of the robotic arm, a huge improvement in the efficiency of the system can be made. With grasp detection algorithms such as the one proposed by Asif et al. (2018), the detection of objects of interest for the robot arm to interact with can be detected through computer-vision algorithms. Using such algorithms, another way of controlling the robotic arm could rely on only three MI tasks: left-hand squeeze, right-hand squeeze and an idle state. Using the left-hand squeeze MI task, the robot arm could be alternated between all possible items to grasp as detected by the grasp detection algorithm. Using the right-hand squeeze MI task, the item could be picked up or set down depending on the current state and computer-vision-aided tools. This limits the amount of successful consecutive steps needed and being a three class MI classification, the accuracy is expected to be higher. Thus, the system would be less intuitive but far more usable. This proposed design is shown in

Figure 1.8b. When comparing both subfigures of Figure 1.8, it becomes apparent that the less intuitive system is far more efficient at the given task, requiring only two actions to pick up the second glass rather than 13. It is noted this is only an educational example and real systems would likely still require a more complex interface. It is also noted that it is common to reduce misclassification by basing the taken actions on multiple classifications rather than one.

Cong Wang et al. (2011) use comparable reasoning to propose a MI-based BCI system to control a robot arm. Cong Wang et al. (2011) use only three distinct MI classifications: imagined right-hand movement, imagined left-hand movement and imagined foot movement. These three controls enable the user to select eight different possible actions through a menu where two options are always shown that can be controlled using either an imagined left-hand movement or an imagined right-hand movement. Scrolling through the menu to show two other possible actions is possible through the imagined foot movement. This shows that with the right system design, few controls can still allow for many actions to be taken. However, they lack a detection of the idle state. This is possible by determining the idle state based on the certainty of the model that a certain prediction is correct. If the model is uncertain, it is likely the user wasn't performing any of the MI tasks and thus the system should not do anything.

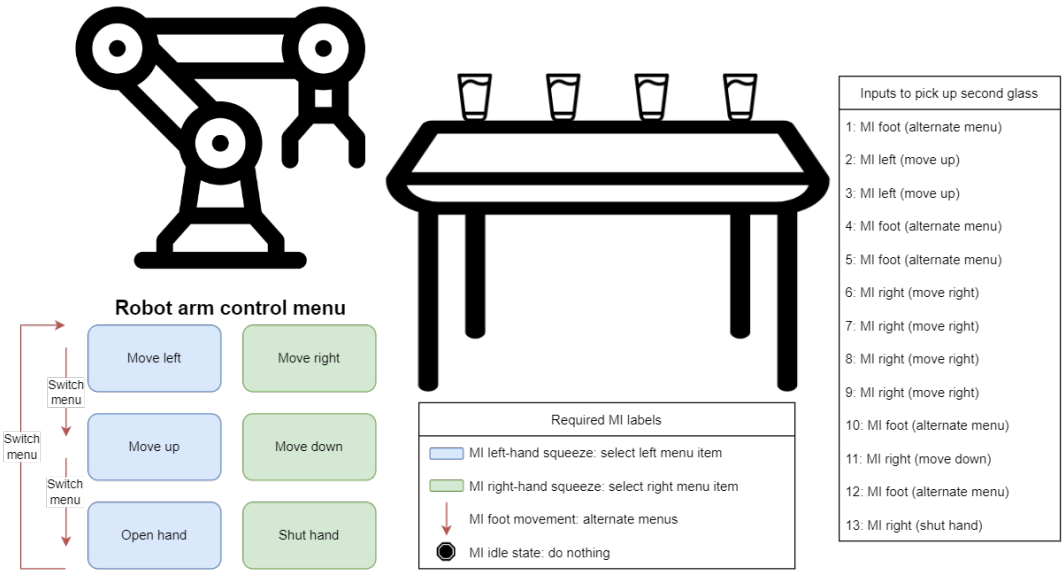
1.4.3 The potential of an AutoML variant for BCI pipelines

People that are first introduced to machine learning (ML) and especially deep learning (DL) might think that these techniques provide a *one-click solution* for data classification. Whilst in some cases, re-using existing complex DL methods can give satisfying results when providing it with enough training data, creating high-quality ML and DL systems still requires a lot of human expertise. However, attempts are being made at providing such one-click solutions through a process called automated machine learning (AutoML), which is reviewed in detail by He et al. (2021). The idea behind AutoML is self-explanatory. Rather than having an expert design a ML pipeline, an automated system tries all sorts of combinations for a pipeline based on provided labelled data and perhaps some general preferences in techniques to be tested. The problem then boils down to an optimisation problem where the combination of fine-tuned components that maximizes a certain metric such as accuracy on test data has to be found. AutoML facilities have not yet been made for BCI specific pipelines, but with Google providing AutoML as part of their cloud infrastructure for computer-vision, the AutoML idea is becoming more and more popular (Dillen et al., 2022).

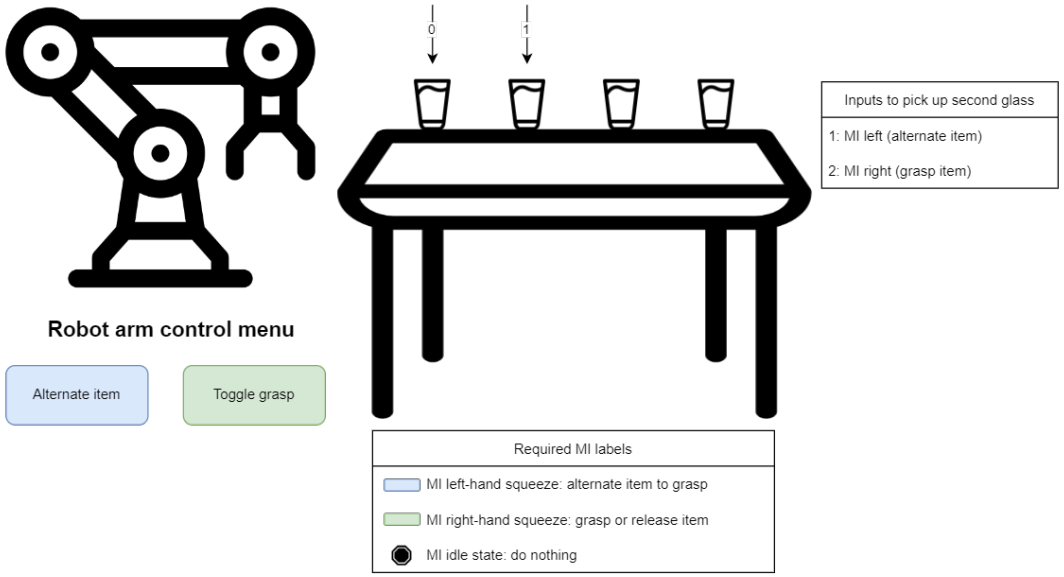
One library to develop a complete BCI pipeline

When taking into account the discussed literature so far, it is clear that many techniques for each component of the BCI pipeline already exist and that certain combinations have proven to be very successful. With more and more of these components being provided through open-source libraries as discussed in Section 1.2.4, it is not unreasonable to think that an AutoML library for BCI pipelines might be introduced in the future. If done right, this pipeline can set standards for the expected output of each component meaning that newly proposed components can be directly provided to the framework by the author of that new component. Whilst this would require a lot of commitment from the authors in the field, it could lead to newly gained insight on uncommon component combinations that provide a pipeline which performs best for certain problems. Such a framework would also enable new research to re-use existing components much faster as they would ideally all be present in the library with some form of a standardized output.

As this library would require a lot of maintenance, it should be provided by a team that has sufficient resources on maintaining it and updating it to support newer techniques in the field.



(a) A BCI system that offers complete control over a robotic arm but with poor efficiency.



(b) A BCI system that offers limited control over a robotic arm but with great efficiency.

Figure 1.8: The contrast between mappings of MI EEG for controlling a robotic arm to grasp an item. One approach allows for full control but requires four MI classes and many steps. One approach offers less control but succeeds in the task with three MI classes and very few steps.

Whilst there is currently no big organisation in the field that is likely to take this task upon themselves, the rise in popularity might attract such big corporations.

One library to develop multiple variants of one component from the BCI pipeline

Besides aiming to provide a complete pipeline, a library that focuses on providing all alternatives of one component following a specific standard would already be incredibly handy. Take for example a library that provides all open-source BCI datasets from one singular library, with all of the datasets following a specific standard. This would allow for testing a model on a whole suite of BCI data, such as MI EEG, and thus make it easier to compare results with existing literature as these would also use datasets from this common library. This has already been proven possible in other fields. For example, in reinforcement learning (RL), the environments used for training, which could be seen as the training data, are often environments provided by the OpenAI Gym library (Brockman et al., 2016). This library is maintained by OpenAI, a big company interested in RL among other AI techniques. Due to its popularity, many other authors contribute to this open-source library by providing their custom environments in a format that is directly supported by the OpenAI Gym library. A similar system for providing all sorts of BCI data would be incredibly helpful for the field of BCI research. It could also cause a chain reaction, with the introduction of other libraries that include state-of-the-art classification algorithms which work directly with this common data-providing library. This is also the case in the RL field with libraries providing algorithms for these mentioned Gym environments (Liang et al., 2017; Weng et al., 2021). Thus, providing such a library as the first step to an AutoML revolution in the BCI field might be a good idea.

1.4.4 A lack of standardized testing and reporting

The evaluation and comparison of different BCI systems is not an easy task. This is in part due to the many components a BCI system consists of, making direct comparisons often impossible. Is comparing the performance of a classification model from a BCI system that works with dry electrodes in a real-world setting to one that works with data from medical equipment in a lab environment fair? Is it fair to compare traditional two-step ML approaches that have great explainability and interpretability with black-box one-step DL approaches? Many more of these questions exist and they remain mainly unsolved for the BCI field.

The issue with individually evaluating BCI pipeline components

One potential approach is evaluating each component of the BCI system individually. However, whilst these may make comparisons between the different components easier, these individual metrics don't tell the whole story for the complete system. To truly test the capabilities of a proposed BCI system in the real world, intensive user studies should be performed. Even in its simplest form, such a user study should validate if the BCI system succeeds in its advertised tasks in both a timely and reliable manner that is pleasant to use for the target audience (Dillen et al., 2022). It should also be compared to existing solutions to ensure that there is an added benefit compared to these alternatives for the user. To avoid biases, these user studies should not only include enough participants but the participant group should also be heterogeneous.

Doing even the most primitive version of such a user study takes significant resources to perform. As such, it is rather uncommon to see these user studies in the same paper that proposes a BCI system. Initial articles of a new proposed BCI system often focus on providing only objective evaluation metrics of the classification system. These objective measures include accuracy, sensitivity and specificity. Simple proof of concept (POC) experiments are sometimes

also included in these initial articles but these are tested only on a small group of participants, which are often not from the target audience. The earlier discussed system by Halder et al. (2018) was user tested on able-bodied participants whilst the target audience was not. As discussed in Section 1.3.4, this study gave promosing results while a comparable system tested for a person in the target audience found that the user wasn't interested in using the proposed BCI system as existing solutions were more pleasant to use (Käthner et al., 2015). Some of the articles included in the review by Dillen et al. (2022) also performed user tests on participants that are not part of the target audience. However, with many current BCI systems focusing on providing systems for medical purposes, finding enough participants from the target audience to form a heterogeneous group can be challenging.

Proposing a complete BCI system throughout at least four articles

The above-mentioned points make the author of this master thesis believe that the development of a BCI system should take place over at least four separate articles. The first article should focus on data acquisition. It should discuss the equipment used for recording the brain signals. It should also address which kind of brain signals are extracted and where they are expected to originate from on an anatomical level. The exact procedure used for effective data collection should also be detailed. This includes details on what participants are included, if there were surveys used to test their capabilities both before or after the experiments, if the participants were trained, how many trials over which periods were performed per participant and much more. Such an article may contribute to the field by providing an open-source dataset, describing a novel training technique, discussing a more user-friendly data collection procedure, proposing new hardware and more. What kind of evaluation should be done in this type of article depends mainly on what the contribution of the article is. When proposing completely new hardware, a comparison between comparable hardware may be done in the same manner as Zerafa et al. (2018) compared different consumer-grade EEG headsets. When testing a new training procedure, the data may be fitted to a common classifier to compare its learnability with the data of participants that followed a different training procedure. Discussing all possible evaluation metrics falls outside the scope of this master thesis, but the most important factor is discussing the exact data acquisition process, as important details such as if live feedback was used often lack in existing work providing open-source datasets.

A second article should focus on the classification pipeline. It can contribute by proposing a novel combination of existing pipeline components or completely new components altogether. The type of components such a pipeline has and the evaluation that can be done will be discussed in detail in Chapter 3. These articles will make use of the data from the first article on data acquisition and should thus briefly summarize the used dataset. The experiments that will be done in this master thesis relate mainly to this type of article.

A third article should focus on linking this system to an external device such as a robot arm. This might involve changes to the classification pipelines, especially in the cutoff values for specific classes. Rather than optimizing global accuracy, it might be preferred to tune the system such that it makes more, but predictable errors. To illustrate this, a pipeline might be tuned such that it has a high true positive rate and a low false positive rate for actions that have significant risk. This means that when these actions with certain risks are performed, the system is almost certain the classification was correct. Doing this will make more false classifications for the other classes, but if these classes are neutral classes that don't perform actions, it has less risk associated. This corresponds to preferring a BCI for biosignal control over a robot arm that has a global accuracy of only 80% but makes almost all of its errors in not moving when movement was desired over one that has 85% global accuracy but has most errors in it moving when no

movement was desired. The focus of this article is not on obtaining the best classification metrics anymore but rather on providing the safest and most predictable system that is still useful. This third article might perform a simple POC to validate that the desired action taking is possible, even if this is only tested for best-case scenarios. It can contribute to the field by proposing more intuitive interaction methods, a new form or external device, a clever way of boosting reliability and predictability and more.

A fourth article should then focus on an in-depth user study with a heterogeneous group from the target audience. Ideally, this study should be performed over multiple sessions to study long-term impacts on the user's life, including an idea of the public acceptance of the proposed system. It should review all previous articles and is likely to require most resources, but is also the most important for bringing the proposed BCI to the real world. It contributes to the field by addressing the good and bad points of a proposed BCI system being used in the real world. It may highlight points that went undetected in the previous articles.

Comparing these different articles

Some metrics for comparing each type of article with each other were named in this section already. However, standardized tests should be introduced to fully ensure comparability between results and that a correct evaluation is done in each article. These are not yet present in the BCI field and proposing them is a difficult task as the proposing authors should be respected so that other authors are even willing to accept their proposals. It is also not easy to propose testing and reporting standards that fit all variants of these proposed articles. Do you train the models on a single patient and test them on the same patient or do you test for generalisability? Do you allow it to run on very capable hardware or limited but very affordable hardware? It should be apparent that there is no straightforward way of proposing standardized testing and reporting for all variants of these articles.

This master thesis only aims to highlight these issues as solving them is an open issue and standardized methods would need the consensus of the field, which a master thesis is unlikely to obtain. For now, creating at least four different papers to propose a complete BCI system and focusing on being as transparent as possible in each is already a good step in the right direction. However, transparency lacks in many papers, albeit due to page constraints or a lack of standardized reporting. Providing the source code with additional notes through an external source such as GitHub may help combat the page constraint issue. Following guidelines from related fields such as PROBAST (Wolff et al., 2019) and TRIPOD (Collins et al., 2015) originally proposed for CADx and CADe research can help in determining a new reporting standard for BCI research. Such a reporting standard will also aid in the repeatability and reproducibility of articles, especially for those where sharing the source code and data is not possible.

1.4.5 Challenges from the highly interdisciplinary nature of BCI systems

The idea of splitting the development of a complete BCI system into at least four separate articles discussed in Section 1.4.4 can also help with the challenges that the highly interdisciplinary nature of BCI systems offers. This is because each of those articles limits itself to a couple of domains rather than all domains present in BCI research. In practice, many articles aim to provide a system that is as complete as possible but this results in a system where many components, that fall outside the knowledge domain of the authors, are adopted from previous literature providing little to no scientific value.

Disciplines needed for each of the four proposed articles

When taking into account the four separate articles needed for proposing a complete BCI system proposed in Section 1.4.4, the following experts may be required. For the data acquisition article, a neuroscientist can greatly improve the value of the article by understanding in detail what brain signals can be measured and what interpretation they may have. They can also help in determining the optimal placement of the electrodes based on the anatomy of the brain and the expected source of the signal of interest. To create the hardware for the headset that captures the wanted brain signals, e.g. an EEG headset, engineers capable of reliably measuring and transmitting these incredibly low voltage and noisy signals are needed. A graphical designer can help with making the headset aesthetically acceptable.

The article related to data classification can be done mainly by a computer scientist specialised in machine learning (ML). However, when making visualisations and features, knowledge from neuroscience may still be needed. Because of this, a neuroscientist may still be very helpful for this article. Likewise, whilst a computer scientist may be able to reduce noise in a software manner, having an understanding of how the headset works and what types of noise and artefacts can be expected is also incredibly helpful. Whilst a well-documented data acquisition article may cover all these points, some aid from engineers in understanding the headset hardware can still be helpful.

For the third article, where the connection to an external device is discussed, a wide variety of experts may be required. For example, the previously discussed robotic arm with example menu's shown in Figure 1.8 requires computer vision knowledge for grasp detection, engineering knowledge to make the robot arm and general computer science knowledge to create a working link between classification labels and action controls. A graphical designer may once again help with making the external device aesthetically acceptable. General user experience (UX) knowledge can help in creating a great mapping between the detected brain signals and the control over the external device.

The fourth article may require all previous experts to finetune and configure the system. Since these trials involve human participants and often people suffering from a handicap, an ethics expert should also be consulted. Adding to this, by surveying the users in a manner that truly captures the UX and net benefits, both a psychologist and an expert of comparable systems may be needed. These discussed experts may vary greatly based on the type of BCI system that is proposed.

The issue with wanting to propose a complete system directly

The idea of proposing a BCI system as a result of four articles is proposed in this master thesis and is currently not present in the field. Whilst some researchers do focus on some components of a BCI system only, in the same spirit the division in multiple articles would do, most researchers aim to also propose a complete working BCI system in terms of a proof of concept (POC). Take for example the BCI system proposed by Herath and de Mel (2021). It has a significantly sophisticated robotic hand that functions almost completely as a human hand does. The effective hardware used for the complete BCI system, including the processing unit, are also well detailed. This shows that the authors have a significant understanding of the hardware a BCI system requires. However, the methodology used for data acquisition, the used classification algorithms and the user interface to the external system as proposed by Herath and de Mel (2021) could benefit from improvements. Since this is mostly a straight copy of existing methods and models not optimized to their proposed BCI system, one could argue they have little scientific value. The simple POC experiments are also of low value as discussed in Section 1.4.4. This means that these added components to create a POC, which take significant time, have a limited scientific

value. This is by no means a criticism to Herath and de Mel (2021) but demonstrates the interdisciplinary nature of BCI systems and how researchers that are specialized in one of these disciplines will outperform certain aspects of a BCI system while leaving room for improvement in other aspects. This interdisciplinary nature of the BCI field is part of what makes it so fascinating yet also very sophisticated with a steep learning curve. The above-proposed idea of splitting the development of a complete BCI system into multiple papers could limit the amount of time put into creating components that fall outside the scope of expertise and thus have less value in the article.

1.4.6 Brain signals are a complex data source

The human brain is an incredibly complex system of billions of neurons and trillions of synapses. Whilst great efforts are being made in mapping all of these neurons and synapses, faster than predicted, a complete mapping remains unavailable. The lack of such a complete mapping directly implies that research is unable to simulate the brain in silico. It also implies that there are still many unknowns in the complete structure of the brain and its internal working. Hodson (2019) talks about the efforts in mapping the brain and the many unknowns still present in the working of the brain in more detail. This limited understanding of the brain makes it a highly complex source of data, even when working with well-understood brain signals.

The most difficulties in using brain signals as a data source arise from the fact that the brain is a dynamic, nonstationary and non-linear system (Abdulkader et al., 2015). Non-linearity is a common challenge in many research fields and many ML and DL models can work with non-linearly correlated data. However, the dynamic and nonstationary nature of brain signals is less common in applications of ML and DL. Intuitively, this dynamic and nonstationary property results in measurements that can differ greatly for the same patient during the same trial doing the same task. For example, an envisioned MI task of the right-hand movement can be repeated by the user multiple times. However, the user's mental and emotional state, their focus and fatigue levels and many other aspects that will influence the brain signals will change continuously. Both due to the limited spatial resolution of measuring modalities and our limited knowledge of the brain, it is impossible to measure the wanted brain signal without measuring any other brain activity. Thus, this change in the brain activity that is not of interest makes the measurements nonstationary and dynamic, making the learning task significantly harder and more complex.

Adding to this, neuroplasticity, which was already discussed in Section 1.3.1, can change the structure and working of the brain. This makes the brain even more dynamic. Adding to what could be seen as internal noise factors of the brain, external noise can also be present in a nonstationary and dynamic manner. This external noise includes common artifacts such as electrical interference and the detection of EMG, which will be discussed in more detail in Section 2.4.6.

All of these factors already makes brain signals an incredibly complex data source for ML and DL, as is used in BCI systems. The limited knowledge of the brain's inner working and the dynamic, nonstationary and noisy nature of brain signal measurements also challenges the explainability and interpretability of these signals. Besides these factors, brain measurements are far harder to collect resulting in an often very limited amount of training data (Abdulkader et al., 2015). This makes many of the techniques used in fields where gigantic datasets are available, unapplicable for BCI research. Take for example the EEG MI dataset by Kaya et al. (2018), which is one of the largest publicly available datasets of this type. This dataset by Kaya et al. (2018) has around 60 000 combined usable samples over many different MI tasks and paradigms. Taking into account that only a subset of these samples can be used for training, this collection is multiple orders of magnitude smaller than datasets from other ML and DL applications. Take

for example the commonly used ImageNet dataset by Deng et al. (2009) which contains over 3 000 000 images for training an image recognition classifier. The high dimensional and time-dependent nature of brain measurements also makes it harder to re-use common ML and DL approaches from other fields.

1.5 Ethical challenges for BCIs

The field of brain-computer interfaces (BCIs) has many promosing aspects, with Sections 1.2, 1.3 and 1.4 discussing the rise in popularity, the growing amount of use cases for BCI systems and some of the opportunities in the field. These evolutions in the field have made BCIs present in more then just research labs (Abdulkader et al., 2015; Bernal et al., 2021). Widespread adoption of BCI systems is still unlikely in its current state, with some obstacles that need to be overcome first. This relates to creating a system that fits Wolpaw's definition of a perfect BCI and overcoming at least the challenges discussed in Section 1.4. However, with the recent news coverage of the Elon Musk company Neuralink (Musk & Neuralink, 2019), discussed in Section 1.2.1, it has become clear that ethical challenges might be tougher to overcome than technical ones. Even if Wolpaw's vision of a perfect BCI can be met and a net benefit to the user has been proven, some users' moral intuition may refrain them from using such a system.

Whilst the author of this master thesis is by no means an ethics expert, he believes that it is important for any researcher interested in the BCI field to be aware of some of the many ethical challenges BCIs have. This section aims to illustrate what kinds of ethical challenges are related to BCI research and how there is no clear solution for them. The interested reader is invited to consult the interesting articles on the ethics of BCIs by Drew (2019) and Bernal et al. (2021) for a more in depth ethical study. Some articles have also been made focusing solely on ethical questions surrounding the Neuralink proposed BCI (Dadia & Greenbaum, 2019; Jawad, 2020).

1.5.1 Painfully confronting users with their brain

When a BCI system is used by patients suffering from a brain disorder, it is easy for the user to be confronted with their disorder by that system. Take for example an invasive medical BCI used to predict when an epileptic seizure would occur such that it can alert the user to take the right medication to prevent the seizure from occurring. Each time the system alerts the user, the user is confronted with the fact that they have epilepsy and a potential epileptic seizure coming. Gilbert et al. (2019) performed a clinical trial on six users suffering from epilepsy that use such a BCI system. Whilst positive results in seizure reduction were found for all patients using the system, the user experience (UX) differs greatly between all six participants.

From all six patients, four expressed how they felt more in control of their daily life thanks to the system. They reported a pleasant overall experience with the system. Another patient, *patient four*, expressed that he saw epilepsy as part of his life and himself. As a result, he wasn't interested in being dependent on the BCI system to warn him and ignored it most of the time. More worryingly, *patient three* said the following:

> [The BCI] made me feel I had no control. So I didn't have control over what I was going to do. [...] It made me feel that I was always different [from] everyone not just in the moment of the seizure [...] I got really depressed
>
> *Patient three from the study by Gilbert et al. (2019)*

Being aware that the system was constantly monitoring her brain activity and an alert could

be sent at any time, the patient felt *different to everyone else* all of the time, whilst without the system, this feeling was only present when a seizure occurred. The BCI system constantly reminded her that she has a brain condition which had a significant psychological burden. Whilst this is only the experience of one patient from a small trial, it was an important finding that has given many ethicists reason to question the use of BCI systems for all patients (Drew, 2019). It is not unreasonable to think that similar feelings might occur when using BCIs for neurological rehabilitation through neuroplasticity, a use case discussed in Section 1.3.1. Neuroplasticity may take a long time or even be impossible for some users. Having constant feedback that shows no or poor progress can be very confronting for the user. Even in commercial settings, when a person tries to calibrate a headset relying on MI tasks and the calibration fails due to the user being poor in envisioning the MI task, a feeling of *lacking this capability compared to others* might cause a dramatic UX. This poor UX for certain users is something that can only be found with thorough trials and further shows how important it is to do real-world experiments.

1.5.2 E-waste inside your skull

Section 1.3.2 already adressed Second Sight, one of few companies providing visual prostheses with approval from the FDA. This company discontinued support for some of its earlier visual prostheses. This resulted in all of the users from those early products losing any form of support. As discussed by Strickland and Harris (2022), this has major consequences. First, the technology that has enabled them to restore their vision to a certain extent can fail at any time. If this happens, this results in an indefinite loss of vision again due to the lack of support. Second, the company refuses to make the required data for the maintenance of the system public. This means that no third party can ever work on it. Third, due to no third party being capable of removing the system, the users are essentially stuck with what is or will become e-waste inside their skull. This has already caused major issues for one user, Ross Doerr. As discussed by Strickland and Harris (2022), Ross can't get the help he needs with a brain tumour due to doctors not having the required information about the brain implant to ensure a safe operation. Even a simple magnetic resonance imaging (MRI) scan can't be planned safely due to this lack of information surrounding the brain implant.

With commercial companies showing interest in invasive BCIs, such as the earlier discussed Elon Musk company Neuralink, it is not unreasonable to think a similar situation might occur from the manufacturer of such systems. What happens to an invasive BCI system once the manufacturer doesn't support it anymore or the patient isn't interested in using it anymore? It becomes clear that some form of regulations should be in place to ensure that users of such invasive systems are always able to get the information needed from the manufacturer to ensure they are not restricted in getting needed care. It also begs the question if invasive BCIs is really the road to go and how dependent we want to be on these technologies, taking into account they might fail for an unspecified period at some point.

1.5.3 Changing peoples personal identities

Section 1.5.1 already adressed the trial conducted by Gilbert et al. (2019). In this trial, six patients suffering from epilepsy using an invasive BCI for epileptic seizure reduction were monitored over a long period. Besides the finding of extreme psychological burden for one patient, the following words from *patient six* are also eye-opening:

> [The BCI] was me, it became me. [...] I found myself changing. [...] I felt like I could do anything. [...] With this device, I found myself.
>
> *Patient six from the study by Gilbert et al. (2019)*

Whilst this is a positive experience according to the user, Gilbert et al. (2019) described this as a radical symbiosis between the patient and the system. The idea of users feeling themselves becoming one with technology, creating a symbiosis, has multiple ethical concerns (Drew, 2019). These concerns fall outside the scope of this small ethical discussion on BCIs. However, this change of personality is something that can influence the patient's life in a manner that perhaps wasn't initially expected. For example, when someone goes through major weight loss, they have more chance to see a marital status change. Both in terms of getting married or having a divorce (King et al., 2022). Similar trends might occur due to behaviour changes as a consequence of having a BCI.

Likewise, some people that suffer from a medical condition might embrace their condition and share some sort of culture with people that have the same condition. People suffering from major hearing loss may enjoy parties oriented toward them where bass-heavy music is played at loud volumes such that they can feel the vibrations of the music throughout their body. For them, this is how they can enjoy music without hearing music. If such a person were to regain hearing through a BCI, it is unlikely they will enjoy hearing the load and constant bass. Similarly, people suffering from major vision impairment might enjoy tactile art. When vision is restored, their impressions of these kinds of artworks are bound to change.

In both of these situations, the interest in the events they once liked might faint and the connection to the community surrounding them might vanish. They lose touch with the *community they were once part of* (i.e. the blind and deaf community). These are things that might go unthought of when looking at the added benefits of a BCI system might offer. However, they can cause major identity issues where users are left in a quest of finding who they are after a major change in their daily life. This quest is one that not many look forward to doing and maybe a reason some people opt to not use a BCI system.

1.5.4 A great risk of sloppy science

With a lack of standardized testing and reporting as discussed in Section 1.4.4, it is easy for researchers to perform *sloppy science*. It allows them to report and compare only those metrics that have the desired outcomes rather than performing time-consuming user trials that have the most scientific value. With many authors wanting to propose complete systems over singular components, the risk of sloppy science is increased even further as the other components are highly likely to receive less thought resulting in poorer quality, as discussed in Section 1.4.5. However, respected journals have appropriate mechanisms in place, such as peer reviews, to stop published articles from performing sloppy science.

More worryingly, private companies aiming to commercialise their BCI systems as soon as possible are less subject to these mechanisms of protecting against sloppy science. Willingham (2021) expresses some of her concerns with Elon Musk's Neuralink company in this regard. Willingham (2021) refers to an interview with ex-employees by Brodwin and Robbins (2020). In that interview, it is discussed how more than half of the initial researchers have left the Neuralink company already and how these ex-employees experienced a rushed timeline which wasn't compatible with the slow pace of science. This led to multiple internal discussions between the initial researchers wanting more time and shareholders wanting a sellable product quickly. Whilst these ex-employees were also quick to state that they see Neuralink as an incredibly

innovative company in the BCI field and believe they will succeed in their mission, these words about a rushed timeline are worrying. As discussed in Section 1.4.4, thorough experiments take a lot of time, so it isn't surprising that this isn't compatible with a commercial aim of providing a product to market as soon as possible. However, when considering that Neuralink's proposed system is an invasive one and the ethical questions about e-waste in your brain discussed in Section 1.5.2, the worry of such companies performing sloppy science to be among the first on the market only grows. Whilst some legislations are in place that should ensure safe human trials, only time will tell how much these commercial companies respect the time-consuming nature of research.

1.5.5 Advertisements based on your thoughts

Big tech companies such as Meta, formerly known as Facebook, aren't known for respecting the privacy of their users and protecting their data (Fuller, 2019; Hu, 2020). Besides multiple court cases, the European Union (EU) and individual governments have made multiple efforts to protect their citizens against the malicious use of their data. This includes legislations such as the general data protection regulation (GDPR). Whilst this is great, it hasn't stopped companies from continuing their excessive data mining to sell it to advertisement companies. As discussed by S. Lee (2020), the advertisement industry functions more or less the same since the introduction of the GDPR. Advertisement companies are still able to purchase a lot of personal and sensitive data from users to use for targeted ads. The only real difference for users is that they are now presented with ambiguous popups and opt-in data sharing they have to accept to enjoy all available content a company offers. In many of these cases, the user isn't aware of what they are accepting, and the user protection that the GDPR tries to offer gets lost.

With Meta having shown interest in BCIs and Steam, a marketplace for games, showing interest in integrating a BCI with a VR headset, the question remains how these companies will use the acquired brain signals. If they process them in the cloud, it is a necessity that the acquired signals are transferred to their servers and the risk of data leaking and other scandals increases. It is also expected for these companies to have long, complex and ambiguous terms of service (ToS). This could be paired with the previously mentioned opt-in data processing agreements that have to be accepted to use all functionalities of their product or which are designed in such a way that it is hard to not opt-in for them. This raises the question if it is desired that a commercially oriented company, which makes a large amount of money from selling user data, sells BCI systems. New legislations are likely required to limit the amount of processing that can be done on acquired brain signals, even if consent is given that the data may be used by the company to *improve their services*.

1.5.6 Hacking BCI systems

It is impossible to fully protect any system against hacking, hijacking or any other malignant use. Locks can be picked, databases can be breached, Bluetooth (BT) connections can be spoofed and even the most secure encryption algorithms are prone to hacking (Castelvecchi, 2022). Martínez Beltrán et al. (2022) have shown that an EEG based BCI working with P300 signals can be hacked by generating fake P300 waves from an external device. These generated waves are recorded similar to other electrical interference, such as the typical 50 hertz (Hz) or 60Hz alternating current (AC) inference, but the processing algorithms see them as effective P300 signals and thus the system is now controlled by those artificially generated waves rather than the user's brain signals.

More traditional hacks such as BT spoofing can also be performed when BT is used in

a wireless BCI. Even if the data transmitted over BT is encrypted, flaws in the encryption methodology may be found that enable hackers to decrypt the encrypted data. Whilst none of this is exclusive to BCI systems, it is something that raises ethical questions. If BCIs are ever used for controlling many devices, some having certain risks, the hacking of these BCI devices can impose fatal consequences. Will all companies put in the required resources and effort to protect their BCI systems from malignant attacks?

1.6 Chapter conclusions

This first chapter gave an in-depth but intuitive introduction to brain-computer interfaces (BCIs). Section 1.2 discussed the growing scientific and commercial interest in BCIs. Sections 1.3 and 1.4 provided some more insight on the common use cases for BCIs and some of the opportunities and obstacles for BCI research. This chapter ended by addressing some ethical challenges any researcher interested in BCIs should be aware of. Throughout this chapter, many articles were discussed and some strong points and weak points were explicitly mentioned. Many of the discussed BCI systems were also compared to Wolpaw's definition of a perfect BCI given in Section 1.1.

From this chapter, it should become clear that BCI research has grown and matured significantly over the last few years. Positive human trials have been performed showing that BCIs can improve the quality of life for people with certain medical conditions. However, many systems can still see improvement and it is incredibly hard to meet Wolpaw's definition of a perfect BCI. This is great news for researchers interested in the field, as it means that there is still a lot of work that can be done. Some of the biggest current challenges with BCIs lie in the fact that the amount of open-source data and code-providing libraries is rather limited and that there is no standardized way of testing and reporting on a BCI system. It was also discussed that BCI research is highly interdisciplinary which makes it hard for individual authors or authors from the same expertise domain to provide a complete BCI system. To tackle some of these challanges, Section 1.4.4 proposed that the proposal of a complete BCI system should happen over at least four different papers. In order, they should focus on data acquisition, data classification, connecting the classifier to an external device and testing the whole system in a thorough human trial.

Since this master thesis is from a computer science student, the following chapters will focus on what is important for a computer science student that wants to contribute to this field. The remainder of Part I will provide some more general BCI knowledge required to fully provide an article on the data processing of a BCI system. To do this, Chapter 2 will discuss in greater detail what biomedical signals (biosignals) such as brain signals are, where they originate from and how they can be measured. With a deeper understanding of what EEG data is, Chapter 3 discusses in greater detail how they can be processed. All of the components from a common BCI pipeline are discussed and some examples of these components are given.

The next part, Part II puts this theory to practice by discussing the development of seven offline MI EEG classification pipelines. This includes the implementation of two CSP based two-step ML approaches and three literature proposed state-of-the-art CNN based one-step DL approaches. Two LSTM extensions to the CNN based EEGNet model are also proposed and evaluated accordingly. Whilst focusing on an offline BCI system in Chapter 4, some of the required steps to go to an online system are discussed in Chapter 5. The implemented system makes use of existing EEG data from multiple MI tasks for classification. All of the implemented variants of the BCI pipeline are compared and the results of this master thesis are further discussed in Part III, the final part.

Chapter 2

Origins and acquisition of biomedical signals

2.1 Introduction

Whilst Chapter 1 has provided an in-depth intuitive introduction to BCIs, some more technical aspects need addressing as well to provide a computer scientist with all of the required foundational knowledge that is necessary for BCI research. This chapter provides the required technical knowledge on the data that BCI systems use, brain signals. Brain signals are only one of the many types of biosignal present in the human body. Whilst from a computer scientist's perspective brain signals may just be another type of input data to a classification model, having at least a basic understanding of this data is crucial in making good classification algorithms for BCI systems. Even when using DL approaches where no manual feature engineering is necessary and where basic models may have satisfactory results, understanding the data will allow for the creation of better models. This understanding of the data also helps in troubleshooting why some models may not have the desired results.

To provide this basic understanding, this chapter starts by briefly discussing what biosignals are and where they originate from in the human body. After this general discussion on biosignals, a focus is put on the different biosignals from the human brain, with brain signals measured using EEG in particular. This EEG measuring technique and other measuring techniques are also discussed in greater detail. Whilst it is addressed that EEG has some fundamental shortcomings over other measuring modalities, it also has some attractive properties over these alternatives. These attractive properties are listed and provide an argument as to why the remainder of this master thesis will focus on EEG and MI EEG in particular.

2.2 Biosignals in the human body

Biosignals are nothing more than a measurement over time of a living being. In practice, these biosignals are closely related to physiological processes. This makes it possible to monitor or detect those physiological processes using biosignals. Biosignals can be produced by different forms of energy, such as the electrical energy form when measuring EEG in μV. Table 2.1 summarizes some of these energy forms and which type of biosignals they can produce. Whilst living beings, including humans, produce many different types of biosignals, this master thesis will only consider time-varying electrical biosignals. These types of biosignals, sometimes reffered

to as electrical biomedical signals (bioelectrical signals), are the ones used as input data for BCI systems. Semmlow (2018) discusses these and other types of biosignals in greater detail.

Energy form	**Variable type**	**Biosignals**
Chemical	Chemical activity and/or concentration	Blood ions, O2, CO2, pH, hormonal concentrations, and other chemistry
Mechanical	Position, force, torque or pressure	Muscle movement or cardiovascular pressures, muscle contractility, valve and other cardiac sounds
Electrical	Voltage or current	EEG, ECoG, ECG, EMG, EOG, ERG, EGG, GSR and EDA
Thermal	Temperature	Body temperature and thermography

Table 2.1: Some of the different energy forms in living beings and the most important measurable biosignals they produce. Data from Semmlow (2018).

2.2.1 Origin of electrical activity inside the human body

Electrical activity in the human body, better known as bioelectricity, can be seen as the generation or action of electric currents and voltages in physiological processes. As shown in Table 2.1, the measurement of this bioelectricity is what enables the monitoring of bioelectrical signals. This is one of the reasons biosignals are closely related to physiological processes since it measures the bioelectricity generated by these processes. A complete understanding of how bioelectricity is generated, maintained and transmitted in the human body isn't required for a computer scientist to contribute to the BCI field. However, a superficial understanding of this process makes it easier to understand the limits of the accompanied bioelectrical signals. As such, the remainder of this section gives a simplified explanation of how bioelectricity is made, maintained and transmitted in the human brain. This explanation is based on chapter 12 of the recently renewed book by Betts et al. (2022) and an EEG focused explanation of bioelectricity by Kirschstein and Köhling (2009).

Resting membrane potential causes negatively charged neurons

As was already addressed in Section 1.2.2, the human brain has billions of neurons with Nunez and Cutillo (1995) stating that around 10^7 parallel pyramidal neurons reside in only a single cm^3 of the brain cortex alone. A neuron, also known as a nerve cell, is an electrically excitable cell. Being an electrically excitable cell, a neuron has a resting membrane potential. This resting membrane potential is around -70 millivolts (mv) and expresses the difference in electrical charge between the inside and the outside of a neuron. This negative difference is maintained by the sodium-potassium pump which is responsible for the hydrolysis of ATP to ADP. During this hydrolysis process the sodium-potassium pump releases three positively charged sodium ions (Na^+) whilst only taking in two positively charged potassium ions (Ka^+), this difference causes the membrane potential to remain negative.

Action potential allows for neuron communication

Whilst the sodium-potassium pump inside the neuron explains why there is a negative resting membrane potential of around -70 mv, it doesn't explain the variable volt measurements of EEG. The change in membrane potential occurs when the neuron gets excited. The most common way a cell gets excited is through the process known as an action potential. An action potential forms the basis for electrical signalling within neurons, enabling some form of communication between them. To do this communication, neurotransmitters released by another neuron bind to receptors on the dendrites of the receiving neuron which has a depolarization effect.

This depolarization causes the neuron to become less polarized, resulting in its membrane potential moving close to zero. When sufficient depolarization occurs, the action potential process could start. This process is visualised in Figure 2.1. For the action potential process to start, the depolarization should be of such a magnitude that the neuron reaches its threshold membrane potential, which is around -55 mv. This is achieved through the repeated binding of neurotransmitters to the receptors. The annotation for "failed initiations" in Figure 2.1 denotes the common situations when the threshold membrane potential is not reached.

When the threshold is reached, a large number of sodium channels open, allowing many positive sodium ions (Na^+) into the neuron, causing the membrane potential to rise quickly. This depolarization is what creates the electrical signal known as the action potential that travels down the neuron to eventually release neurotransmitters itself. Eventually, a peak is reached, after which the sodium channels close, not allowing any further sodium ions (Na^+) to enter the neuron. To return to its resting membrane potential, the neuron opens its potassium channels to release many potassium ions (Ka^+). This is known as the falling phase where the neuron repolarizes. However, the release of positive potassium ions (Ka^+) happens so quickly that the membrane potential falls below the resting membrane potential. The neuron is now hyperpolarized, denoted as undershoot in Figure 2.1. During this hyperpolarized state, also known as the refractory period, failed initiations occur more often as it is very difficult to fire the neuron again. Eventually, the resting membrane potential is reached again and the neuron functions like before.

EEG measures postsynaptic potentials

Whilst the action potential explains how most changes in membrane potential of an individual neuron occur, it is highly unlikely to be measured by EEG and other measuring modalities. This follows from the fact that, as discussed earlier, many billion neurons make up the brain making it impossible to monitor a singular neuron. Since action potentials are such rapid current flows, it is highly unlikely enough neighbouring neurons will have an action potential at the same time resulting in a measurable signal. However, whilst it was discussed how neurotransmitters can cause depolarization which can initialize action potentials, these neurotransmitters can also influence the membrane potential in the opposite direction by causing further polarization. The release of these neurotransmitters and the binding to the receiving neuron also causes currents known as postsynaptic potentials. Whilst these are not action potentials, they are essentially what causes an action potential to take place and the action potential can also cause the release of neurotransmitters. These postsynaptic potentials are present for a longer period than action potentials. Thus, it is more likely for many neighbouring neurons to have active postsynaptic potentials simultaneously. The summation of these postsynaptic potential currents from many millions of neurons is what is detected by EEG. However, neurons experiencing action potential at that same time are among many other things sources of noise in this summation of these currents. Biasiucci et al. (2019) explain in greater detail what exactly is measured with EEG.

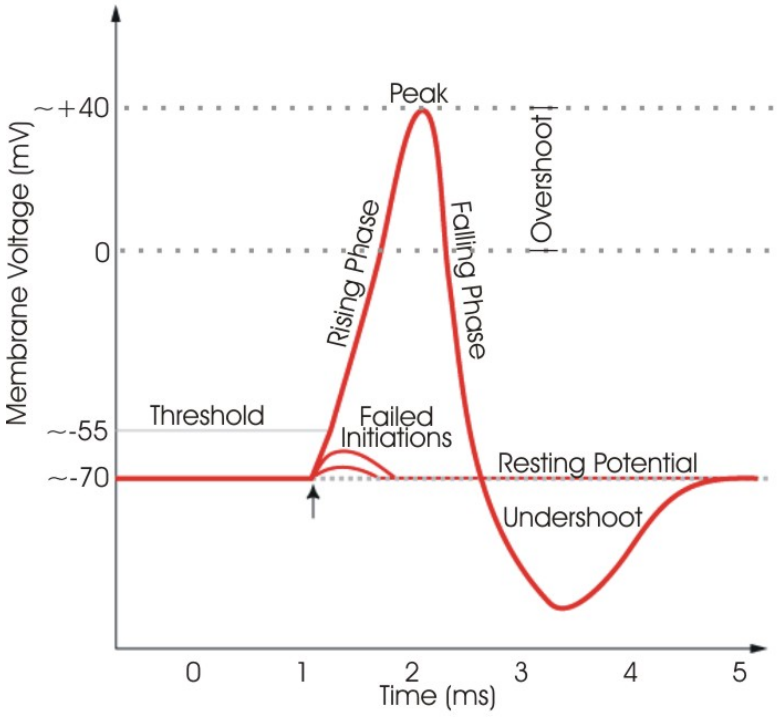


Figure 2.1: Chart of the membrane potential during the action potential process in a neuron. Free to use figure by Synaptidude, GFDL 1.2, via Wikimedia Commons.

Biosignals originating from bioelectricity in the human body

Both voltage and current can be measured from bioelectricity, providing many different biosignals. The measurement of voltage originating from the brain, measured in a non-invasive manner with electrodes on the scalp is known as EEG. Likewise, electrocardiography (ECG) revolves around the non-invasive measurement of voltage originating from the heart. Electrooculography (EOG) and electroretinography (ERG) are techniques used for measuring bioelectricity that originates from the eyes. Table 2.1 provides the most important biosignals related to the measurement of bioelectricity. The most important bioelectrical signals in BCI research are EEG and the invasive alternative ECoG.

2.2.2 Other energy forms and their related biosignals

Section 2.2.1 illustrated that bioelectricity is omnipresent in the human body. The human nervous system relies on bioelectricity to quickly carry information from the human body to the brain and the other way around. The process of muscle contraction is started when the discussed action potentials release neurotransmitters from motor neurons to muscle fibres. As discussed by Robinson et al. (2021), bioelectricity can even be *hijacked* as a way of novel medicine and treatment, i.e. for cancer treatment.

However, bioelectricity is only one energy form present in the human body. As shown in Table 2.1, other energy forms provide many other types of biosignals. Whilst these are of great

importance in many fields, BCI research is mostly interested in brain activity and thus the bioelectrical signals EEG and ECoG. However, as discussed in Section 1.3.3, hybrid systems are being explored which may combine these bioelectrical signals with biosignals from other energy forms. These types of hybrid systems and the other types of biosignals they use fall outside the scope of this master thesis.

2.3 Measuring biosignals from the brain

Neuroimaging is a discipline focused on capturing the anatomy and function of the central nervous system (CNS), which includes the brain (Fulham, 2004). As such, the modalities for measuring biosignals in the brain are mostly a subcollection of neuroimaging techniques. Many different types of these modalities exist, both invasive and non-invasive. There is no singular best measuring modality and often a trade-off has to be made between affordability, the signal-to-noise ratio (SNR), the ease-of-use and the risks involved. The SNR metric provides the ratio of the meaningful input to that of meaningless or unwanted input. Section 2.3.1 will discuss these modalities in further detail. The remainder of this master thesis then focuses on the EEG modality as it remains one of the most attractive modalities for BCI research.

2.3.1 Common measuring modalities in BCI systems

Research by Berger (1929) is the first in describing the measurement of brainwaves from the human brain in a non-invasive manner. Because of this, the German neuroscientist and psychiatrist Hans Berger is often seen as the inventor of EEG. Whilst he was one of the first to use the term *elektrenkephalogramm*, it was Richard Caton who first described the findings of bioelectricity in brains in general. He found this phenomenon in animal brains as early as 1875 (Haas, 2003). Since then, the neuroimaging field has been looking for new ways to capture these electrical signals coming from the brain and monitor the anatomy and function of the CNS in general. This has caused the introduction of many different measuring modalities, each with its strengths and weaknesses. However, many of these modalities require bulky equipment which makes them non-portable and thus non-attractive for use in general BCI systems (Martini et al., 2020).

Table 2.2 provides an overview of the most common measuring modalities considered for use in BCI applications together with some of their properties. Many more modalities exist but these are currently not portable enough or have not yet been proven useful for use in BCIs (Martini et al., 2020; Ramadan & Vasilakos, 2017). Modalities based on the presence of bioelectrical signals in the human brain seem most probable to be used in BCI systems for the foreseeable future (Ramadan & Vasilakos, 2017). This includes EEG, ECoG, intravascular electrodes and implemented electrodes among other modalities. However, there is also a growing interest in using modalities working with non-electrical signals for BCI applications. Whilst these are still limited due to poor portability and affordability, they have promising aspects for future research. This section focuses on some of the most important modalities for BCI research. The works by Martini et al. (2020) and Ramadan and Vasilakos (2017) cover some more modalities in greater detail.

Electroencephalography (EEG)

Electroencephalography (EEG) is a non-invasive technique used to measure the electrical activity of the brain, often expressed in μV over time. Electrodes are most commonly placed on the scalp with as main goal to detect the activity of neurons in the cerebral cortex. The cerebral cortex, better known as grey matter, is the outmost layer of the brain, closest to the electrodes placed

Modality	Invasive	Risks	Temporal resolution	Spatial resolution	Affordability	Ease-of-installation	Ease-of-use	Comfort
EEG	No	Low	50ms	10mm	€	Good	Good	Medium
ECoG	Yes	High	5ms	1mm	€€-€€€	Poor	Good	Good
fNIRS	No	Low	1000ms	10mm	€-€€€	Good	Good	Medium
Implemented microelectrode	Yes	High	3ms	0.05mm - 0.5mm	€€€	Poor	Good	Good
Intravascular electrode	Yes	Medium	5ms	2.4mm	€€-€€€	Poor	Good	Good
fTCD	No	Low	1000ms - 5000ms	10mm - 30mm	€	Medium	Medium	Poor
MEG*	No	Low	1ms - 5ms	1mm	€€€+	Good	Poor	Poor

Table 2.2: Overview of the most common neuroimaging modalities for use in BCI systems. *: Currently not available in a portable manner. Data based on Martini et al. (2020) and Ramadan and Vasilakos (2017).

on the scalp. Whilst more inner layers of the brain also produce information-rich bioelectrical signals, these signals do not reach the scalp in a clear detectable manner for EEG to pick up. As discussed in Section 2.2.1, EEG is not capable of measuring the individual activity of neurons in the cerebral cortex but rather the activity of large groups of neurons that are active at the same time. Postsynaptic potentials contribute most to these measurements as action potentials are too short in duration. The location at which electrodes are placed on the scalp often follows specific standards which are further explained in Section 2.4.1. These electrodes can be divided based on whether they are wet or dry, active or passive and whether they are wired or wireless. The difference between these types has already been covered in Section 1.2.2.

The output of these electrodes is processed through a differential amplifier which takes two electrical inputs and displays the output as the difference between these inputs. This means that an EEG measurement displayed from one electrode is not absolute but rather the relative difference between that electrode and another. The selection of the other electrode for passing through the differential amplifier can differ per device through what is known as the montage of the EEG device. The most common montages are common reference, average reference and bipolar. The common reference montage uses a predefined electrode as a reference electrode and compares each electrical signal from an electrode with it. Most commonly this reference electrode is placed on the ear as this is least influenced by brain activity. The average reference montage compares each electrical signal from an electrode with the averaged electrical signal from all electrodes. The bipolar montage follows a certain scheme of comparing two always differing electrodes with each other. Most often this is done in a sequence from the front of the scalp to the back of the scalp. All of these montages are summarized in Figure 2.2. The output of the differential amplifier is what is known as the EEG measurement.

EEG is one of the most popular neuroimaging modalities and also the most common modality in BCI research (Dillen et al., 2022; Martini et al., 2020; Ramadan & Vasilakos, 2017). This success is due to the good temporal resolution, its non-invasive nature and the relatively low cost of EEG hardware, further discussed in Section 2.3.3. Many different types of headsets and caps exist which hold the electrodes in place and with minimal modification, these headsets are easy to install and use, which is also an attractive property of EEG.

However, some of the headsets for dry electrodes, such as the Ultracortex Mark IV discussed in Section 1.4.2, can be unpleasant to wear over prolonged sessions. Other disadvantages include the poor spatial resolution as discussed in Section 1.2.2. This poor spatial resolution combined with multiple fluids and structures blocking the electrical signals coming from the brain means that the data collected from EEG is relatively low in quality, both from a SNR perspective and a

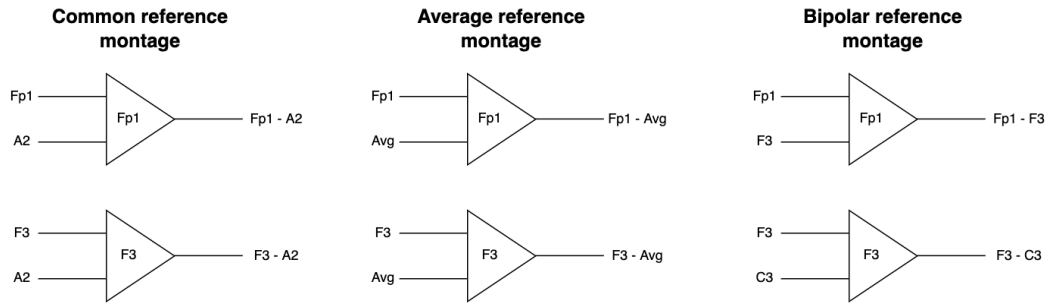


Figure 2.2: Illustration of the three most common montages for the differential amplifier in EEG. The annotations are the electrode names used in the international 10-20 system.

source localisation perspective. Since EEG measurements are limited to the activity of neurons in the grey matter, not all brain activity can be measured by EEG either. Section 2.2.1 and Biasiucci et al. (2019) explain in greater detail what EEG effectively measures. Some studies have also shown that extensive EEG use can cause temporary hair loss and the electrolytic gel could also cause allergic effects for the user (Morris et al., 1992; Tseghai et al., 2021).

Electrocorticography (ECoG) and other invasive methods

With most shortcomings of EEG being related to the non-invasive nature of the modality, invasive variants of EEG have been studied extensively as well. One such invasive modality to measure the bioelectrical signals of the brain is electrocorticography (ECoG). When using ECoG, electrodes are placed directly on the cerebral cortex, the outer layer of the brain. ECoG was first explored in the 1950s to determine the exact origin of epileptic seizures so that this part of the brain could be surgically removed (Palmini, 2006). The procedure used involved opening a portion of the skull, placing the electrodes on the exposed area of the cerebral cortex, waiting for a seizure to occur such that ECoG can be used to determine the exact location, removing that part of the brain, removing the electrodes and closing the skull again. This was done in this manner as the electrodes are physically closer to the brain, improving the spatial resolution and SNR significantly when compared to EEG. To this day, ECoG is still used in certain brain surgeries and this type of use of ECoG is called *intraoperative* ECoG.

However, this superior spatial resolution, far better SNR and the relatively fixed electrode positioning combined with a system that is invisible to the human eye has made it an attractive variant of EEG for post-surgery measurements as well. This is known as *extraoperative* ECoG. Many successful steps have been made to lower the risk and costs associated with the invasive nature of ECoG. When only interested in *extraoperative* ECoG, a craniotomy, also known as an open skull surgery, is not needed anymore. Instead, tiny holes can be made in the skull to implement the electrodes in that way. The size and associated surgical risks of those holes have been shrinking.

The Neuralink company discussed in Section 1.2.2 is responsible for the most recent advancements in this area, with the implantation of hair-sized electrodes being capable using only a robot and minimal anaesthesia. At this stage, it is unclear whether their system will be considered *extraoperative* ECoG or a different type of invasive measuring modality. Since these electrodes are so tiny, many thousands could be used for measuring the brain signals, compared to EEG

where a system with 21 or 64 electrodes is most common. Whilst for EEG having more than 1024 electrodes is almost impossible due to the physically limited space on the scalp, invasive techniques are aiming to allow for thousands of electrodes in only a small region of the brain. Whether this increase in electrodes will be useful has yet to be determined, as for EEG it is known that adding more electrodes does not improve the measured information past a certain point (Aflalo et al., 2015; Lebedev, 2014). However, many of these procedures still need significant medical trials before widespread use will be possible and as such, ECoG and other invasive methods such as intravascular electrodes (minimally invasive) and implemented microelectrodes (best spatial resolution) are yet to see widespread BCI adoption outside medical applications. These latter invasive methods are covered in more detail by Martini et al. (2020). There are also some ethical questions surrounding the use of invasive BCI, as discussed in Section 1.5.2.

Other modalities which do not rely on bioelectricity

As discussed in Section 2.2, there are multiple energy forms that produce biosignals. Measuring electrical signals from the brain is by far the most common in BCI research as these bioelectrical signals carry much information surrounding brain activity whilst also being relatively easy to measure. The discussed EEG, ECoG, intravascular electrodes and implemented electrodes all measure bioelectrical signals. Besides measuring biosignals which are being emitted by the human body all the time, it is also possible to expose the human body to an external energy form and measure the reaction of the human body. These types of modalities are common for medical imaging in hospitals. The popular X-ray, computerised tomography (CT), positron emission tomography (PET) and single-photon emission computerized tomography (SPECT) imaging modalities all rely on different forms of ionising radioactive exposure to the human body. Non-ionising methods are also common, such as the MRI modality which relies on exposing the body to strong electrical fields.

Some of these medical imaging modalities have seen use in highly medical BCI applications, but they require expensive and non-portable equipment, which makes them unsuited for use in general use BCI systems and conflicting with Wolpaw's definition of a perfect BCI given in Section 1.1 (Martini et al., 2020; Ramadan & Vasilakos, 2017). One very promising medical imaging technology for measuring brain activity is magnetoencephalography (MEG). Intuitively, MEG could be seen as an alternative to EEG that measures the magnetic activity associated with brain activity rather than the electrical activity. MEG distinguishes itself from other modalities by offering excellent temporal and spatial resolution whilst being non-invasive and without any mentionable risk. These properties make MEG an incredibly attractive modality for use in BCI research. However, the device required to perform a MEG scan costs millions, is non-portable and requires high maintenance in its current form. This makes it unsuited for general BCI usage and only highly medical and temporary BCI applications have been made with it (Ramadan & Vasilakos, 2017). If MEG would ever become portable and affordable it is bound to revolutionise the BCI field. Koshev et al. (2021) explains the MEG modality in greater detail.

A technique that has seen such a decrease in price and increase in portability is functional near-infrared spectroscopy (fNIRS). As shown by Tsow et al. (2021), a comfortable and portable fNIRS headset can be made for only a few hundred euros, compared to previously available solutions which cost thousands and are less portable and comfortable. FNIRS thanks its name due to the use of a light source that operates near the infrared wavelength for detecting changes in haemoglobin species present in the brain. FNIRS does not expose the body to ionising radiation and is proven to be a safe modality for repeated use (Pinti et al., 2020). Whilst the poor temporal resolution of fNIRS limits it usability for online BCI systems where a fast response time is desired, it has been used in combination with EEG to form a successful hybrid-BCI (W.-L. Chen et al.,

2020; Martini et al., 2020). Pinti et al. (2020) explains the fNIRS modality in more detail.

2.3.2 Non-invasive EEG for BCIs

It is clear from table 2.2 that invasive modalities for measuring bioelectrical signals from the brain generaly have excellent spatial and temporal resolution. Adding to this, the invasive device is always present and ready to be used once implanted and if the implantation procedure has no side effects, the user comfort and easy-to-use factor of these approaches should also be among the best. With electrodes being physically closer to the brain, the SNR and other data quality metrics are also often superior compared to non-invasive alternatives. All of these properties point towards invasive solutions being superior to non-invasive alternatives for BCIs applications.

However, there are also some major drawbacks to invasive solutions. As discussed in Section 1.5.2, ethical questions on what happens to an invasive BCI once it eventually stops working or stops receiving support are challenging to answer. Other ethical questions present for BCIs in general, such as data privacy discussed in 1.5.5, are even more challenging for invasive approaches, as they can't simply be removed when the user wants to ensure the device is not active. Besides these ethical issues, there are also still significant medical risks and high costs associated with these invasive solutions. Whilst an automated implantation procedure is likely to reduce both costs and risks, as already demonstrated by Musk and Neuralink (2019), there is no widespread availability possible yet.

Given the risks, high costs and limited availability still associated with invasive approaches, this master thesis and most research active in the field still focus on non-invasive solutions. From these available modalities, EEG is the oldest but also the most studied. It's been proven capable of making multiple BCI applications possible as discussed throughout Chapter 1. The good temporal resolution, affordability and UX for headset installation, use and comfort make EEG attractive over other non-invasive modalities. The limited spatial resolution and poorer SNR are downsides but can be taken into account when developing a BCI system. This is what makes EEG attractive in current BCI research and why it remains so common and the modality of focus for this master thesis.

The choice between invasive and non-invasive solutions is an open discussion in the BCI field as Steyrl et al. (2016) and Waldert (2016) discuss in greater detail. Once invasive systems become safer and more affordable, only time will tell how the ethical challenges will influence a potential future dominance of invasive solutions. Given non-invasive solutions such as MEG have proven to be far superior to EEG and even some invasive alternatives, further evolution in affordability and portability of these modalities are likely to leave the debate between invasive and non-invasive solutions ongoing for a long time.

2.3.3 Available EEG measuring equipment

With EEG being one of the oldest brain signal measuring modalities, many devices for the acquisition of EEG have been introduced. Some consumer-grade EEG devices, such as the Neurosky Mindwave Mobile 2, only use a singular electrode for capturing brain signals whilst some medical-grade devices, such as the g. Pangolin High density EEG, boast up to 1024 electrodes (Ratti et al., 2017). Likewise, most consumer-grade EEG devices offer a sensible 256 Hz to 512Hz sampling frequency whilst some medical devices, such as the ANT Neuro eego my lab 256, support up to a 16000Hz sampling frequency and more. As discussed in Section 2.3.1 and 1.2.2, there are also various types of electrodes that can be used. This means that hardware capable of measuring EEG can start anywhere from a couple of hundred euros to multiple tens or even hundreds of thousands of euros.

Many comparisons between different types of measuring equipment, often with greatly differing costs, have already been made (David Hairston et al., 2014; Krigolson et al., 2017; LaRocco et al., 2020; McCrimmon et al., 2017; Nijboer et al., 2015; Pathirana et al., 2018; Ratti et al., 2017). Table 2.3 shows some of the available EEG measuring equipment and their specifications. Comparing all of these devices falls outside the scope of this master thesis but the remainder of this section will highlight the meaning of the provided specifications.

Pricing of EEG equipment

Prices of EEG measuring equipment vary a lot. Cheap consumer-grade systems like the InteraXon Muse 2 can be bought for under €300. These types of systems are often considered novelty items, as their limited electrode counts combined with poor SNR is believed to limit their capabilities as part of a BCI system. However, the InteraXon Muse 2 has been proven to have good support in Python libraries and other programming languages. This has caused many individuals and researchers to try the headset, which has resulted in a broad range of commercial BCI applications, proving both its educational and overall usefulness is greater than the specifications would suggest (Krigolson et al., 2017). Research-grade equipment like the OpenBCI Cyton, OpenBCI Cyton + Daisy and Emotiv Epoc Flex can be bought for around €2000, which includes the main board, electrodes and a headset or cap for holding those electrodes. Medical-grade EEG hardware has to have specific licenses and approvals which increases the costs significantly. Since these medical-grade devices are bought for a specific medical application and thus are often bundled with software, exact prices are rarely stated online. However, known prices for medical EEG equipment are well above €10000, with more than double that not being uncommon when paired with capable software.

Electrode type and count of EEG equipment

Cheap consumer-grade hardware often focuses on a pleasant UX rather than raw data quality. Because of this, these cheaper systems often make use of only a small number of dry electrodes in an all-in-one package. Research-grade hardware is often more flexible in this regard, with main boards that can work with both wet and dry electrodes through universal connectors that allow for many different types of electrodes to be connected. Most of these research-grade headsets even support other types of sensors, such as EMG electrodes, to be connected as well. The electrode count of these research-grade devices is often around 20, as it aims to be compatible with the international 10-20 system discussed in Section 2.4.1. Medical grade systems often aim for the highest data quality possible, using expensive golden wet electrodes. Whilst most of these systems offer interchangeable electrodes as well, they often use proprietary connectors which limits the choice of electrodes significantly. Electrode counts of some expensive medical-grade systems are also multitudes higher than those of other hardware, with 256 electrodes not being uncommon. However, this increased electrode count does not guarantee an increase in information captured, as already touched upon in Section 1.2.2. Experiments by Montoya-Martínez et al. (2021) revealed that 32 well-placed electrodes could be sufficient for most EEG applications.

Sampling frequency of EEG equipment

The sampling frequency of EEG equipment is often expressed in hertz (Hz) and corresponds to the number of samples taken from every electrode per second. When taking into account the Nyquist sampling theorem, the sampling frequency should be at least twice the highest frequency contained in the signal that is being measured (Lévesque, 2014). Considering the

Company	Product	Price estimate	Electrode count	Electrode type	Sampling frequency	Headset aesthetics	Customizability	Wireless	Python support
NeuroSky	Mindwave	€150	1	Dry	512Hz	Good	Poor	Yes	Minimal
InteraXon	Muse 2	€270	4	Dry	256Hz	Good	Poor	Yes	Good
OpenBCI	Ganglion	€800	4	Variable (active dry)	200Hz	Variable (poor)	Good	Yes	Good
OpenBCI	Cyton	€1500	8	Variable (active dry)	250Hz	Variable (poor)	Good	Yes	Good
Emotiv	Epoc Flex	€2000	32	Variable (passive wet)	128Hz	Decent	Decent	Yes	Good
OpenBCI	Cyton + Daisy	€2500	16	Variable (active dry)	125Hz-250Hz	Variable (poor)	Good	Yes	Good
Neuroelectrics	Enobio 20	€10000+	20	Wet	500 Hz	Decent	Decent	Yes	Good
Advanced brain monitoring	B-Alert X24	€10000+	20	Wet	256Hz	Poor	Decent	Yes	Poor
Wearable Sensing	DSI 24	€20000	24	Dry	600Hz	Poor	Poor	Yes	Minimal
ANT Neuro	eego mylab 256	€25000+	256	Wet	16000Hz	Decent	Poor	No	Minimal

Table 2.3: Some of the available EEG measuring equipment with their most important specifications. Some equipment consists of multiple configurable components, these are annotated as "variable" with the default option provided in brackets. Prices are estimates based on buying a minimal total configuration of the product. Data was gathered from the original manufacturer's website and community blogs. More information on the provided properties is given in Section 2.3.3.

highest frequency of the five widely recognized brainwaves discussed in Section 2.4.3 is 100Hz for the gamma waves, a minimal sampling frequency of 200Hz is common for EEG equipment. However, based on the applications, the desired sampling rate may be higher or lower (Davis et al., 2018; Jing & Takigawa, 2000). The high sampling rates boast by some medical-grade EEG equipment caries a significant computational and storage overhead making it unneeded for most but highly specific medical applications. It should be noted that some headsets, such as the Emotiv Epoc Flex and OpenBCI Cyton + Daisy, have a higher internal sampling frequency which is downsampled before transmission.

Headset aesthetics and customizability of EEG equipment

Cheap consumer-grade EEG equipment often focuses on headset aesthetics as it is an important factor for the sales of a commercial product. However, these systems are often an all-in-one solution that offers limited to no customizability. Traditionally, research-grade and medical-grade hardware have not focused on the aesthetics of the headset or cap that holds the electrodes in place but have provided more customizability. This difference in customizability was already present when the electrode types were discussed earlier in this section. With Wolpaw's definition of a perfect BCI, discussed in Section 1.1, including a requirement of being aesthetically acceptable and the increasing competition in the field, an increasing focus has been put on aesthetics for both research-grade and medical-grade systems as well. The Emotiv Epoc Flex offers a sleek black cap that has clear thought put into cable management of the wet electrodes and a nice holder for both the main board and batteries. Whilst still covering the entire scalp and thus hard to call fashionable, it is far more visually pleasing than traditional medical caps where no thought has been put into the visual design at all. The research-grade OpenBCI offerings come standard with a 3D-printed headset, the Ultracortex Mark IV. This headset has a lot of exposed wires and looks more like a prototype than a finished product. The comfort of this one-size-fits-all headset is also far from ideal. However, the 3D model is made available publicly and as a result, the OpenBCI community has provided multiple alternative designs that are far more attractive and comfortable. In terms of customizability, OpenBCI products are among the best in the field. A comfortable and aesthetically pleasing design will be important for a widespread BCI.

Other properties of EEG equipment

Many other properties of EEG equipment play an important role in deciding which one to choose. For BCI researchers, support of the system in their preferred programming language will play an important role. Most hardware providers have at least an SDK that allows for a live connection between the headset and a Python environment. Systems having only a Python compatible SDK are labelled 'minimal' in Table 2.3. Research-grade equipment often provides these SDKs in an open-source manner which allows other developers to provide additional libraries and support for the equipment as well. Equipment that has multiple supporting Python libraries, both from the manufacturer and third parties, is labelled 'good' in Table 2.3. Besides this, factors such as connectivity, guarantee policies, parts availability and more can also influence the decision of which equipment to pick.

2.4 Working with EEG

This section discusses a few guidelines and standards that should be known when working with EEG data, specifically in a BCI setting. Some details on what electrode locations correspond roughly with which part and function of the brain are also given. Some of the common methods

to induce specific electrical brain activity that can be used for decoding specific user intentions, such as motor imagery (MI) and the evocation of P300 signals, are also addressed. Finally, some common issues with EEG including common artefacts and generalisation problems are also addressed.

2.4.1 Standards and guidelines for EEG measurements

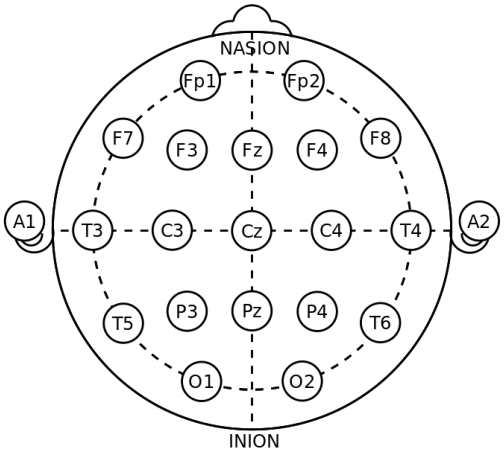
There is no singular method nor device to measure EEG as discussed in Section 2.3.1 and 2.3. Whilst this leaves freedom for both hardware manufracturers and users to find a setup that is best for them, it can be challenging to compare clinical results between setups and to make algorithms that support all possible variations. Because of this, multiple standards and guidelines have been proposed. This section summarizes the most important guidelines and consensus statements on EEG by the American Clinical Neurophysiology Society.

Sinha et al. (2016) provide guidelines on the minimal technical requirements for performing EEG in a clincial setting. Their guidelines propose the use of an EEG machine with support for at least 16 electrodes capable of measuring signals with a frequency range between 0.5Hz and 70Hz. Electrical shielding of the EEG machine is not required given the wiring for the AC within the building meets labratory standards. Other guidelines of interest for this master thesis included in this document relate to the use of standardised electrode placements and montages. The most common electrode montages were already discussed in Section 2.2 and illustrated in Figure 2.2.

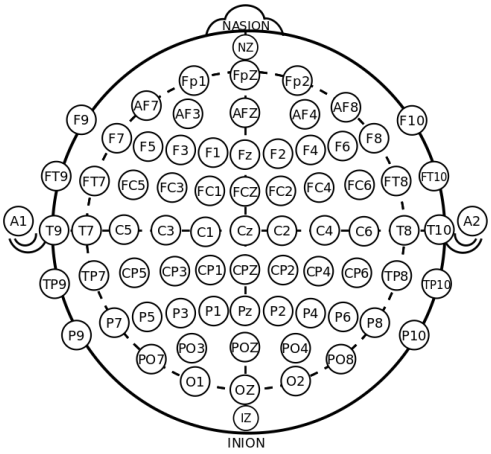
Acharya, Hani, et al. (2016) discusses the standardised electrode placements in more detail. The most commonly used scheme for electrode placement is the international 10-20 system, which was first introduced in the late 1950s (Klem et al., 1999). The 10-20 system, shown in Figure 2.3a, makes use of 21 electrodes placed at relative distances between four main positions on the head that are easily transferable between subjects. These four points are the nasion (bridge of the nose), inion (the midline bony prominence at the back of the heaf) and two preauricular points near the front of each ear. These roughly correspond with the nasion, inion, A1 and A2 annotations on figure 2.3a.

To go from these four identifiable points on the skull to the locations where electrodes should be placed, a specific process based on the proportion shown in Figure 2.4 is followed. First, a line is drawn from the nasion to the inion and marks are placed on the following proportional intervals: 10%, 20%, 20%, 20%, 20%, 10% as shown by the blue line and marks in Figure 2.4. Next, the same procedure is followed but going from the preauricular point near the left ear to the preauricular point near the right ear as shown by the red line and marks in Figure 2.4. Then, a circle is drawn with the centre mark used as the middle and a diameter that corresponds with the distance between the most outer marks on the line drawn between both ears. In this circle, a total of 10 marks are placed, each at a distance of 10% from each other. The most outer marks from the horizontal line between the ears are included, but the most outer points of the first drawn vertical line from the nasion to the inion are not. This step corresponds with the green circle and marks in Figure 2.4. After this, parasagittal measurements are made as illustrated by the yellow curves from the nasion to the onion in Figure 2.4. At a 25% interval from the outside of the parasagittal measurements, transverse measurements are made, as illustrated by the remaining yellow curves in Figure 2.4. The intersections of both parasagittal measurements and transverse measurements are the locations for the final four makers. All markers, except for the two most outer ones on the line from the nasion to the inion, represent a location for electrodes. This totals 19 locations, with the last two simply being electrodes connected to the ears.

Each electrode is then represented by a combination of letters and numbers. The numbers



(a) Electrode placement and channel names of the international 10-20 system. Free to use figure by TOMATON, Public domain, via Wikimedia Commons.



(b) Electrode placement and channel names of the modified combinatorial nomenclature for the 10-10 system. Free to use figure by Brylie Christopher Oxley, CC0, via Wikimedia Commons.

Figure 2.3: The original international 10-20 system for EEG electrode placement and its modified combinatorial nomenclature (the 10-10 system) which supports more electrodes.

indicate the side of the head, with odd numbers on the left side of the skull and even numbers on the right side, when looking from the top with the nose in front. Lower numbers correspond to electrodes closer to the midline drawn from the nasion to the inion. The names for the electrodes on this midline do not contain a number but rather end with a 'z' denoting zero. The letters are indicators of the position on the head, with F(rontal), C(entral), P(arietal), T(emporal), O(ccipital) and A which refers to the prominent bone found behind the outer ear.

Besides highlighting this international 10-20 system, Acharya, Hani, et al. (2016) also addresses the 10-10 system as a common standard for when more electrodes are desired. This system is shown in Figure 2.3b and follows a similar approach to the 10-20 system but by using smaller intervals. The other guidelines by the American Clinical Neurophysiology Society are more specific to the medical world and highlight the importance of choosing the right montage, considerations when measuring EEG with young individuals or individuals who are suspected to be brain dead, how to store EEG data and how to report on EEG measures among other things (Acharya, Hani, et al., 2016; Halford et al., 2016; Kuratani et al., 2016; Stecker et al., 2016; Tatum et al., 2016).

2.4.2 Using the anatomy of the brain for electrode subsampling

The human brain is a complex organ which has multiple functions and responsibilities. For this reason, the brain is often divided into regions based on the motor and sensory activity known to be most present in that region. These regions are illustrated in Figure 2.5. Whilst EEG has limited spatial resolution, as discussed in Section 2.2, and works with a blocked and blurred view, as discussed in Section 1.2.2, this information on brain regions can still be used in EEG-based BCI systems.

For example, much research has been put into localising the region of interest when performing MI (Amunts & Zilles, 2015; Kasess et al., 2008; Kotegawa et al., 2020; Miller et al., 2010; Sanchez-

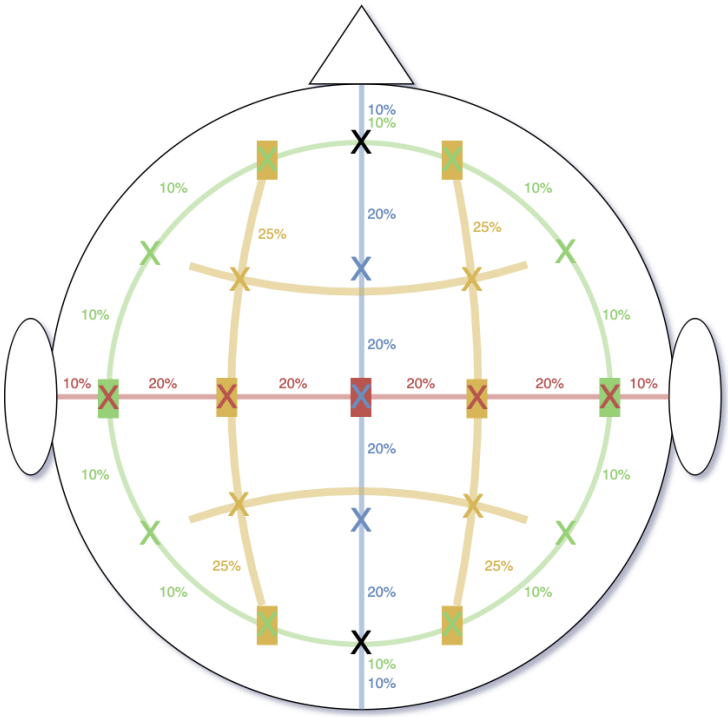


Figure 2.4: Illustration of how the placement of electrodes in the 10-20 system is determined based on the four identifiable points of the skull around the nasion, inion and ears. Black crosses are used for identification of the blue interval only and no electrodes are placed here. Crosses with a coloured background are used for determining two intervals, the cross colour and background colour intervals.

Panchuelo et al., 2012; Schnitzler et al., 1997). Whilst there is some variance between subjects and experiments, it is generally believed that most information-rich activity is centred around the precentral gyrus and premotor cortex. When using the 10-20 system, this would correspond roughly with the following electrodes: C3, Cz, C4, P3, P4, F3, Fz and F4. As discussed by Kotegawa et al. (2020), the prefrontal cortex has also been shown active during specific MI tasks and as such is sometimes also considered to be a region of interest. When using the 10-20 system this corresponds roughly to the Fp1 and Fp2 electrodes. Similar studies have been done for other types of brain activity, such as localising both visual and auditory P300 signals (Linden, 2005), the former of which is mostly active in the visual cortex. Chapter 2 of the BCI book by J. Wolpaw and Wolpaw (2012) explains these brain regions and the activity correlated to them in more detail.

Removing electrodes that are not close to the region of interest can help in reducing data dimensionality without losing too much useful information. It can also help in reducing the learning of activity that is concurrent with the desired activity. For example, EEG artefacts caused by eye movement and the movement of facial muscles produce far stronger signals than those related to a MI task. If a user is asked to perform left-hand and right-hand motor imagery tasks but twitches the eyebrow of the left or right eye corresponding to the envisioned task,

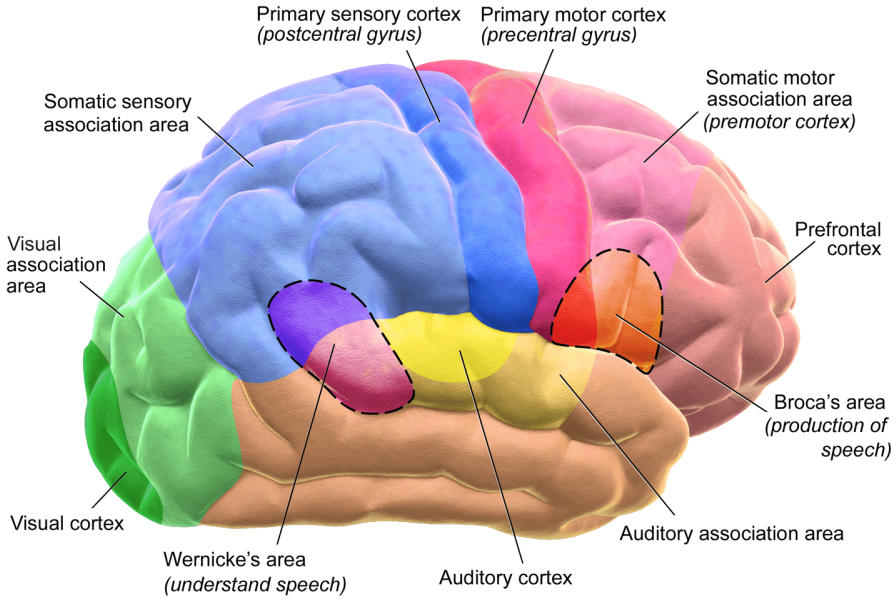


Figure 2.5: Sideview of the brain with motor and sensory regions highlighted. The left side corresponds to the back of the head, with the visual cortex being just above the back of the human neck. Free to use figure by Blaus (2014), CC BY 3.0, via Wikimedia Commons.

most algorithms will learn to detect this twitching rather than the desired problem. Whilst this can have beneficial improvements for individuals that are persistent in this behaviour, it will cause poor generalisation as it isn't learning the desired tasks. As such, the removal of the data from the frontal electrodes which capture these muscle movements might be a smart approach. However, due to the blurry view EEG receives and the poor spatial resolution combined with the smaller related activity present in other regions of the brain, removing too many electrode inputs can have a significant degrading effect on performance.

2.4.3 The five widely recognized brainwaves

Brainwaves are defined as the oscillating bioelectrical signals from the brain of which modalities such as EEG measure a superposition. There are five widely recognized brainwaves, delta (δ), theta (θ), alpha (α), beta(β) and gamma (γ) waves (Priyanka et al., 2016). These can be further divided into some low and high variants of which Table 2.4 provides the properties (Kawala-Sterniuk et al., 2021). Besides these five widely recognized brainwaves, other waves with known properties exist such as the sensorimotor mu (μ) wave with a frequency range from 7Hz to 12Hz (Kirar & Agrawal, 2018). Disproportions in these waves are the basis for multiple medical diagnostics as discussed by Kawala-Sterniuk et al. (2021).

For BCI applications, knowing which frequency range is of interest for a given task may help in preprocessing the signal such that only that frequency range is included. For example, EEG data from MI tasks is often filtered to only include signals from the alpha and beta wave and thus a frequency range of 8Hz to 30Hz (Afrakhteh & Mosavi, 2020). Sometimes the lower threshold is set to 7Hz such that the mu wave is also included (Kirar & Agrawal, 2018). However, the exact thresholds of these frequencies are known to be subject-dependent (Kirar & Agrawal, 2018).

Because of this, subject-wise finetuning may be required for obtaining optimal results.

Brainwave (Frequency band)	Frequency range	Brain states
Delta (δ)	0.5Hz-4Hz	State of deep sleep, when there is no focus, the person is totally absent, unconscious.
Theta (θ)	4Hz-8Hz	Deep relaxation, internal focus, meditation, intuition access to unconscious material such as imaging, fantasy, dreaming.
Low alpha (α)	8Hz-10Hz	Wakeful relaxation, consciousness, awareness without attention or concentration, good mood, calmness.
High alpha (α)	10Hz-12Hz	Increased self-awareness and focus, learning of new information.
Low beta (β)	12Hz-18Hz	Active thinking, active attention, focus towards problem solving, judgment and decision making.
High beta(β)	18Hz-30Hz	Engagement in mental activity, also alertness and agitation.
Low gamma (γ)	30Hz-50Hz	Cognitive processing, senses, intelligence, compassion, self-control.
High gamma(γ)	50Hz-70Hz	Cognitive tasks: memory, hearing, reading and speaking.

Table 2.4: Overview of the five widely recognized brainwaves according to data from Priyanka et al. (2016), with further subdivision in low and high brainwaves by data from Kawala-Sterniuk et al. (2021).

2.4.4 Common methods for inducing electrical brain activity

For general BCIs to consistently work with brain signals, it is often desired that the expected signals can be induced by the user. There are two possible approaches for this, either by using external stimulus to trigger evoked potentials (EPs) or by relying on specific tasks performed by the user which causes event-related synchronization (ERS) or ERD without the need for external stimulation. Figure 2.6 provides an overview of these methods.

Evoked potentials (EPs)

Evoked potential (EP) are measurable bioelectrical signals from the brain that occur after a certain stimulus is given to the subject (Abdulkader et al., 2015). An EP can be of exogenous or endogenous nature. Exogenous potentials are influenced by the physical properties of the stimulus, for example, visual exogenous potentials can be influenced by the brightness intensity, the refresh rate and more. Endogenous potentials are different as they are related to the psychological significance of the stimulus to the subject, for example, a feeling of joy for the user (Abdulkader et al., 2015; Picton, 1988).

EPs can be further divided into ERPs and steady state evoked potentials (SSEPs). ERPs can be seen as a change from previous brain activity after the user is exposed to a stimulus change. ERPs can originate from visual (visual evoked potential (VEP)), auditory (auditory evoked potential (AEP)) and somatosensory (somatosensory evoked potential (SSEP)) stimulus changes. These changes can be either positive or negative in amplitude which is reflected in the name of the potential together with the expected latency from the stimulus change. For example, the commonly used P300 potential denotes a positive electrical change around 300 milliseconds after the stimulus change. Likewise, an N400 potential denotes a negative change 400 milliseconds after the stimulus change.

SSEP are potentials which occur at a frequency that is equal to or related to the frequency of the stimulus. This SSEP can also be from visual (SSVEP), auditory (steady-state auditory evoked potential (SSAEP)) and somatosensory (steady-state somatosensory evoked potential (SSSEP)) origin. A SSEP begins and ends with the presence of the stimulus.

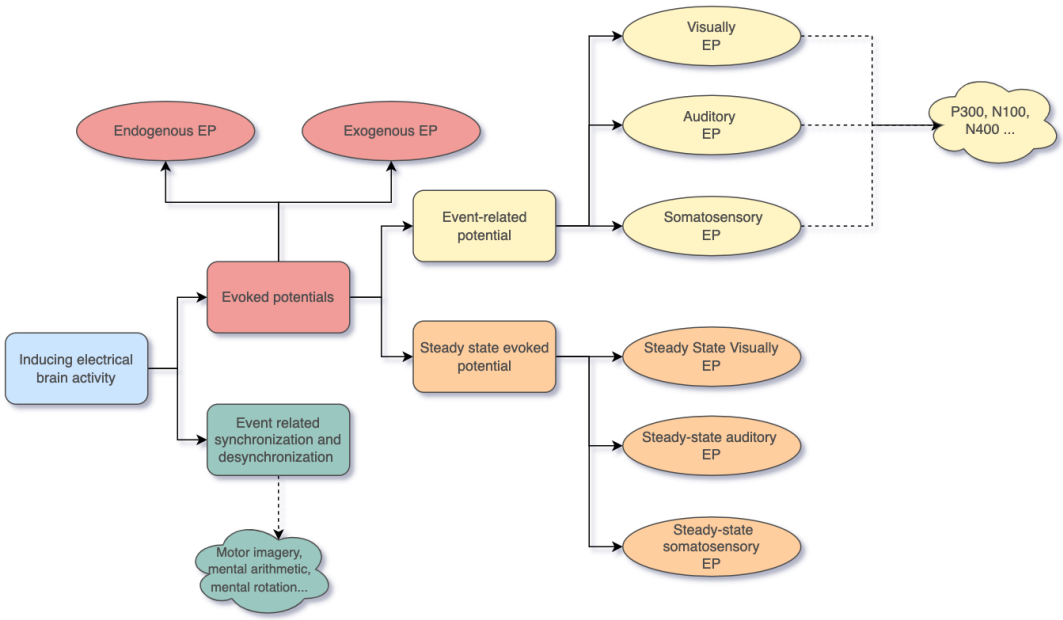


Figure 2.6: Overview of the most common methods for inducing electrical brain activity and the signals they produce.

Event related synchronisation and desynchronisation

Besides using external stimuli for inducing EP, the user can induce changes in the ongoing brain activity by performing certain tasks. These are known as ERS and ERD. As discussed by Abdulkader et al. (2015), ERS and ERD are time-locked to the event but not phase-locked, making them harder to detect and requiring non-linear methods. The intuitive idea behind ERS and ERD is that certain tasks performed by the user change the synchrony of measured neurons through EEG and thus the measured signal. If the power of this synchrony increases, the term ERS is used, whilst ERD denotes a decrease in this power.

One commonly used method for causing these types of electrical activity changes is motor imagery (MI). MI is the process in which a person imagines motor movements such as the closing or opening of the right hand. MI-based BCIs are interesting because they don't require any external stimulus nor effective motor movements. This makes MI-based BCIs applicable for many people, including those with motor disabilities. Pfurtscheller et al. (1997) were the first to experiment with the idea of using MI in an EEG classification task. Since then, many MI-based BCIs have been proposed. However, the MI capability of users can differ greatly and some may require significant training as was already discussed in Section 1.4.2. Other methods comparable to MI include mental arithmetic and mental rotation.

2.4.5 Generalisation issues of brain activity

In an ideal world, one singular model would classify MI tasks performed by any individual with high accuracy. However, intersubject classification has proven to be difficult and even intersession classification from the same subject is challenging when working with multiple classes. Most of

the generalisation issues lie in the non-linear and non-stationary nature of the brain which is reflected in the recorded EEG data. Kowalik et al. (1996) describes the non-linear nature of the brain in greater detail. As discussed by Abdulkader et al. (2015), the most important consequence of the non-linearity of the brain is that no linearity assumptions should be made when working with EEG signals and as such non-linear classification models are likely to perform better.

Compared to stationary signals whose properties do not change over time, brain signals measured with EEG do see significant change over time. This change may be caused by slight displacements of the EEG electrodes during use and varying noise but also by the working of the brain itself. Since EEG can not record the activity of singular neurons, there are always several unknown and constantly changing latent variables captured even if the signal of interest is identical. These changing latent variables can cause two identical MI tasks performed only seconds apart to look drastically different on an EEG. Even if no latent variables were present, neuroplasticity and other neurological properties of the brain can change the shape and location of the signal of interest over time. This means that the non-stationary nature of brain signals and EEG should always be taken into account when working with EEG.

Finally, the human brain can differ greatly between subjects, both anatomically and functionally. As a result, precise models learned for one subject are not guaranteed to work well on another subject in an intersubject setting. When working with ERS and ERD, there is also going to be user-variability in how these changes are induced. For example, even after sufficient training, one user may perform a MI task of the hand with the hand envisioned in front of them, whilst another may envision the movement with the hand located next to them. Such subtle changes are bound to add up and cause a change in the generated ERS and ERD. These factors significantly limit generalisability between subjects and sometimes even between sessions of the same user.

2.4.6 Common EEG artefacts

EEG artefacts can be seen as recorded signals which do not originate from the brain or the region or signal of interest. Such artefacts cause minor to significant changes from the expected EEG measurement. A distinction is made between artefacts not caused by the user, nonphysiologic artefacts, and artefacts caused by the user, physiologic artefacts. This section discusses the most important artefacts found in both categories. Chapter 8 from the book by Sazgar and Young (2019) provides a detailed overview of these artefacts and many more accompanied by visual examples of these artefacts in the time domain.

Nonphysiologic EEG artefacts

By far the most common nonphysiologic EEG artefact that should always be taken into account is the electrical and electromagnetic interferences caused by AC electrical lines and devices present in the environment. These artefacts are easily detectable in the frequency domain as there will be a great spike at either 50Hz or 60Hz dependent on the AC standard present in the geographical location of measurement. Since artefacts from this type are so common, most EEG measuring equipment has a built-in filter to eliminate this frequency.

The second most common nonphysiologic artefact is electrode pop caused by a change in potential between the scalp and an electrode. It is often short in duration and can vary greatly in appearance in both the time and frequency domain. In general, these types of artefacts are most visual in the time domain with an abrupt and unusually high amplitude change. Some EEG equipment has additional pressure sensors to notice if the location from the electrode to the scalp changes, the information of which can be used to temporarily ignore that channel. Likewise,

postprocessing may temporarily suppress electrodes with such abrupt changes in amplitude. However, certain medical conditions can have properties comparable to the artefacts caused by electrode popping and thus some applications might opt to not filter away all abrupt changes. Other nonphysiologic artefacts include additional movement of the EEG equipment such as the electrode cables as well as poorly placed electrodes or grounds and defective components in general.

Physiologic EEG artefacts

Muscle activity underneath or nearby electrodes causes significant artefacts in the EEG measurements and are by far the most important type of physiologic EEG artefacts. Whilst these types of artefacts are actual EMG measurements rather than EEG measurements and thus should not be desired in EEG, they are commonly used in BCI applications, both deliberately and accidental. As shown by McDermott et al. (2021), muscle artefacts are far easier to detect and classify then effective brain signals. As a result, systems such as the ones by W.-L. Chen et al. (2020) use these easier-to-detect EMG activities in EEG-based BCI, with repeated blinking of one or both eyes and smirk movements being most common. However, certain users suffering from medical conditions that limit control over voluntary movements of these body parts are directly excluded from using such BCI systems. As such, using these types of artefacts deliberately is often seen as a hacky solution to BCIs. Accidental use of these artefacts includes unknowingly using or learning consistent concurrent muscle movement by a user performing the goal task, as already described in Section 2.4.2.

Other movements such as eye movement, tongue movement during speech and general movements during breathing or walking also cause significant artefacts in the EEG measurements. Besides muscle artefacts in the form of EMG measurements made with EEG equipment, ECG measurements due to electrical activity from the heart can sometimes also be detected with EEG. A last important physiologic artefact is caused by sweat. Not only can perspiration cause distinct very low-frequency electrical artefacts, but it can also short electrodes when a user is sweating heavily.

2.5 Chapter conclusions

This chapter detailed the origin and function of biosignals and bioelectricity in the human body. Multiple modalities for measuring brain activity, both invasive and non-invasive, were discussed. Whilst invasive solutions were determined to be superior in all criteria related to data quality, they still impose significant risk and are not yet widely available in an affordable manner. As such, they were found unfeasible for use in current and foreseeable future general BCI applications. From the non-invasive methods, MEG was the most capable modality in terms of temporal and spatial resolution. However, MEG has an incredibly high cost and maintenance and is non-portable in its current state. Two cheap and portable non-invasive modalities, fNIRS and EEG, where also discussed. It was discussed that EEG with its great temporal resolution is most promising in the current and foreseeable future BCI application. As such, this master thesis will further focus on the EEG modality and data from it alone. Some of the available EEG equipment was discussed showing that capable research-grade EEG equipment is obtainable for around 2000 euros and usable cheap consumer-grade EEG equipment starts at only a couple of hundred euros.

Next to this technical discussion on measuring modalities and EEG equipment, some practical guidelines and pointers for working with EEG data were given. From the discussed methods for inducing electrical brain activity, ERS and ERD seem most promising as they don't require

an external stimulus. In this regard, this master thesis will focus on EEG data related to MI tasks. Lawhern et al. (2016) considers this as one of the hardest to learn BCI paradigms since the associated ERS and ERD are weak and highly variable, both across and within subjects. However, using the pointers given and knowing the common artefacts present in EEG, it is possible to develop a capable offline classifier for MI EEG data, as discussed further in the following chapters.

Chapter 3

Decoding brain signals with machine learning

3.1 Introduction

The previous chapter, Chapter 2 discussed biosignal and how bioelectricity is made in the human body. Multiple modalities for measuring brain signals were discussed and it was discussed how EEG seems like the most promising measuring modality for capturing bioelectrical signals from the brain in a BCI setting. Chapter 2 also discussed that MI is one of many methods to induce such bioelectrical signals activity in the brain, namely by causing ERS and ERD. MI and ERS/ERD were deemed interesting as it doesn't require the external stimuli that EP do and applies to many people, even those with limited mobility. However, it was also discussed that ERS/ERD following from methods such as MI require extensive user training and are harder to detect than most EP alternatives for inducing such bioelectrical signals activity in the brain.

This chapter will discuss how brain signals can be decoded using both traditional two-step ML and one-step DL in a supervised manner. Whilst Chapter 1 already highlighted multiple breakthroughs in this regard on an intuitive level, this chapter will provide more technical details. First, the general pipeline for classifying brain signals and MI EEG in particular is discussed. Then the role of ML and DL in this pipeline and some of the most important concepts from these technologies are discussed. This chapter then goes over the process of evaluating and using the created and trained pipelines. The chapter concludes by discussing some of the common issues encountered while creating these pipelines and the conclusions that can be made from this chapter.

3.2 A general pipeline for classifying brain signals

The general pipeline of classifying brain signals and EEG-based MI tasks in particular is similar to that of a CADx systems which was discussed in Section 1.2.4 and shown in Figure 1.5. Whilst the training pipeline and prediction pipeline for classifying brain signals consist of the same components, their input and output are different. For this reason, this master thesis considers them as two separate pipelines. The remainder of this section will discuss the components used in these brain signal classification pipelines and the techniques commonly used in each of these components. Figure 3.1 provides a visual overview of these pipelines and their components.

(a) Components of general training pipeline for brain signal classification annotated with common examples per component.

(b) Components of general prediction pipeline for brain signal classification annotated with common examples per component.

Figure 3.1: General training pipeline and prediction pipeline used for classifying brain signals.

3.2.1 Data source

Assuming supervised learning, a ML paradigm further discussed in Section 3.3.1, input data for the training pipelines should include both the independent variables as wel the dependent variable. When working with an EEG MI classification problem, these independent variables are the EEG measurements of each channel whilst the dependent variable is the MI task performed at a specific time point. Multiple possibilities exist for providing these variables and the link between them. Most open-source MI EEG datasets, such as the one by Kaya et al. (2018) used in this master thesis, do this by providing the EEG recordings of an entire session as a single 2D matrix (channels x measurement per time point) and the desired labels as an equal width vector containing the marker at any given time point. The time points are dependent on the sampling frequency and denote the sample number counting from the first sample of the session. The marker may be the current content of the screen which provides tasks to the user or other event-related information. Figure 3.2 combines both the independent and dependent variables into a singular visualisation.

The prediction pipeline only expects the independent variables as its task is to predict the dependent variable. In theory, these independent variables should be of the same format used during training, but in practice, they might originate from a different source or recording and as such might require additional steps during preprocessing to ensure at least an equal sampling frequency and channel distribution. This master thesis assumes the device used during training and prediction is identical with equal settings used and as such doesn't require this type of preprocessing.

The prediction pipeline may work in an offline manner, as is the case for the experiments in this master thesis, or in an online manner. When using offline prediction the data was already recorded and stored before being provided in its totality to the prediction pipeline. In an online setting, the data is streamed to the prediction pipeline as it is being measured and the prediction pipeline should merge this incoming data to an object that is usable in the next stages itself. For example, a buffer may be used to collect samples until one second is obtained and pass that to the next step. In an online setting, windowing is often directly performed on the stream, as discussed next. Other types of data formatting are possible but they should all provide the same information.

3.2.2 Windowing

The data source as described in Section 3.2.1 provides a continous signal over multiple channels. To process these signals a mechanism has to be in place so that this continuous signal is split into discrete segments. Such techniques are often referred to as windowing, but the neurophysiological field also refers to it as epoching. The latter should not be confused with the meaning of epochs in a ML setting. Different types of windowing exist and three common approaches are illustrated in Figure 3.3. Using a fixed window surrounding known event points is trivial on the training data and results in the simplest classification task with the most consistent window labels from the three windowing techniques shown in Figure 3.3. However, when trying to predict outcomes, the point at which an event occurs has to be known as well. This information may be known, for example during an offline MI classification task or when using a fixed feedback loop in an online manner where actions from the user are accepted at fixed time intervals.

A more intuitive but computationally harder windowing technique is using sliding windows. A non-overlapping sliding window technique is shown in Figure 3.3b and an overlapping technique is shown in Figure 3.3c. Both are trivial to apply to the independent variables but deciding which dependent variable should be related becomes a difficult task. How much activity from the event should be included for the training window to be considered a sample of that event?

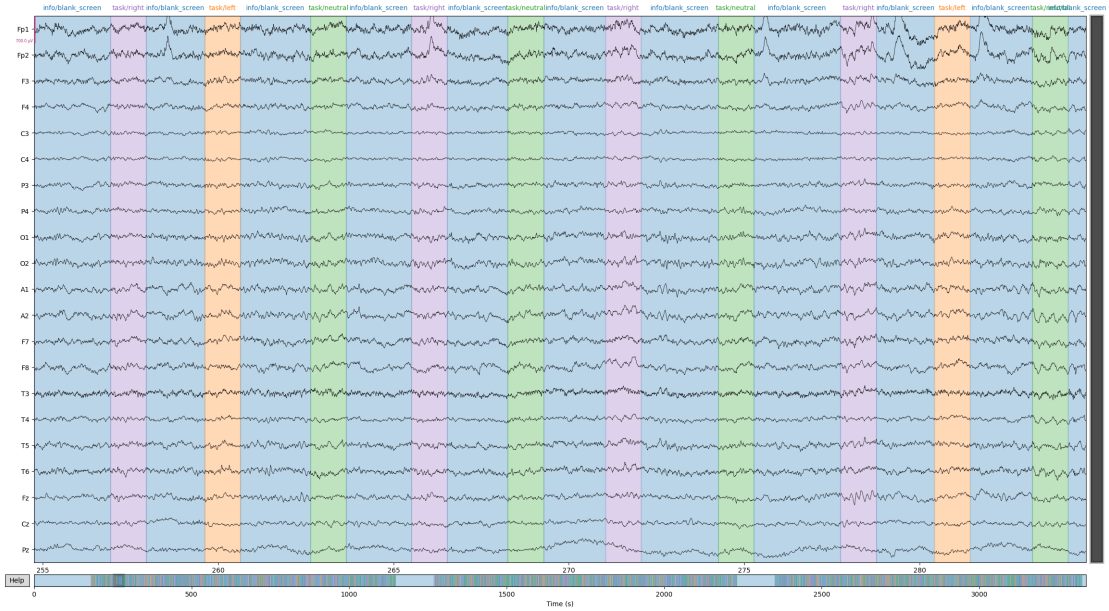


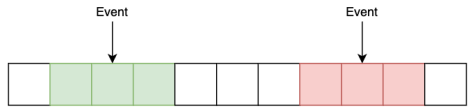
Figure 3.2: Visualisation of the EEG recording from December 15 2015 overlayed with the provided markers for subject B in the open-source dataset by Kaya et al. (2018). The x-axis depicts the time in seconds, the y-axis depicts which channel of the recording is visualised and the colour overlay represents the active marker.

What happens when a window includes two distinct events? These are questions that should be answered within the context of the application. When using sliding windows, singular labels such as "task right" may be split into "task right start", "task right hold" and "task right end".

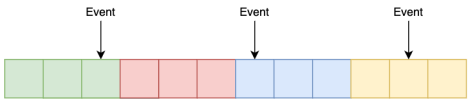
Many different windowing techniques exist that use far more complex strategies than the ones illustrated in Figure 3.3, especially for controlling the boundaries of the window. Podder et al. (2014) discuss some of these more complex windowing techniques in more detail. When done right, certain sliding window techniques can improve the performance of a MI EEG classification pipeline compared to even fixed windowing surround a known event, as shown by Gaur et al. (2021). However, starting a new sliding window at each time point may cause significant computational overhead, increasing both training and prediction time. The latter can make the prediction pipeline too complex to be run on affordable, low-energy and portable computational hardware as desired by a BCI system doing local processing. This master thesis will consider a fixed window of 0.5 seconds and 1.5 seconds surround a known event for both training and prediction.

3.2.3 Preprocessing

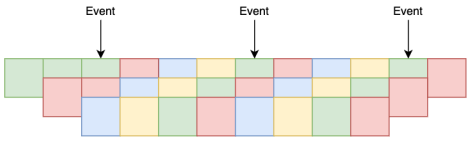
Brain signals are non-linear as well as non-stationary signals and exact execution of the performed MI tasks are bound to differ per subject as already discussed in section 2.4.5. Combining these properties with the poor SNR of EEG, as discussed in Section 2.2, means that raw EEG measurements are hard to intepret, even by machines. Luckily, as discussed by Dillen et al. (2022), a DL approach making use of sufficient layer, nodes and training should be capable of learning any mapping from input to output, including any manual preprocessing that can be



(a) Fixed window surrounding a known event.



(b) Non-overlapping sliding windows.



(c) Overlapping sliding windows

Figure 3.3: Different types of windowing techniques.

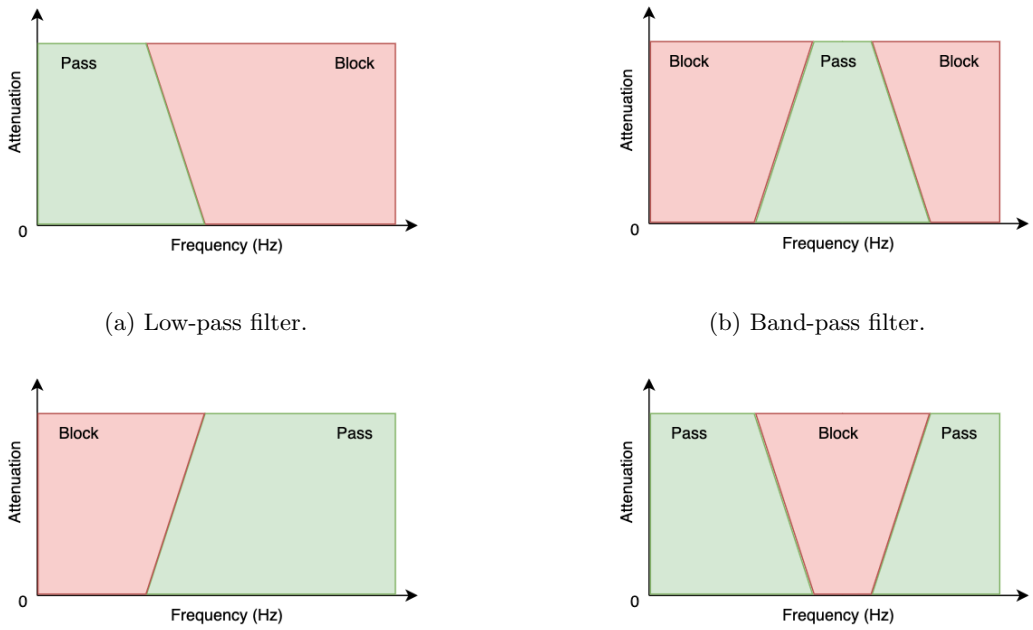
done. As such, DL approaches can often work directly on this raw EEG signal and raw signals in general. This is one of the most promising aspects of DL in multiple fields and it is the reason the one-step DL approaches from this master thesis will use no preprocessing besides the AC artefact removal that was already performed by the suppliers of the open-source database (Kaya et al., 2018). Traditional two-step ML approaches do not have this property of being able to learn any mapping from input to output and as such require at least minimal preprocessing of the data to obtain usable results. For this reason, many libraries providing the most basic EEG preprocessing steps have been developed with MNE by Gramfort (2013) being the most popular for Python and used in this master thesis. Some automated pipelines specifically for EEG preprocessing have also been proposed, such as the PREP pipeline by Bigdely-Shamlo et al. (2015).

Frequency filtering

One of the most common preprocessing operations done to EEG signals is frequency filtering. Frequency filters come in four main categories: low-pass filters, high-pass filters, band-pass filters and band-stop filters. The working of these filters in the frequency domain is shown in Figure 3.4. Again, many variants on how to exactly perform the filter exist. Some use a harsh filter with no transition band whilst others use a transition band as visualised in Figure 3.4. As further discussed by Pal (2017), a first distinction is made between finite impulse response (FIR) and infinite impulse response (IIR) filters. Most common filter operations are using a band-stop filter to cancel out the AC artefacts discussed in Section 2.4.6 and to filter out frequencies not of interest for the application. Since these operations are so common they are often included directly in EEG equipment with a hardware filter. This master thesis uses a FIR filter design using the Blackman window method in some of its experiments to band-pass filter the signal to only include the MI frequencies as discussed in Section 2.4.3. Further details of this exact filter are not of interest for this master thesis and the MNE library supplied functionality is used to obtain the desired filter (Gramfort, 2013).

Baseline correction

Another common preprocessing operation is baseline correction on a window. Baseline correction consists of taking a baseline period and determining the mean voltage of each electrode's



(a) Low-pass filter.

(b) Band-pass filter.

(c) High-pass filter.

(d) Band-stop filter.

Figure 3.4: Different types of filtering techniques.

measurements during that baseline. This baseline is most often one or more seconds before an event occurs. This mean voltage is then subtracted from the remaining signal of each respective channel. Doing this normalizes each window such that it has a centre closer to zero. This can help in reducing the non-stationary problem and is used for the traditional two-step ML experiments in this master thesis with a baseline period of one second before the event.

Channel and data downsampling or upsampling

Another type of preprocessing is channel subsampling and augmentation through interpolation. This consists of removing channels not of interest as discussed in Section 2.4.2 or adding augmented channels by clever interpolation or other approaches from the existing channels. When a certain channel is known to be bad or artefacts described in Section 2.4.6 are detected, it may be replaced by an augmented channel during the artefact period as a way of resolving the artefact. Likewise, if the sampling frequency of EEG equipment is too high or low, data subsampling or upsampling may be performed.

Other preprocessing techniques and preprocessing ordering

There exist many other preprocessing techniques for signals and MI EEG signals in particular but these fall outside the scope of this master thesis. Preprocessing can happen in other places of the pipeline than the one shown in Figure 3.1, for example, channel subsampling reduces dimensionality and is thus often done as soon as possible to limit unneeded computational overhead. It is important to note that multiple preprocessing steps may be performed in sequence. This means

that the output of a previous preprocessing step is used in the next and as such the ordering of preprocessing steps can be important depending on the techniques used in the sequence. It is also important to note that certain preprocessing techniques which learn parameters from the training data should use the same parameters for the prediction pipeline, as redetermining them on the prediction data may alter the data in a way unknown by the already trained classifier later in the pipeline. Whilst often not the case for preprocessing techniques, this is a subtility of great importance in feature engineering, the next step in the pipeline.

3.2.4 Feature engineering

Feature engineering, or feature extraction, is the process of representing raw or preprocessed data by numerical features that carry information from the original data related to the problem. The goal of feature extraction is to simplify a complex data structure by representing it as one or many features that are easier to interpret and/or learn. This step is crucial in traditional two-step ML classification approaches as even the preprocessed data is often too complex for traditional ML classifiers to directly learn from. This differs from preprocessing which aimed to improve the raw data quality, although for some of the discussed preprocessing techniques it could be debated that they are also a type of feature extraction. As already addressed in Section 3.2.3, some features rely on learned parameters, such as the $\mathbf{W}$ matrix for spatial filters discussed later in this section, these learned parameters should be reused in the prediction pipeline and not redetermined on the prediction data, as this will confuse the classification model and produce unwanted behaviour.

A simple but poor feature extraction technique for EEG data would be representing the channels' measurements not by their raw data but by their minimum, maximum, median, first quartile and third quartile, much like a boxplot would. Whilst this would be easy to interpret, it would carry too little information for a classifier to effectively learn anything but the simplest problems. Finding appropriate feature extraction techniques is a hard task which has taken years of refinement and has ongoing refinement in many fields, including EEG classification and other medical imaging fields (van Ginneken, 2017). EEG feature extraction methods can be categorized by the domain they work in, namely the time domain, frequency domain, time-frequency domain or spatial domain. The experiments in this master thesis will only consider CSP for feature extraction in the traditional ML classification approaches, a spatial filtering technique closely related to principal component analysis (PCA) used for mainly spatial domain feature extraction. As discussed in Section 1.2.4, CSP derived features are commonly used in traditional two-step ML classification and the CSP technique has seen many extensions over the years. CSP is further discussed together with its extensions in Section 4.2.

The remainder of this section briefly discusses the feature extraction possibilities in each domain but a more detailed explanation falls outside the scope of this master thesis. The reader is reffered to chapter 7 of the BCI book by J. Wolpaw and Wolpaw (2012) for an in-depth overview of many feature extraction techniques for BCI applications in far greater detail. Boonyakitanont et al. (2020) compares the performance of multiple feature extraction techniques for epileptic seizure detection using EEG. Geethanjali et al. (2012) compares multiple feature extraction methods for EEG-based BCI applications in the time domain whilst Al-Fahoum and Al-Fraihat (2014) does the same for feature extraction methods in the frequency and time-frequency domain.

Time domain feature extraction

Temporal feature extraction methods work in the time domain, where the EEG data is analysed as a time series of voltage measurements per channel. This is most likely the representation of the

data as it comes from the previous preprocessing step in the pipeline and as such doesn't require an additional transformation. Geethanjali et al. (2012) describes some relatively simple feature extraction methods in the time domain, namely the windows's mean absolute value (MAV) per electrode, the amount of times a zero crossing (ZC) and slope sign changes (SSC) occurs per channel and the cumulative waveform length (WL). When all four of these relatively simple features are combined, surprisingly good accuracies are obtained through intrasession testing (Geethanjali et al., 2012).

Many more time domain feature extraction techniques exist and some more examples are given by Boonyakitanont et al. (2020). Boonyakitanont et al. (2020) discusses some other simple features, including those used by Geethanjali et al. (2012). For example, the WL feature described by Geethanjali et al. (2012) is called *total line length* by Boonyakitanont et al. (2020) and provided in Equation 3.1. Boonyakitanont et al. (2020) also details more complex features based on the entropy of the signal such as permutation entropy (PE), approximate entropy (ApEn), fuzzy entropy (FuzzEn) and more.

$$L(X) = \sum_{i=1}^{N-1} |X[i] - X[i-1]| \tag{3.1}$$

Frequency domain feature extraction

As the name suggests, frequency domain feature extraction happens in the frequency domain. The frequency domain represents the measured EEG signals in terms of frequency rather than time as was the case for the time domain. To extract features in the frequency domain, the signal represented in the time domain must first be transformed to its frequency domain representation. The most common way of going from the time domain to the frequency domain is through the use of a Fourier transform (FT) (Bracewell & Kahn, 1966). A FT has the nice property that it can be converted back to the time domain by using the inverse Fourier transform (IFT). The theoretical FT and IFT make use of integrals from $-\inf$ to inf and as such can't be directly used on real data. Many methods have been proposed to estimate this full integral solution, with the discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) being the most common. The DFT and IDFT equations are given in Equation 3.2 and 3.3 respectively. The output of the DFT (X_k) is a complex number that represents the amplitude and phase of a sinusoidal wave, representing the signal in the frequency domain. The frequency of this sinusoidal wave is $\frac{k}{N}$ derived from Euler's formula. Binder et al. (2009) discusses the FT, IFT, DFT and IDFT in more detail. Binder et al. (2009) also discusses that the computation of DFT can be too complex for many applications and introduces the fast Fourier transform (FFT), a faster variant of DFT. FFT is the algorithm that is most commonly used for effective conversion between the time and frequency domain in computer applications.

$$X_k = \sum_{n=O}^{N-1} x_n e^{\frac{-2\pi i k n}{N}} \tag{3.2}$$

$$x_n = \frac{1}{N} \sum_{k=O}^{N-1} X_k e^{\frac{2\pi i k n}{N}} \tag{3.3}$$

Once the signal is represented in the frequency domain, most of the frequency domain features are extracted from the power spectral density (PSD). The PSD is a visualisation of the power levels of the frequency component present in the signal. Figure 3.5 show the PSD of the EEG signals previously shown in Figure 3.2. Note the sharp trough present at 50Hz due to the AC

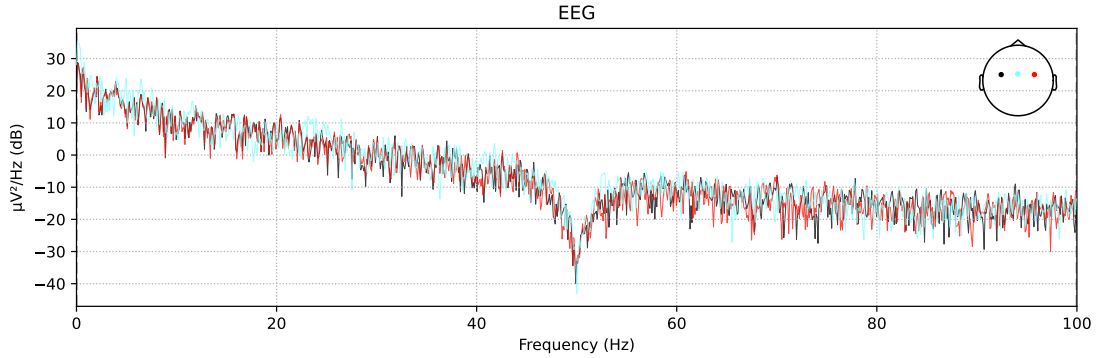


Figure 3.5: Power spectral density (PSD) for the EEG signal previously shown in Figure 3.2. Only the C3, Cz and C4 channels are shown.

artefact removal that is done by a band-stop filter at that frequency. As discussed further by Boonyakitanont et al. (2020), the features extracted from the PSD include the energy, intensity weighted mean frequency (IWMF) and intensity weighted bandwidth (IWBW) among others.

Time-frequency domain feature extraction

Whilst it is possible to extract features from both the time and frequency domain and rely on the classifier to learn a mapping between them, attempts have been made at combining both domains. For example, when further windowing the time domain signal into short time segments and using a FT on those shorter time segments, a short-time Fourier transform (STFT) is obtained. STFT assumes that those shorter time segments are stationary, even if the complete signal from which those time segments are taken was non-stationary. This allows for using frequency-domain feature extraction on those shorter FT signals. This will result in features that represent frequency domain characteristics but are ordered in a time respective order. These features could then be seen as individual data points in the time domain and thus can be further processed with time-domain feature extraction techniques.

However, the use of STFT in EEG applications is limited as finding a balance between long enough time segments such that the FT is meaningful but short enough so that enough time domain information is retained has been proven challenging. A better alternative is using wavelet transformation (WT), a technique that has been proven powerful in image compression and many other fields (Antoine, 2003). Compared to the FT that works with sinusoidal waves, WT works with wavelets. These wavelets are characterised by their scale and location. The scale relates to the frequency domain whilst the location relates to the time domain, showing a clear link with the time-frequency domain. Compared to STFT, WT does not make any assumptions about stationarity. The book by Mallat (2009) discusses WT in detail. The technical details and related feature extraction methods of which fall outside the scope of this master thesis.

Spatial domain feature extraction

As discussed in Section 2.3.1, EEG has a relatively poor spatial resolution. This would suggest that feature extraction based on the spatial domain is bound to perform poorly as well, bet this is not the case. Many BCI systems rely on features extracted from EEG signals that were spatially filtered with techniques such as PCA, independent component analysis (ICA) and CSP

(J. Wolpaw & Wolpaw, 2012). To understand spatial filters, it should be repeated that EEG channels don't represent the value of a singular electrode but rather the output of applying a differential amplifier on two channels, as was discussed in Section 2.3.1. If all these channels have one electrode in common, an alternative set of channels can be created by weighing and combining the original channels (J. Wolpaw & Wolpaw, 2012). This is the main idea behind spatial filtering and can be mathematically represented as the matrix equation shown in Equation 3.4. In this equation, $\mathbf{X}$ represents the 2D matrix of N channels and P samples. $\mathbf{W}$ represents the weight matrix with M spatial filters and N channel weights. Finally, $\mathbf{Y}$ represents the alternative set of channels with M spatially filtered channels and P samples.

There are both data-independent and data-dependent techniques for determining the weight matrix $\mathbf{W}$ of Equation 3.4. In some regards, the average reference montage shown in Figure 2.2 can be considered a data-independent spatial filter, as it uses a fixed, data-independent $\mathbf{W}$ matrix, namely one that averages all channels. A more complex but comparable data-independent spatial filter is using one of many variants of surface Laplacian spatial filtering. Intuitively, these types of Laplacian filters use only the electrodes at a certain spatial distance for determining the average. Simple variants may consider each electrode in the region of interest equally important whilst others might use a weighted average based on the spatial distance of the electrode. Surface Laplacian spatial filtering has proven successful in improving the source localisation capabilities of EEG by filtering out the signals present in multiple electrodes and keeping only those unique to the electrode of interest (Liu et al., 2020; Srinivasan, 1999; Srinivasan et al., 1996).

$$\begin{bmatrix} y_{11} & y_{12} & \dots \\ \vdots & \ddots & \\ y_{M1} & & y_{MP} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \dots \\ \vdots & \ddots & \\ w_{M1} & & w_{MN} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \dots \\ \vdots & \ddots & \\ x_{N1} & & x_{NP} \end{bmatrix} \tag{3.4}$$

However, by far the most commonly used spatial domain feature extraction happens on data after a data-dependent spatial filter such as PCA, ICA or CSP has been applied. These data-dependent spatial filters optimize the $\mathbf{W}$ matrix from Equation 3.4 based on measured EEG signals. PCA creates a $\mathbf{W}$ matrix such that the resulting signal has the highest proportion of amplitude variance from the original signals matrix $\mathbf{X}$ (J. Wolpaw & Wolpaw, 2012). Whilst this can be successful, PCA only uses the values from the independent variables for creating the $\mathbf{W}$ matrix and as such the created matrix is not optimized for best discrimination of the dependent variable. Imagine for example a $\mathbf{X}$ matrix that was not preprocessed to remove the AC artefacts described in Section 2.4.6. Since this line current noise is so present, PCA will consider it as one of the most important PCA components even though it carries no information for the problem. ICA could be seen as a variant of PCA that wishes to create a $\mathbf{W}$ matrix such that the resulting signal consists of independent channels. However, as described by J. Wolpaw and Wolpaw (2012), using ICA forms multiple challenges in BCI applications due to its computational cost and often imperative way of determining the number of independent channels to be found. Besides this, ICA also only rely on the independent variable and as such is not further considered for this master thesis.

CSP is a type of spatial filtering closely related to PCA that does use the dependent variable to create a set of new channels that are optimized to solve the problem. This makes CSP incredibly powerful and studied in BCI research, as already addressed in Section 1.2.4. CSP will be discussed in further detail together with its extensions in Section 4.2. J. Wolpaw and Wolpaw (2012) provides a more in-depth overview of all three of these spatial filtering techniques.

The promise and downfall of feature extraction

Manual feature extraction is done in what this master thesis refers to as traditional two-step ML. These traditional two-step ML approaches have some attractive properties but also some fundamental limitations. Manual feature extraction such as the one discussed in this section mostly tries to implement human knowledge about a problem as an algorithm. Features that are derived are those that an expert deems fit for describing and learning the problem. This is great in terms of explainability and interpretability, as combined with the right ML classification algorithm such as random forest (RF) it can give direct insight into how a prediction pipeline came to its conclusion. This is desired in many fields, especially the medical one. As such, traditional two-step ML still has significant ongoing research as is shown by the many alternatives on CSP that have been proposed, as further discussed in Section 4.2.

However, as discussed by van Ginneken (2017) for medical lung imaging applications, DL that learns some form of features from data in an automated manner outperform traditional two-step ML classification in many tasks. This lies in the power of DL being able to learn features that might be unknown by experts and thus can't be encoded for manual feature extraction. Research in disease detection, such as Attia et al. (2019) findings of using a CNN to detect atrial fibrillation (AF) from normal sinus rhythm ECG has shown just that. Given the human understanding of the brain is still limited as discussed in Section 1.4.6, it is likely that the right DL models will learn features that are far more descriptive of the problem but not yet understood or discovered by experts. This is also likely the reason that CNN based models, such as those by Lawhern et al. (2016) and Schirrmeister et al. (2017), have matched or even outperformed the capabilities of state-of-the-art traditional two-step ML approaches that have taken years of refinement. The main limiting factor for DL is the lack of explainability and interpretability since these approaches are seen as black box approaches. However, as 4.3.4 will discuss in more detail, attempts have been made at offering some insight into the working of these models. These attempts at clarifying what a DL model learns can provide experts with a deeper understanding of the problem (Goodfellow et al., 2016).

3.2.5 Classification

The last step of the general pipeline for classifying brain signals is the classification itself. In the training pipeline, the classification model used will be trained by providing it both the independent variables processed in the previous pipeline steps and the dependent variables. The goal of the classification model is to learn a mapping from those independent variables to the dependent variable. As discussed, the independent variable is either the features extracted from preprocessed data in case of traditional two-step ML or the minimally preprocessed raw EEG signal in case of one-step DL approaches. The dependent variable is the label of the MI task that is being performed. The prediction pipeline uses this trained model to predict the dependent variable based on only the independent variables.

A regression pipeline for processing brain signals would differ in this part, as it would not output the label of MI task performed but a continuous value. Classification and regression are types of supervised learning, the ML paradigm assumed by the proposed pipeline. The difference between regression and classification and why classification is more popular for BCI applications relying on MI EEG data is further discussed in Section 3.3.1.

Many different classification models exist with the popular traditional ML algorithm providing Python library scikit-learn (sklearn) by Pedregosa et al. (2018) having over ten different ML classifiers each with their own tunable parameters available. For DL classification models, Keras and Tensorflow (Python DL libraries by Chollet et al., 2015; Martín Abadi et al., 2015) provide tens of layers that all can be combined to create an almost endless amount of unique classification

models. Countless combinations of both these traditional ML and DL approaches have been tried in EEG-based BCIs. As such, discussing all these different classification models falls outside the scope of this master thesis. The interested reader is referred to both the original review paper on classification models for EEG-based BCIs by Lotte et al. (2007) and the updated version (Lotte et al., 2018).

The traditional two-step ML classifiers used in the experiments of this master thesis will be discussed in more detail in section 3.3.3. In particular, linear discriminant analysis (LDA), C-support vector classification (SVC) (based on SVM) and random forest (RF) are discussed. Likewise, section 3.3.4 will discuss the most important concepts of DL used for the DL experiments in this master thesis.

3.3 Machine learning and deep learning for classification

Machine learning is a broad field in computer science that focuses on algorithms aiming to relate specific data with a task at hand. Many different ML paradigms exist, with this master thesis focusing on supervised learning with classification in particular. Throughout this master thesis, the terms *traditional two-step ML* and *one-step DL* were used quite often. This section will explain how DL is a subfield of ML and the most important difference between DL and ML are given. Three traditional ML classifiers are discussed in more detail: LDA, SVC and RF. The most important concepts of DL for this master thesis will also be discussed, including fully connected layers (as seen in ANNs), convolutional layers and pooling layers (as seen in CNNs) along with various important concepts such as dropout, batch normalization and activation functions. For even further insights, the interested reader is referred to the ML book by Burkov (2019) and DL book by Goodfellow et al. (2016) which cover all ML and DL concepts needed for this master thesis and much more. The practical book on both traditional two-step ML model development using sklearn and one-step DL model development using Keras and Tensorflow by Géron (2019) covers all practical aspects needed to understand the Python implementations of the experiments performed in this master thesis.

3.3.1 Machine learning paradigms

In ML four major categories of learning are distinguished: supervised, unsupervised, semi-supervised and reinforcement learning. The following will briefly discuss the difference between these four major ML paradigms.

Supervised learning

In supervised learning, the dataset used for learning consists of N *labeled samples*: $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$. The feature vector $\mathbf{x}_i$ contains the independent variable(s) and y_i contains the label of a specific sample i. The feature vector can contain all sorts of information in all sorts of data structures but the order of elements in the feature vector $\mathbf{x}_i$ must be respected for all samples. For examples, if the second feature of sample i in the feature vector represents the age of a subject i ($\mathbf{x}_i^2$ = age of subject i), it should also represent age for any other subject j ($\mathbf{x}_j^2$ = age of subject j). The label y_i can also be any data type, although it most belongs to a finite set of classes (e.g. whether an email is considered spam or not expressed as a string or integer) or an infinite set of continuous values (e.g. a decimal denoting the risk of having a disease).

Within supervised learning, a further division can be made between classification and regression. Classification consists of predicting the class of a sample (y_i is a class) based on its feature vector ($\mathbf{x}_i$). Regression consists of predicting the real-valued label (y_i is a continous value) based

on its feature vector ($\mathbf{x}_i$). In both cases, the goal of supervised learning is to find a model that maps the feature vector $\mathbf{x}_i$ to the corresponding label y_i as good as possible.

Supervised learning is one of the most commonly studied ML paradigms, especially for EEG-based BCI where it is almost exclusively used for classification (Dillen et al., 2022). In the proposed pipeline for MI EEG data classification from Section 3.2, the feature vector contains $\mathbf{x}_i$ either the features extracted in the feature extraction phase when using two-step ML or the minimally preprocessed EEG signal when using one-step DL. One sample i corresponds with the data of one window. The classes possible for y_i are the MI tasks that the user is allowed to perform.

Unsupervised learning

In unsupervised learning the dataset only contains the feature vector of samples, it is a collection of unlabeled samples: $\{(\mathbf{x}_i)\}_{i=1}^{N}$. Multiple goals for unsupervised learning exist, with clustering being the most common. In clustering, the model is asked to find groupings (clusters) in its data and return the group (cluster) a new sample belongs to. In a clustering task, the model is expected to find patterns in the data which might reveal interesting outcomes. Other common unsupervised learning tasks include dimensionality reduction and outlier detection amongst others (Burkov, 2019). Unsupervised learning is most commonly used in the BCI field for calibrating an already trained model to a new subject or session without requiring explicit calibration from the user where the user is asked to perform a specific task. This is a promising technique in the BCI as it uses the data already collected by a BCI to further improve itself. Hubner et al. (2018) reviews and tests this technique in an ERP-based BCI setting.

Semi-supervised learning

The above-given example of using unsupervised learning to perform calibration without requiring specific actions from a BCI user shows that unlabeled data can enhance the quality of what is otherwise a supervised learning problem. Semi-supervised learning builds on the same idea, both labelled and unlabeled samples are provided for learning what is otherwise a supervised learning goal. This is especially handy in cases where a large amount of unlabeled data is available but only a small amount of labelled data is present. Since this is the case for most applications of the BCI field, this type of learning has seen a growing interest in the field (Ko et al., 2022; Schwarz et al., 2019; Y. Wei et al., 2022). These approaches deliver promising results compared to traditional supervised learning approaches with similar classification results but requiring fewer training samples and thus shorter calibration. However, more research is needed to ensure the reliability of these methods in learning useful properties from the provided unlabeled data.

Reinforcement learning

RL is a completely different approach to ML as those described before. In RL an *agent* ("an algorithm") lives in the real world or a simulation. The world the agent lives in, albeit simulated or not, is called the environment and the agent can perceive this environment through a feature vector that is called a *state*. The agent can then perform *actions* in the environment which may change the state. After a sequence of actions, the agent receives a *reward*, which is comparable to a label and can be dependent on multiple factors including the final state and the time it took to get there. The goal of the agent is to learn a policy for choosing the best action for each possible state which yields the best reward. Whilst RL is an interesting approach that has yielded algorithms which outperform human experts in multiple games such as Go (Silver et al.,

2016), its usability in the BCI field is limited as of now and as such this paradigm is not discussed further.

3.3.2 Traditional two-step ML vs one-step DL classifiers

Throughout the discussion of the general brain signal classification pipeline in Section 3.2, it was discussed how DL approaches differ from traditional ML approaches by being able to learn directly from the raw, or minimally preprocessed, EEG signal. Since DL is a subdivision of ML, the difference between the two approaches is often emphasized by calling these ML classifiers that require explicit feature extraction *tradition two-step ML* whilst DL approaches who don't require this step are referred to as *one-step DL*.

As described by Sonoda and Murata (2017), DL models are universal function approximaters. This is the foundational reason why DL can learn any preprocessing or feature extraction step from labelled data, given sufficient layer, nodes and training as discussed in Section 3.2.3. Thus, DL models learn some form of feature extraction during training. As discussed in Section 3.2.5 this makes it possible to gain additional knowledge of a problem by interpreting these features created by the DL model. Whilst this is a non-trivial task due to the black-box property of DL, it has been achieved for multiple types of DL models in multiple fields (Buhrmester et al., 2019; Schirrmeister et al., 2017; Sheu, 2020). For some of the experiments in this master thesis, the CNN based EEG classification models DeepConvNet and ShallowConvNet by Schirrmeister et al. (2017) are used. For these models, a visualisation of their learned features exists, as further discussed in Section 4.3.4.

3.3.3 Common traditional two-step ML classifiers

As discussed in Section 3.2.5, many different tradition two-step ML classifiers exist. The experiments in this master thesis make use of three such classifiers: LDA, SVC (based on SVM) and RF. This section will introduce these ML techniques in what follows. For more theoretical details, the interested reader is referred to the work by Burkov (2019). For more practical implementation details, Géron (2019) provides a discussion on how to use these models with sklearn, the same Python library used for the experiments in this master thesis (Pedregosa et al., 2018).

Linear Discriminant Analysis (LDA)

As the name suggests, LDA uses linear decision boundaries to classify data points. The most attractive feature of LDA is the fact that it has no hyperparameters to tune. For this reason, LDA is often used when already hyperparameter tuning many parameters for the feature extraction components, as eliminating classifier finetuning can speed up to process significantly. Adding to this, LDA is inherently multiclass, has an easy-to-compute closed form solution and has been shown to work well with CSP-based feature extraction (Alotaiby et al., 2017; Shang-Lin Wu et al., 2013). These are the exact reasons LDA is used in this master thesis when using complex hyperparameter tuning setups for CSP and frequency filtering. However, LDA assumes that the features from each feature vector $\mathbf{x}_i$ have a normal (Gaussian) distribution and thus the data is a multivariate Gaussian. It also assumes that each feature $\mathbf{x}_i^j$ has a comparable variance around their mean on average. These assumptions are not super strict and as discussed, it has been shown that LDA works well with CSP-based feature extraction (Alotaiby et al., 2017; Shang-Lin Wu et al., 2013).

LDA was first proposed by Fisher (1936) as a dimensionality reduction technique. It is inspired by PCA but makes use of both the feature vector $\mathbf{x}_i$ and class labels y_i, where PCA only used the feature vector $\mathbf{x}_i$. The dimensionality reduction of LDA works by maximizing the

ratio of the intra-class scatter, given in Equation 3.5, to the inter-class scatter, given in Equation 3.6 (Li et al., 2006). In these equations, n denotes the number of classes, c_i denotes the class i, m_i denotes the number of training samples for class i, $\bar{\mathbf{x}}_\mathbf{i}$ is the mean for class i, and $\bar{\mathbf{x}}$ is the total mean vector derived from Equation 3.7. From these values, the linear transformation Φ can be derived by solving the eigenvalue problem shown in Equation 3.8. This linear transformation (Φ) can then be used for predicting the class c_{pred} of a new sample $\mathbf{x}_{new}$ in the transformed space by using any arbitrary distance measure d in equation 3.9. Li et al. (2006) discusses LDA classification in further detail.

$$\hat{\Sigma}_w = S_1 + ... + S_n = \sum_{i=1}^{n} \sum_{x \in c_i} (\mathbf{x} - \bar{\mathbf{x}}_\mathbf{i})(\mathbf{x} - \bar{\mathbf{x}}_\mathbf{i})' \tag{3.5}$$

$$\hat{\Sigma}_b = \sum_{i=1}^{n} m_i (\bar{\mathbf{x}}_\mathbf{i} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{i} - \bar{\mathbf{x}})' \tag{3.6}$$

$$\bar{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^{n} m_i \bar{\mathbf{x}}_\mathbf{i} \tag{3.7}$$

$$\hat{\Sigma}_b \Phi = \lambda \hat{\Sigma}_w \Phi \tag{3.8}$$

$$c_{pred} = \arg\min_j d(\mathbf{x}_{new}\Phi, \bar{\mathbf{x}}_j \Phi) \tag{3.9}$$

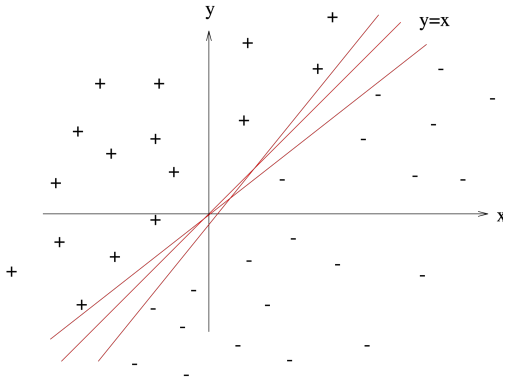
Support vector machines (SVM)

SVM are a popular supervised ML algorithm for classification (SVC), regression, outlier detection and more (Cristianini & Ricci, 2008). The fundamental idea behind SVM was first introduced by Boser et al. (1992) but has since seen some changes to become the algorithm it is known as today (Chervonenkis, 2013). The general ideas of using SVM as a classifier will be discussed here. For a more detailed explanation or more information about the use of SVM for regression and other tasks, the reader is referred to the work by Cristianini and Ricci (2008).

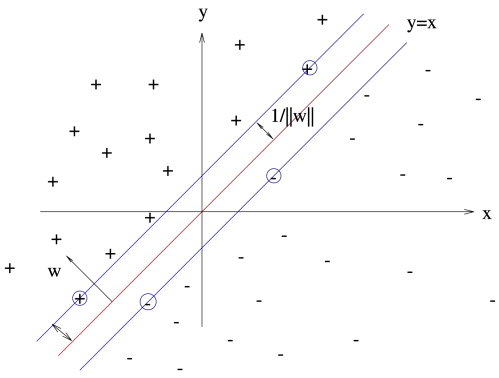
Just like LDA, SVM also uses linear decision boundries, however, it is not inherently multiclass like LDA. multiclass classification using SVM can be established using a wide variety of methods, such as a one-vs-rest scheme or a one-vs-one scheme. The sklearn implementation of the SVM classifier, SVC, uses a one-vs-one scheme (Pedregosa et al., 2018). This means that multiple SVM are made, one for each possible pair of classes, and the prediction of a new sample is made based on a (potentially weighted) majority vote.

There are three important concepts to address about SVM classifier: linear boundary optimisation, soft margins and the kernel trick. When a problem is linearly separable there exists not one but many linear decision boundaries, as is shown in Figure 3.6a. Whilst intuitively one decision boundary shown in Figure 3.6a might be better than another, they all separate the training data equally well and without making any assumptions, no *best* decision boundary can be chosen.

SVM assumes that the best decision boundary is the one which maximizes the distance from the *support vectors* to the decision boundary. These support vectors are the points closest to the decision boundary. Figure 3.6b shows the decision boundary chosen by SVM based on the maximized distance from the support vectors circled in blue to the decision boundaries shown in red. In higher dimensionality, this decision boundary becomes a hyperplane which can be described by Equation 3.10 where $\mathbf{W}$ is the weight vector normal to the hyperplane and b is the



(a) Multiple equally good linear decision boundaries for the train data. Free figure by Sylenius, CC BY-SA 3.0, via Wikimedia Commons.



(b) The decision boundary chosen by SVM is based on the maximized distance to the support vectors. Free figure by Sylenius, CC BY-SA 3.0, via Wikimedia Commons.

Figure 3.6: A linearly separable problem with multiple possibilities for decision boundaries.

bias. The distance from a point $\mathbf{p}$ to this decision boundary can be calculated using Equation 3.11. $||\mathbf{w}||$ denotes the euclidean norm of the weight vector $\mathbf{W}$.

Since we assume a linear separation is possible, we know that for the training samples $\mathbf{x}_i$ with binary class label y_i that Equation 3.12 holds. Since scaling $y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b)$ won't change the position of the hyperplane, it is possible to state that for a certain set $\mathbf{W}$ and b the minimal value is 1 since the result will always be different from zero due to the linear separability assumption and the possibility of scaling. As such, the distance to the hyperplane at the support vectors can be simplified from $\frac{y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b)}{||\mathbf{w}||}$ to $\frac{1}{||\mathbf{w}||}$. Maximizing $\frac{1}{||\mathbf{w}||}$ corresponds to minimzining $||\mathbf{w}||$ which is computationally easier and what is done by the SVM algorithm in an iterative way.

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{3.10}$$

$$\text{Distance to point } \mathbf{p} = \frac{\mathbf{w}^T \cdot \mathbf{p} + b}{||\mathbf{w}||} \tag{3.11}$$

$$\mathbf{w}^T \cdot \mathbf{x}_i + b \begin{cases} < 0 & \text{if } y_i = \text{-1} \\ > 0 & \text{if } y_i = 1 \end{cases} \iff y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) > 0 \tag{3.12}$$

With the decision boundary in place, making predictions is trivial. However, the calculations shown all assume a perfectly linearly separable problem, which is often not the case with real-world data and especially not with EEG data. When making this assumption, the SVM approach is reffered to as *hard margin* SVM classification. To allow for classifying in non-linearly separable problems, *soft margin* SVM classification has to be used. Soft margin SVM classification assumes the samples that stop the linear separation from being possible are noise and can be ignored. Mathematically this is done by changing the minimum condition from $y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) = 1$ to $y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) = 1 - \xi_i$. ξ_i is called the slack variable and Cristianini and Ricci (2008) explains in greater detail how it can be controlled by the hyperparameter C, which makes a tradeoff between the maximizing the margin and the number of misclassification on the training data.

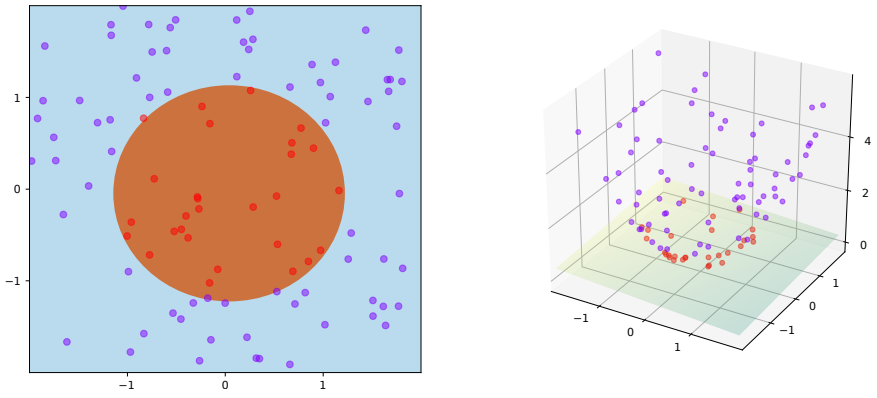


Figure 3.7: An intuitive example of SVM's kernel trick where a 2D non-linearly separable problem is transformed into a 3D problem where it is linearly separable. Free to use Figure by Shiyu Ji, CC BY-SA 4.0, via Wikimedia Commons.

Whilst the slack variable allows non-linearly separable problems to be learned by SVM, it assumes that the classification errors it makes are noise in the data. As such, it only works well for data that is *almost* linearly separable. However, SVM has a clever mechanism for reducing non-linear problems to (almost) linear problems. By finding a higher dimensional representation of the data, SVM can reduce a non-linear problem to a linear one. This is known as the *kernel trick* and an intuitive example is shown in Figure 3.7. The specific transformation done in this figure is from 2D to 3D with the kernel function shown in Equation 3.13. In practice, much more complex kernels are often used, with the most common ones being the linear kernel, the polynomial kernel and the radial basis function (RBF) kernel. From these, the RBF kernel is the most notable as its feature space has an infinite number of dimensions (Burkov, 2019). Equation 3.14 shows the RBF kernel function. By tuning the hyperparameter σ, the smoothness of the decision boundary in the original space can be controlled.

$$\Phi((x_1, x_2)) = (x_1, x_2, x_1^2 + x_2^2) \tag{3.13}$$

$$\Phi((\mathbf{x}, \mathbf{x}')) = \exp(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{2\sigma^2}) \tag{3.14}$$

Random forest (RF)

To understand the RF classifier, decision trees, bootstrapping, aggregation and ensemble learning should first be understood. Decision trees are a type of classifier that learns how to best split the training data based on certain feature value comparisons. For example, in a dataset of healthy patients and patients with lung cancer, it might be found that the best first split to make for dividing the data into two most discriminative sets is splitting on whether the subject is younger or older than 60. As subsequent splits happen, a tree-like structure is made where every split has a rule, a *decission* and where steps in a particular direction are called branches.

The branches eventually lead to a collection of training data each from the same class. At this point, further splitting of this branch stops and this last part is called a leaf. When predicting using the decision tree, each subsequent condition is checked and the tree is decent in the right direction. When reaching a leaf, the model returns the class label that is most dominant among the training samples in that leaf. When no additional restrictions are in place, a decision tree can be made that is so big and precise that all training samples are classified correctly, assuming no two training samples are identical but from another class. Since this is a clear sign of overfitting, common methods for limiting this behaviour include having a fixed maximum depth (amount of consecutive splits), a minimal amount of samples needed to perform an additional split and more. The criteria used to determine what split to make is dependent on the implementation, Javed Mehedi Shamrat et al. (2022) compares to the most important ones: ID3, C4.5, and CART.

Decision trees on their own are often limited in performance when applied to a complex dataset. RF works by combining many different decision trees to form a forest, where each decision tree is trained using a bootstrapped version of the training data and an often limited random subselection of features. Bootstrapping is the simple technique of building a new dataset based on randomly picking examples with replacements from the original training set. This results in distinct trees, as each tree had different training sets and different features to work with. In essence, a complex problem is divided into many simplified ones. having 100 to 2000 decision trees in a random forest is not uncommon. To decide the final classification, each decision tree's output is collected through a process called aggregation where a majority vote or other technique can be used to determine the final individual classification output of the RF. As RF uses both bootstrapping and aggregation, it is often called a *bagging* method. Since it uses many weaker models to create a strong model, it is often called an *ensemble* method.

3.3.4 Common one-step DL classifier layers

Deep learning (DL) is a subset of ML that can be described as any artificial neural network (ANN) model that has at least three layers. ANN, often simply reffered to as NN, are named after the biological neural network present in the brain and in part discussed in Section 2.2. Where it was discussed the biological neural network consists of many connected neurons that can activate each other in varying strengths based on the number of neurotransmitters released, an ANN consists of neurons connected through edges which have a learnable weight associated that specifies how much one node activates another. Figure 3.8 visualizes one of the most basic NN possible, a multilayer perceptron (MLP) with three layers, the minimum amount to be considered a DL model. As discussed in Section 3.3.2, NN are universal function approximators. For the MLP shown in Figure 3.8, it is a nested function where each layer is seen as a different function. As such, this MLP can be mathematically written as Equation 3.15, where y denotes the scalar returned by the singular output node. The function of the first two layers, $\mathbf{f}_1$ and $\mathbf{f}_2$, are vector functions as these layers consist of multiple neurons. This mathematical notation is further detailed by Burkov (2019).

$$y = f_{NN}(\mathbf{x}) = f_3(\mathbf{f}_2(\mathbf{f}_1(x))) \tag{3.15}$$

Sometimes DL networks are reffered to as *deep* or *shallow*, as is the case for the DeepConvNet and ShallowConvNet CNN based classificiation models considered for the experiments of this master thesis. These definitions of shallow and deep NN are rather ambiguous as following the provided definition of DL, any NN with more than three layers is considered a deep one. Concerning the models used in this master thesis, the difference between a shallow and a deep NN is that the latter has a significantly large amount of layers whilst the former doesn't.

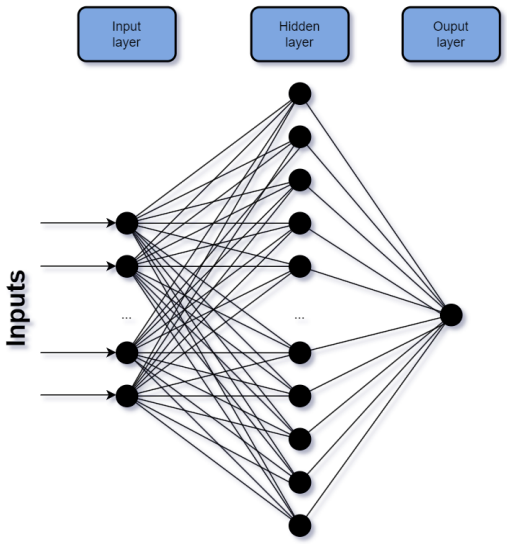


Figure 3.8: An example of a three-layer MLP with a multidimensional input layer and a one-dimensional output layer. Dots denote neurons and lines denote edges which have an associated weight that can be learned.

A DL model is defined by the layers it has, of which there consist many different types, too many to be listed here. As a result, this section only glosses over the layers of most interest for this master thesis. A focus is put on the intuitive explanation of the concepts with resources provided if more technical detail is desired. The DL book by Goodfellow et al. (2016) and the more general book on ML by Burkov (2019) also provide more insight on these layers. Géron (2019) provides a more hands-on book where the implementation of DL models using Keras is discussed. For the understanding of the LSTM architecture, the reader is referred to the work by Staudemeyer and Morris (2019). Due to time constraints, it was not feasible to provide a detailed discussion on this layer, as it requires the explanation of some general RNN concepts as well, which fall outside the scope of this master thesis. It is important to know that these layers are created with time series in mind and as such are good at working with time series, which EEG is. Their direct focus on this kind of data is why these layers are used for both the extension to EEGNet proposed by this master thesis.

Multilayer perceptron (MLP)

a multilayer perceptron (MLP) is one of the simplest types of a feed-forward neural networks (FFNN) and it is often reffered to as a vanilla NN (Burkov, 2019). An MLP consists of an input layer, one or more hidden layers and an output layer, all of which can be multidimensional of differing sizes. Al neurons are fully connected to the neurons in the consecutive layer in a one-directional manner, explaining why they are a type of FFNN. Depending on the activation function used in the output layer, MLP can be a function approximator for a classification issue or a regression issue. For a MLP to learn non-linear mappings from input to output, a non-linear activation function has to be used in at least one of its layers.

The architectural design of a MLP is conceptually simple. Despite its simplicity, being a universal function approximator and providing non-linearity through a correct activation function

choice, a MLP can be very capable. However, given each neuron is connected to all neurons of the following layer, it is a full-connected NN and as such can quickly blow up in the number of trainable parameters when sufficient neurons and layers are used. Thus, considering the number of edges and neurons present, there are more interesting approaches available. Whilst none of the models considered in this master thesis made explicit use of a MLP, they did make use of a dense layer, which is a fully connected feed-forward (FF) layer and as such is similar to a singular layer of the MLP.

Activation functions

As discussed, the activation function is what provides the MLP with non-linearity. Intuitively, this is easy to understand as all the neuron has is a sum of weighted inputs and a bias. None of these operations is non-linear and as such, no non-linear functionality can be learned as is. When providing the neuron with the activation function, it uses this activation function and passes its sum of weighted inputs to it to obtain the output. If this activation function is non-linear then the network can learn non-linear mappings. As such, using non-linear activation functions has almost become a requirement for obtaining capable NN.

In theory, any arbitrary function can be used as an activation function and many different activation functions exist. Choosing the right one can influence the training behaviour of the model a lot. Whilst not required in theory, most of the time each neuron of a layer uses the same activation function. This is also denoted in the notation of the MLP in Equation 3.15.

As described by Goodfellow et al. (2016), the default go to activation function currently is rectified linear unit (ReLU), the definition for which is given in Equation 3.16. ReLU was first described by Fukushima (1975) Whilst this equation looks linear, it is not given the change from a flat line on 0 for $x <= 0$ and a linear function for $x > 0$. ReLU is a simple activation function as it either passes zero or its input value. As such, using ReLU is computationally efficient, which is one of the reasons it is so popular. ReLU can also be used as classification function as described by Agarap (2018).

$$ReLU = f(x) = max(0, x) \tag{3.16}$$

However, by far the most famous classification function, especially for multiclass classification is softmax, which is used by all DL networks considered in this master thesis and was first introduced by Bridle (1989). Softmax provides a value between 0 and 1 for each of the output nodes, the sum of which totals 1. Strictly taken it is not a probability but rather a certainty metric of the network that happens to total 1. The network would have to be calibrated on another dataset such that this output indeed represents probabilities (Gupta et al., 2020). The softmax function is given in Equation 3.17

$$\text{softmax} = f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{3.17}$$

Another commonly used activation function is exponential linear unit (ELU) and is used by most of the DL models considered for this master thesis. It was first proposed by Clevert et al. (2015) who argued that despite being slower to calculate than ReLU, it causes faster convergence overall. It also eliminates the "dead ReLU" problem, where a ReLU activated neuron gets stuck in a position where it always outputs zero and thus is "disabled". The ELU function is given in Equation 3.18. The α parameter is a tunable value between 0 and 1. It was left at its default of 1 in this master thesis.

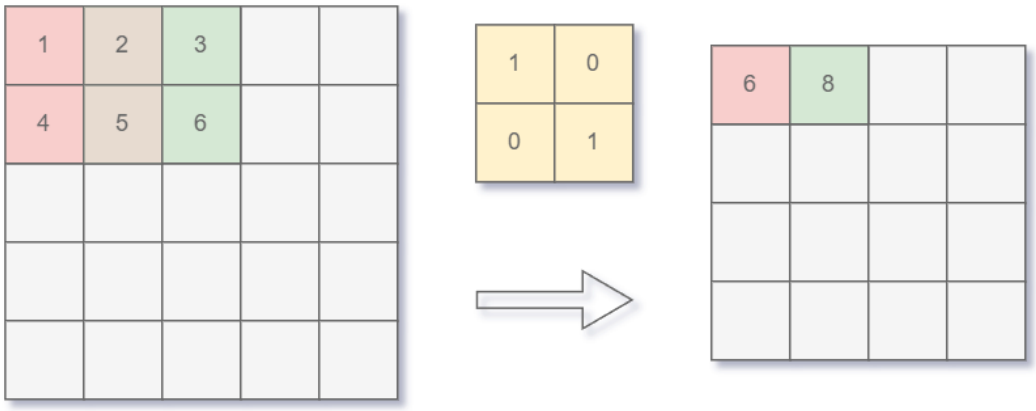


Figure 3.9: Visualisation of a convolutional layer.

$$ELU = f(x) = \begin{cases} x & if x > 0 \\ \alpha(e^x - 1) & otherwise \end{cases} \tag{3.18}$$

Many other activation functions exist but are not used in this master thesis. ShallowConvNet, as proposed by Schirrmeister et al. (2017), makes use of the self-explanatory square and log activation function. Activation functions are still an active area of research where gains are still possible, as discussed by D. Lee (2020). These gains revolve mainly around having more stable and faster convergence rather than obtaining better accuracy results.

Convolutional neural networks

CNN are a type of NN consisting of multiple consecutive convolutional layers often having intermediate pooling layers. A convolution using a 2x2 kernel and stride 1 working over a 2D input is shown in figure 3.9. The idea of a convolutional layer is as follows. For the specified amount of output channels n, n different kernels of a specified kernel size are made. Each of these kernels has trainable parameters which represent the values to use for the convolution. The kernel passes over the input in a left to right, top to bottom order, stepping by a step size that can be configured as the stride. On each step, the convolution is calculated and a singular value is placed on the appropriate location of the output. Taking the convolution of a kernel over the input consists of multiplying the overlapped values and taking the sum. When no padding is used, as is the case for the example in Figure 3.9, downsizing occurs due to the kernel reaching the border of the input space. This reduction corresponds to the size of the kernel in the corresponding orientation minus 1. Thus for Figure 3.9, both horizontal and vertical dimensions are reduced by 1. Settings exist to apply padding, which adds zeros or some other values around the image such that the output dimensions are equal to the input.

It is important to stress that the same kernel weights are used for processing the whole input. This makes convolutional approaches an interesting approach, as the amount of trainable parameters is limited and it has been shown that convolution layers are excellent at extracting spatially correlated features (Albawi et al., 2017). When considering the 2D matrix shown in Figure 3.9 as an EEG signal, using a kernel of height 1, a 1D convolution is done over the

EEG signals, without modifying the channels dimension. This is done by many of the models considered in this master thesis. The convolutional approach is further described in detail by Albawi et al. (2017). When using convolutional networks on a timeseries, it can detect local patterns and as discussed in Section 4.3.1 it can even learn frequency filters.

Besides a convolutional layer, a pooling layer is often used right after a convolutional layer. Pooling layers are made to reduce the dimensions of their input. As such, a pooling layer could be seen as a convolutional layer but where the kernel used a predefined fixed function that doesn't learn or adapt itself. Famous examples of this are max-pooling and mean-pooling.

The proposed models in this master thesis make use of a variant of the conventional layer called depthwise separable convolution and depthwise convolution. Depthwise separable convolution consists of two steps. First, a depthwise spatial convolution is performed, this acts on each input channel separately. Afterwards, a pointwise convolution is performed, this merges the resulting output channels. These variants are mostly used due to the speed increase they offer, as their outputs can also be achieved by regular multiple regular convolutional steps. These extension of CNN are further discussed by Lu et al. (2022) who also discuss further computational optimisations.

Dropout and batch normalization

Dropout is a layer most commonly used to combat overfitting. A dropout layer disables random neurons of its input by setting them to zero. This is done at a given probability. By doing so, you are forcing neurons to learn from many different neurons, as relying on one neuron for its output will degrade its performance when this neuron is turned off due to the dropout functionality. Intuitively, this incentivizes neurons to consider the input of more neurons and as such has proven to significantly help in reducing overfitting. A more detailed explanation as to why this is, is provided by Srivastava et al. (2014).

Batch normalization is a normalization layer. It collects a batch of samples, determines the mean and standard deviation and provides the normalized samples as output by having them subtracted by the mean and then divided by the standard deviation. In doing so the values have a zero mean and unit variance. This helps combat the internal covariate shift problem which is further discussed by Ioffe and Szegedy (2015).

3.4 Evaluating the performance of the pipeline

Section 1.4.4 discussed how a lack of standardized methods for both evaluation and reporting in biosignal control systems and BCIs in particular is one of many open issues. As discussed, the main problem resides in evaluation strategies made for testing specific parts of the BCI system, such as the classification pipeline, not representing the final performance of the complete system. However, due to the complexity of proposing a complete BCI system and performing an in-depth user study of that complete system, it is not feasible for every paper on BCI related research, including this master thesis, to do such extensive testing on a complete system.

For this reason, it was proposed in section 1.4.4 that the proposal of a complete BCI system should happen over at least four distinct papers, each focusing on a specific part. As discussed, the contents of this master thesis relate most to the second of those four, where classification pipelines are compared and proposed. As such, most of the evaluation metrics revolve around general ML metrics derived from the confusion matrix (CM) on either validation and/or test set. However, as further discussed by D'Amour et al. (2020), it should be noted that these metrics don't include certain deployment factors and are most often an overestimation of real-world performance. Besides these general ML metrics, some metrics mostly used in testing the performance of medical

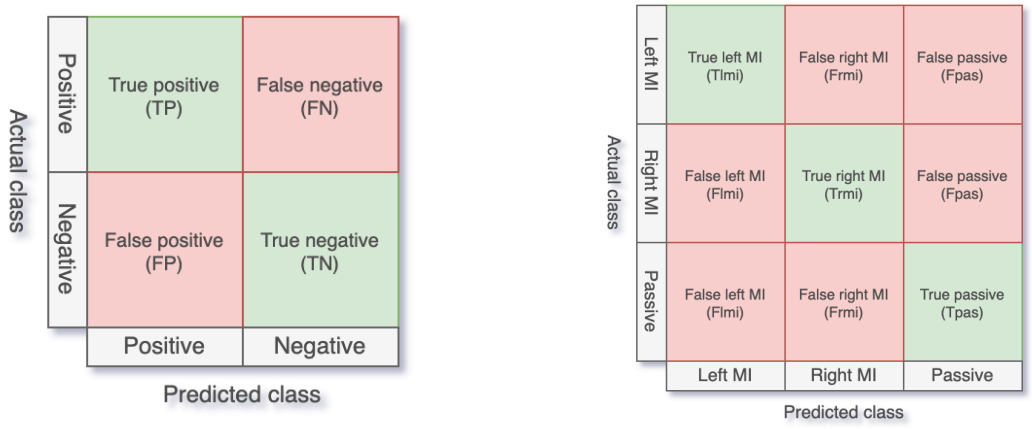
ML can be insightful for evaluating the risks associated with misclassification by the pipeline. The remainder of this section discusses these topics in more detail.

The train, validation and test set split

To test the performance of a classification pipeline, it is important that testing happens on data that was not used for training. Intuitively, this has to be done as asking the classifier to classify a sample it has already seen the label for during training, is not going to give any indication of the real-world performance of the algorithm. A simple 1-nearest neighbour algorithm would score perfectly in such a case and more complex algorithms would be incentivised to overfit to the data rather than learning the problem.

Most commonly, a complete dataset is split into three distinct sets: a training set, a validation set and a test set. The difference between this validation and test set is their intended use. For deciding the right parameters of a model, having some data that was not used for training, allows for testing which parameters are performing better. The validation set is made for these kinds of assessments. However, extensive use of the validation set will degrade the effectiveness of this set. Given enough variations of a model, one is bound to perform well on this validation, even if it doesn't learn anything or it is a completely random one. As such, another set, the test set, is needed for one *final* performance assessment of the model. It is important to stress that this test set should only be used once for this final assessment as further changes after obtaining an initial test score to improve this test score results in the test set becoming a validation set. In medical research, a 7:1:2 ratio for training, validation and test split is relatively common (Attia et al., 2019). However, this is heavily dependent on the amount of data available and the problem at hand.

One issue with splitting the data into a train, validation and test set is that the obtained split is not representative of the data it is taken from or that it suffers from *data leakage*. Depending on the dataset, it can be a complex task to combat this. First, samples should be sampled randomly such that each sample has an equal chance of being selected. Second, the sampling should happen in a stratified manner such that the distribution of classes or other attributes present in the data is respected. This will reduce the risk that a certain set contains an unreasonably high or low amount of samples from a certain class or attribute. Take for example a test set for the MI EEG data where due to unfortunate random sampling the test set consists of only healthy subjects (attribute) and mostly samples from a certain MI task (class) whilst the original dataset was balanced across all classes and had stroke patients included as well. A stratified approach based on the class label and health status of the subject eliminates such a scenario from happening. Third, no data leakage should happen, which requires taking into consideration certain groupings present in the data. Data leakage occurs when the training data contains information that directly helps in classifying samples from the validation or test split that is not present for a true unseen sample. For example, in a 3D medical lung imaging setting for classifying cancerous patients, having images or slices from the same patient in both the training and validation/test set is a common example of data leakage. Even though the training set does not contain that specific slice or image, it contains closely related data that would not be present for a true unseen sample in real-world applications. Grouped splits can solve this issue by ensuring all data of a subject is in the same group. However, when data is limited, ensuring all three of these things can be challenging or even impossible. Take for example data where all samples of a certain class are also from the same group, making it impossible to satisfy both the grouped or stratified conditions discussed.



(a) Confusion matrix for a universal binary classification problem.

(b) Confusion matrix for a three-class MI classification problem.

Figure 3.10: Design of a binary and three-class confusion matrix.

K-fold cross validation

Rather than taking a fixed train and validation set, using k-fold cross-validation (CV) can provide better results. The intuitive idea behind K-fold CV is simple. Rather than making a train/validation/test split, a train/test split is made using the same reasoning. This training data is then further divided into K folds, with five to ten being relatively common depending on the number of training samples available. The classification pipeline is then trained K times on K-1 of these folds. The fold that is not used for training, is used as a validation set and this set differs in each of the K training steps. This results in K different scorings for the validation set and thus allows for determining variance as well. It should be noted that the folds created have to use the same strategy as a fixed validation split to ensure each split is representative of the data without data leakage. The algorithm behind K-fold CV is given as pseudocode in Algorithm 1. One major drawback of using K-fold CV is that it requires training the classification pipeline K times. Considering some pipelines can take hours, days or even weeks to train, K-fold CV is not always feasible.

Algorithm 1 K-fold cross-validation

```
Divide training data in K folds          ▷ Respecting a certain splitting strategy
val_scores ← []                          ▷ Initialize list of validation scores
for fold k_i in K folds do
    Train model on K \ k_i
    val_scores ← val_scores+ metrics for validation set k_i
end for
process val_scores as desired
```

The confusion matrix and its derived scores

The CM is a simple but powerful tool for visualising the predictions made by a classification pipeline. The structure of a CM for both a universal binary classification problem and a three-class MI classification problem is shown in Figure 3.10. Many of the metrics used for evaluating the performance of a classification pipeline are derived from this CM. Take for example the most commonly used classification accuracy metric, it can be derived by equation 3.19.

$$\begin{aligned} \text{Binary accuracy} &= \frac{TP + TN}{TP + FN + FP + TN} \\ \text{Three-class accuracy} &= \frac{\sum\limits_{x \in \text{classes}} T_x}{\sum\limits_{x \in \text{classes}} F_x} \end{aligned} \tag{3.19}$$

However, reporting on accuracy alone is often not enough as it is also interesting to know what the per-class performance is. For this, sensitivity (also known as recall) and specificity are also commonly reported. Their binary formulea are given in Equation 3.20 and 3.21 respectively. Sensitivity gives information on the fraction of true samples being predicted as true whilst specificity provides the fraction of false samples being predicted as false. For a CM where more classes are present, the specificity and sensitivity per class can be calculated similarly.

$$\text{Binary sensitivity} = \frac{TP}{TP + FN} \tag{3.20}$$

$$\text{Binary specificity} = \frac{TN}{TN + FP} \tag{3.21}$$

Whilst specificity and sensitivity already provide more information, these metrics don't report nicely on unbalanced datasets. Assume a binary test setting where only 5% of the samples are of the true class, something that is relatively common for medical conditions such as AF. Papers studying the classification of this phenomenon, such as the one by Attia et al. (2019), will often boast a good specificity and sensitivity, i.e. Attia et al. (2019) reported both to be above 80%. However, due to the class imbalance, these metrics are not sufficient on their own or in combination with the classification accuracy. A mention of the positive predictive values (PPV) and negative predictive values (NPV) is also needed. The binary formulea for which are given in Equation 3.22 and 3.23 respectively. The PPV specifies how many of the samples labelled as true are actually true, the NPV does the same but for negatives. Whilst not reported in the previously mentioned paper by Attia et al. (2019), the PPV was found to be under 30% (Isaksen et al., 2022). In contrast with the high sensitivity and specificity, this would mean that for less than one in three times the model predicted true, the sample was indeed true. Relating this back to a BCI system, if the sensitivity and specificity of a risky action are good but the PPV is not reported and turns out to be poor due to an imbalance in the testing strategy, the model can not be used in a real-world application as so many unwanted executions of risky actions are not acceptable. The PPV is also referred to as the precision of a model. The PPV and NPV can also be determined for each class independently in case of a multi-class classification problem.

$$\text{PPV} = \frac{TP}{TP + FP} \tag{3.22}$$

$$\text{NPV} = \frac{TN}{TN + FN} \tag{3.23}$$

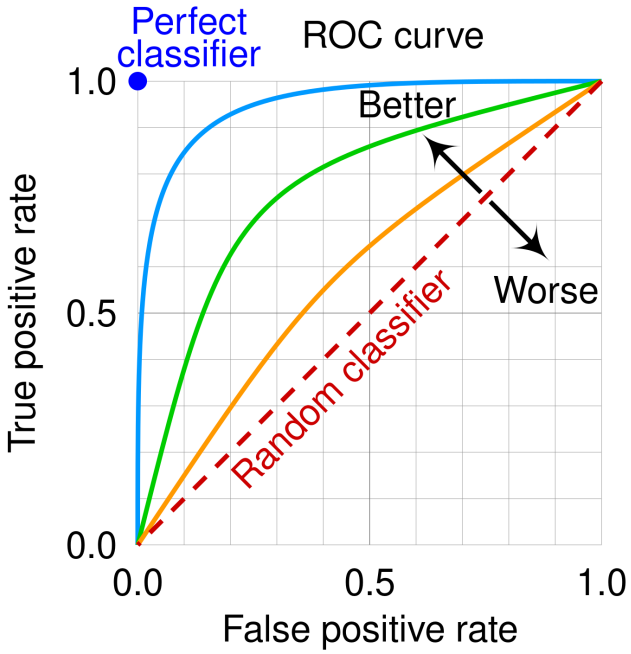


Figure 3.11: ROC curve with some arbitrary examples. Free to use figure by cmglee, MartinThoma, CC BY-SA 4.0, via Wikimedia Commons.

A final metric that is often reported is the F1 score. The F1 score is also derived from the idea that sensitivity alone isn't enough for imbalanced data. To solve this, it provides the harmonic mean of the PPV and the sensitivity. It's binary formula is given in Equation 3.24.

$$F1 = 2 * \frac{PPV * \text{sensitivity}}{PPV + \text{sensitivity}} \tag{3.24}$$

The ROC curve and choosing classification thresholds

Besides the CM, the ROC curve is also a powerful tool for visualising the performance of a classification pipeline. The idea behind the ROC curve curve is to visualize the model's performance when making a trade-off between sensitivity (also known as the true positive rate) and the false positive rate (1 - sensitivity). This trade-off can be controlled by changing the threshold at which a prediction made by the model is considered a prediction of the true label. By default, this is when the model returns a prediction of 0.5 or higher for the true class, but changing this value can yield more preferred behaviour. For example, lowering the threshold will make the model classify samples as positive more often, resulting in an equal or higher sensitivity (desired) but also an equal or higher true negative rate (not desired). Lowering the threshold value does the reverse. Finding the best balance between both is application specific. An example of this graph with some examples to demonstrate its working is shown in Figure 3.11. The ROC curve curve is often used to determine the optimal threshold value. The area under the (ROC) curve (AUC) is also derived from the ROC curve curve as a common evaluation metric.

3.5 Chapter conclusions

This chapter revolved around the discussion of a general pipeline for classifying brain signals. By discussing five different components and providing some examples of the techniques available within each component, all required insights for understanding the seven proposed classification pipelines in the next chapter should be provided. A discussion on which evaluation metrics are recommended to report was also given. This included a discussion on both general ML metrics and metrics more tailored towards the biomedical field where risk factors are more important than overall accuracy results. It was reasoned that this is also the case for BCI systems, especially those that can execute risky actions. Unfortunately, due to time constraints, it was not feasible to report on the LSTM layers, although the references provided explain this layer in great detail.

Part II

Development of motor imagery EEG classifiers

Chapter 4

EEG-based offline classification system for motor imagery tasks

4.1 Introduction

This Chapter discusses the seven different EEG based MI classification pipelines that were considered in this master thesis. In particular, two traditional two-step ML approaches were considered, both based on the CSP feature extraction technique. These two approaches differ by the frequency filter they use in their preprocessing step, where one is fixed throughout all experiments and the other is configured using hyperparameter tuning. The proven to be better performing extension to CSP, FBCSP, is also discussed but not considered for the experiments in this master thesis.

The other five classification pipelines that were considered in this master thesis are one-step DL approaches, as this master thesis focuses primarily on these kinds of approaches. First, three literature proposed state-of-the-art CNN-based approaches are discussed: EEGNet, DeepConvNet and ShallowConvNet. Both their architectural design and some implementation details are discussed. The visualisation possibilities of these CNN-based EEG classification algorithms are also addressed.

The final two DL classification pipelines that were considered in this master thesis are an extension to EEGNet proposed by this master thesis. Both of these extensions aim to provide additional memory to EEGNet by incorporating LSTM functionality. One reuses all components of the EEGNet architectural design but adds a tunable LSTM layer right before the softmax layer. The other approach only reuses part of the original EEGNet architectural design and makes use of a convolutional LSTM layer. Again, some of the implementation details are also addressed.

All of the implemented pipelines together with the performed experiments and saved results are available on the GitHub page of this master thesis (Bontinck, 2021). The state-of-the-art literature proposed CNN-based DL models are a modified version of the Keras reimplementation provided by Schirrmeister et al. (2022) and are made available in the utillity file: `EEGModels.py`. The EEGnet extensions are provided in the utility file: `EEGNet_with_lstm.py`. A custom filter that is hyperparameter tunable with sklearn has also been made and is available in the utility file `custom_sklearn_components.py`. More general Keras and TensorFlow tools are also made available in a separate utility file: `TF_tools.py`. All these utility files can easily be imported and have the required inline documentation to make reusing them easy.

4.2 CSP-based two-step ML approaches

As discussed, the focus of this master thesis lies on one-step DL MI EEG classification pipelines, as these are most widely used in biosignal control systems according to Dillen et al. (2022). However, considering the significant role CSP played in allowing ML approaches to effictively learn from EEG data, accelerating the field of BCI research, this master thesis also considers two traditional two-step MI EEG ML classification approaches based on CSP. The architectural design of both of these pipelines is discussed in what follows.

It is noted that since the introduction of CSP by Koles et al. (1990), many extensions have been proposed that have proven to be far superior to this original version (Kai Keng Ang et al., 2008; Lemm et al., 2005). As such, the results obtained from these two regular CSP pipelines discussed in Chapter 6 should not be taken as an indicator of the performance for traditional two-step MI EEG ML classification approaches in general.

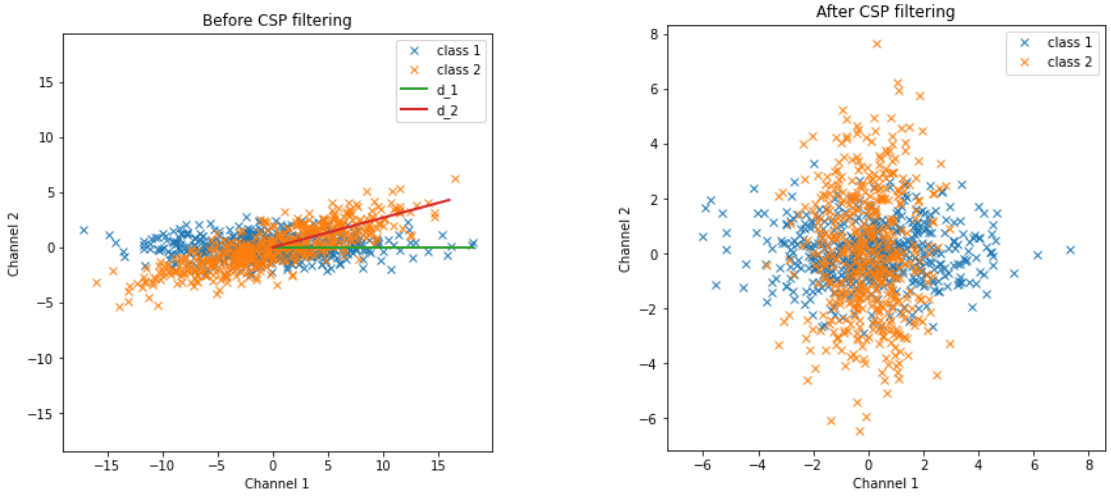
4.2.1 The idea behind common spatial pattern(s) (CSP)

Just as SVM was a classification method originally proposed to solve binary classification tasks (see Section 3.3.3), CSP was originally proposed to find a spatial filter for feature extraction that leads to optimal variances between two classes of EEG data. The general idea of a spatial filter for feature extraction was already discussed in Section 3.2.4. This is a non-trivial task due to ERS and ERD signals, such as MI, being time-locked to the event but not phase-locked and them being oscillatory processes. To achieve an effective optimal variance between two classes of EEG data, CSP has to make some assumptions. CSP assumes the frequency band of interest is known, as discussed in Section 2.4.3, this was around 7Hz to 30Hz in case of MI tasks. Another assumption that is made is that the time window is known, which is more difficult to ensure as windowing the data around an event in the prediction timeline is not always possible, as further discussed in Section 3.2.2. Luckily, CSP can still work reasonably well even if both frequency band and time window estimations are off. Besides these assumptions, CSP also assumes that the band-passed signal is jointly Gaussian within the time window and that there is a difference in the oscillatory signals between both classes. Both of these assumption hold for most MI EEG data.

As discussed in Section 3.2.4, the goal of a spatial filter is determining an optimal weight matrix W for Equation 3.4. In that section, it was also discussed how some data-driven solutions such as PCA and ICA do not use the class information, which can have degrading effects. The goal of CSP is to find a weight matrix W so that the variance of one signal is minimal, whilst the variance of the other is maximal. This is shown in Figure 4.1 where before filtering (Figure 4.1a) the signals are highly overlapping and though to discriminate, but after using CSP the signals are far easier to discriminate.

There are multiple approaches to calculating the optimal weight matrix W based on the optimal variance criteria. Most commonly are using an optimisation problem, describing it as a generalized Eigenvalue problem and a more geometric approach. The remainder of this section will discuss the generalized Eigenvalue problem approach. B. Wang et al. (2021a) discuss an optimisation approach in more detail, Barachant et al. (2010) does the same for a geometric approach.

The optimal weight matrix W for Equation 3.4 based on the CSP criteria can be mathematically described as shown in Equation 4.1. In this equation, X_i denotes the windows of class i represented as a 2D matrix of n channels by t_i EEG measurements per channel. Due to the jointly Gaussian assumption, the covariance matrix of the signals per class can be easily calculated using Equation 4.2 where i denotes the class of interest. From this, the generalized



(a) An arbitrary two-channel signal for two classes before CSP feature extraction. Free figure by MalcolmSlaney, CC BY-SA 4.0, via Wikimedia Commons.

(b) An arbitrary two-channel signal for two classes after CSP feature extraction. Free figure by MalcolmSlaney, CC BY-SA 4.0, via Wikimedia Commons.

Figure 4.1: Visualisation of the spatial transformation performed by CSP.

Eigenvalue problem can be created to solve the problem. Namely, the goal is to find the Eigenvectors $\mathbf{P}$ which contain the Eigenvector of any channel j ($\mathbf{p_j}$) such that Equation 4.3 holds. In this Equation, $\mathbf{I}_n$ denotes the identiy matrix for i channels. This problem can be further reduced to the Eigenvalue decomposition shown in Equation 4.4. From this, $\mathbf{w}$ can finally be found by taking $\mathbf{w} = \mathbf{p}_1^T$. B. Wang et al. (2021b) provides a more detailed explanation of the generalized Eigenvalue problem approach for solving CSP and some additional steps that help make it more applicable to BCI settings.

$$\mathbf{w} = \arg\max_{\mathbf{w}} \frac{||\mathbf{w}X_1||^2}{||\mathbf{w}X_2||^2} \tag{4.1}$$

$$\mathbf{R}_i = \frac{\mathbf{X}_i\mathbf{X}_i^T}{t_i} \tag{4.2}$$

$$\begin{aligned} &\mathbf{P}^T\mathbf{R}_1\mathbf{P} = \mathbf{D} \\ &\text{and} \\ &\mathbf{P}^T\mathbf{R}_2\mathbf{P} = \mathbf{I}_n \end{aligned} \tag{4.3}$$

$$\mathbf{R}_2^{-1}\mathbf{R}_1 = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} \tag{4.4}$$

4.2.2 Traditional CSP with LDA, SVM and RF

Having explained the idea behind CSP, the pipelines using CSP in the experiments of this master thesis can be discussed. The first pipeline uses the standard CSP feature extraction method for classifying MI EEG data in an offline setting and is visualised in Figure 4.2. As discussed further in Section 6.2 the used data considers a fixed 3-second window surrounding a known event,

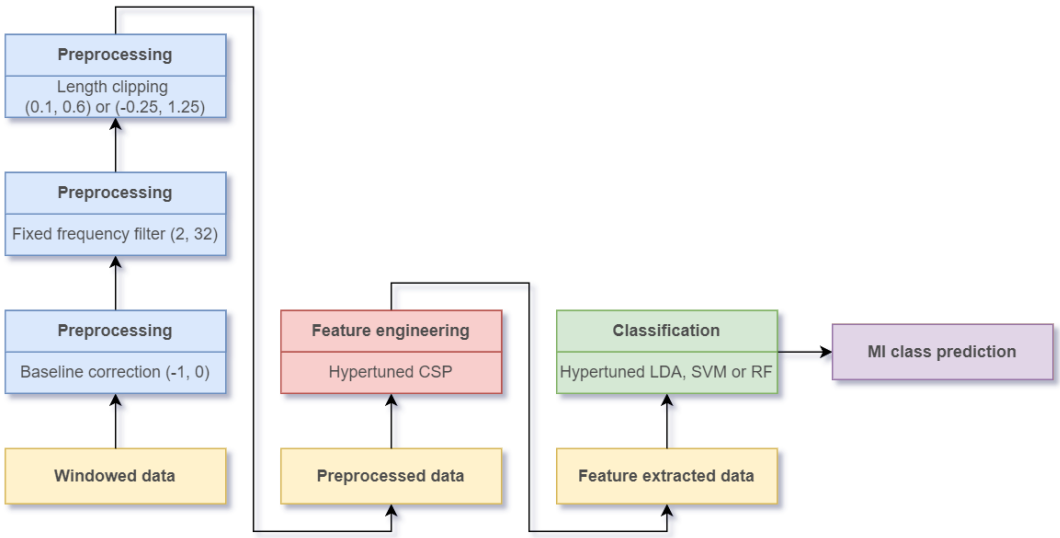


Figure 4.2: Visual overview of the pipeline used for MI EEG classification using the standard CSP method and multiple traditional two-step ML classifiers.

including 1 second before the event onset and two seconds after the event onset. As discussed in Section 3.2.2, different windowing stragies exist but are not epxlored in this master thesis, in part due to CSP assuming a known fixed time window as discussed in Section 4.2.1.

The windowed data is first preprocessed with multiple preprocessing techniques that were already discussed in Section 3.2.3. First, the window is baseline correct on one second of data before the event onset. The resulting data is further preprocessed to only include frequencies between 2Hz and 32Hz. This is done by the use of a FIR filter design using the Blackman window method. Whilst Section 2.4.3 discussed MI is mostly present in the frequency range of 7Hz to 30Hz, the boundaries were loosened to account for a fixed approach that is not hyper-tuned on a subject basis and the inclusion of a *neutral* class, which is not a specific MI task and most likely has lower frequency brain waves when taking into account the brainwave subdivision given in Table 2.4. Finally, the resulting data is clipped to a 2D matrix (channels x EEG measurements) including the samples from either 0.1 seconds after the event onset to 0.6 seconds after the event onset or from 0.25 seconds before the event onset to 1.25 seconds after the event onset, depending on the experiment setup. It should be noted there is no explicit band-stop filter in place to cancel out the AC artefacts discussed in Section 2.4.6 as the used dataset has already filtered out these frequencies (Kaya et al., 2018).

For feature engineering, the multiclass CSP method described by Grosse-Wentrup and Buss (2008) is used. The use of a variant is needed as the CSP method discussed in Section 4.2.1 only support binary classification where the experiments from this master thesis use three-class MI EEG data. The number of CSP components is hyperparameter tuned based on the possible selection of 2, 3, 4, 6 or 10 components. This number may seem limited but providing more components will make overfitting more likely and in practice, the differences between six or more CSP components were found to be minimal after some pilot studies done in the experimental

notebooks available on the GitHub repository of this master thesis (Bontinck, 2021).

For classification, three different classifiers were considered, LDA, SVM and RF, all three of which were discussed in detail in Section 3.3.3. As discussed, part of the reason the LDA classifier is so attractive is due to it not requiring specific hyperparameter tuning. As such, the LDA classifier was not hyperparameter tuned in all but a few experiments that validated there was indeed negligible difference when changing the solver and tolerance values from their default parameters. For the SVM classifier, the earlier discussed C value for controlling the smoothness of the boundary when using the kernel trick was hyperparameter tuned by trying multiple values between 0.01 and 100. The kernel was also hyperparameter tuned by trying the RBF, sigmoid and linear kernels. For the RBF and sigmoid kernel, the gamma parameter denoting the kernel coefficient was also hyper-tuned on values between 0.001 and 10, including some calculated values provided by sklearn. Finally, the RF classifier was hyperparameter tuned for the number of decision trees (values between 10 and 500) in its ensemble as well as the maximum depth (3, 10 or no limit), minimal sample split (2, 5 or 10) and features (different percentual proportion including no limit) to be used by each of those decision trees.

The used preprocessing techniques and CSP methods are supplied by the Python MNE library (Gramfort, 2013). The traditional two-step ML classifiers are provided by the sklearn library from which the hyperparameter tuning functionalities were also used to hyperparameter tune the required components of this pipeline (Pedregosa et al., 2018).

4.2.3 The issue with a traditional CSP approach

It has been discussed that in the classification of MI EEG data using CSP feature extraction paired with almost any classifier, even simple one such as LDA, can produce pleasant results for some experimental settings. However, CSP in itself is a limited feature extraction method compared to the proposed extensions that have far outperformed it (Abdeltawab & Ahmad, 2020; Kai Keng Ang et al., 2008; Khan et al., 2020; Lemm et al., 2005). To understand the intuitive reasoning behind most CSP extension, a different pipeline is configured where the fixed frequency filtering in the preprocessing step of the pipeline shown in Figure 4.2 is now hyperparameter tunable. Practically, this is done by providing an equal filtertering strategy as the one described in Section 4.2.2 through a sklearn transformer. For this custom sklearn transformer provided in the `custom_sklearn_components.py` utility file of this GitHub project, the lower bound and upper bound can be configured (Bontinck, 2021). Being a sklearn transformer, it can be used in the same hyperparameter tuning strategy as described in Section 4.2.2 to find optimal lower bound values and upper bound values.

From the results of these experiments, which are further discussed in Chapter 6, it becomes clear that the performance of CSP is heavily dependent on the used frequency filtering. Accuracy for the worst found parameters and best parameters in 4-fold cross-validation ranged from 30% to 75% in an intrasession testing setup. This huge fluctuation in performance follows from the assumption of a know frequency band that CSP makes as discussed in Section 4.2.1. The best found frequencies were also subject dependent, as Section 2.4.3 already suggested. Whilst incorporating this hyperparameter tuning as a training procedure in the CSP approach would already increase the performance of CSP by automating the finding of the best frequency range for a subject, even better solutions have been proposed. These follow mainly from an equal intuitive idea of learning the frequency band(s) of interest. The next session will discuss the commonly used FBCSP extension of CSP.

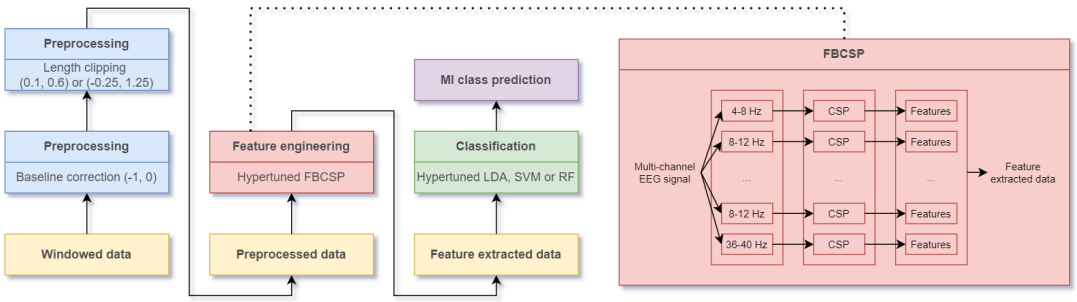


Figure 4.3: Visual overview of a proposed FBCSP pipeline used for MI EEG using multiple traditional two-step ML classifiers.

4.2.4 Improving traditional CSP with FBCSP

Since the frequency band has such a significant influence on the performance of CSP, as discussed in Section 4.2.3, many extensions to CSP have been introduced to automate the learning of the best frequency band(s). One popular extension to CSP that builds upon the idea of learning frequency-bands from labelled data is FBCSP as proposed by Kai Keng Ang et al. (2008). Figure 4.3 shows how a potential FBCSP pipeline may look. There are two main differences with the previous pipeline that was shown in Figure 4.2: the removal of the manual frequency filter and the replacement of CSP feature extraction by FBCSP feature extraction.

Intuitively, FBCSP performs multiple frequency band filters with different boundaries and stores the results on separate independent threads. For each of those threads, traditional CSP is then performed, resulting in many different CSP features. Since this data would likely be too complex to be handled by a traditional two-step ML classifier, a smaller subset of features is selected from the CSP filtered signal. The list of these final features, with multiple features per original thread, is then provided to the classifier as before. This FBCSP concept is also illustrated in Figure 4.3. It should become clear that many different approaches for creating potentially overlapping frequency bands are possible and that the final feature extraction from the CSP transformed signals can also differ greatly. These are all aspects that can be optimized based on the implementation of the proposed algorithm. It should also be noted that FBCSP does not explicitly learn frequency bands but rather provides multiple features from multiple frequency bands for which it relies on the classifier to ignore those that have no or little information.

Whilst being a relatively simple extension to CSP, FBCSP combined with a classifier such as SVM or RF has been proven successful at classifying MI EEG with far greater accuracy then the traditional CSP pipeline used in this master thesis (Kai Keng Ang et al., 2008; Khan et al., 2020; X. Wei et al., 2021). The performance of FBCSP even rivals state-of-the-art one-step DL classifiers such as DeepConvNet and ShallowConvNet (Schirrmeister et al., 2017). However, the implementation and evaluation of such a pipeline fall outside the scope of this master thesis. It should also be noted that whilst FBCSP is often used in literature it is not included in any of the popular Python EEG processing libraries. Providing this feature extraction method as an sklearn compatible component can be a helpful future work to facilitate pipeline development of future research. It is noted that some open-source Python implementations of FBCSP are available but they have limited support and documentation (Chouhan, 2020; Zancanaro, 2021).

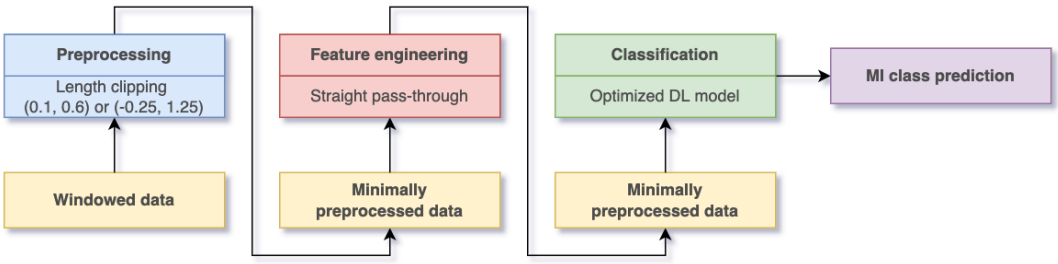


Figure 4.4: Visual overview of a pipeline using a DL classifier for MI EEG classification.

4.3 Convolutional-based one-step DL approaches

The majority of the models considered in this master thesis use a DL classifier. Their pipelines are simplified from those used in traditional two-step ML classification since they require minimal preprocessing and no feature extraction as was discussed in Section 3.3.2. Figure 4.4 shows the general pipeline structure used for the DL MI EEG classifiers that will be discussed in greater detail below. All of these classifiers are made using Keras and are available via the appropriate utility files on the GitHub repository of this master thesis (Bontinck, 2021; Chollet et al., 2015).

It should be noted once again that the used open-source dataset has already performed a band-stop filter to cancel out the AC artefacts discussed in Section 2.4.6. However, it could be argued that even this filtering of the signal to remove the AC artefacts can be learned by the DL model. Although due to this frequency being overpopulated by the AC artefact there is likely very slim to no information present in the filtered-out frequency range. As such, removing the AC artefact and other artefacts manually can facilitate the learning process, decreasing the computational time it takes to train the model. For these reasons, minimal preprocessing, such as AC artefacts removal is often still done in DL pipelines.

4.3.1 EEGNet

EEGNet is a CNN based EEG classification model proposed by Lawhern et al. (2016). There are some key properties of EEGNet that make it different from other DL models proposed in the BCI field. First, rather than using traditional convolutional layers, it uses depthwise and separable convolutional layers to reflect known performant feature extraction methods from the BCI field. These special types of convolutional layers were already discussed in Section 3.3.4. As a result of using these special types of convolutional filters and by being a compact CNN, EEGNet can learn from relatively few samples (Lawhern et al., 2016).

Second, instead of focusing on one specific BCI paradigm such as ERS and ERD or even more specifically MI, the model is created in such a way that it is applicable to many BCI paradigms. In the paper proposing the model, four distinct datasets from different paradigms were used to compare the performance with other models. The first dataset consists of visual P300 ERP signals. The second dataset revolves around another type of ERP signal which occurs after an unusual event occurs in the subject's environment or task: an error-related negativity (ERN). The third datasets contains both ERP signals and oscillatory components as a consequence of ERS or ERD. This type of potential is known as a movement-related cortical potential (MRCP).

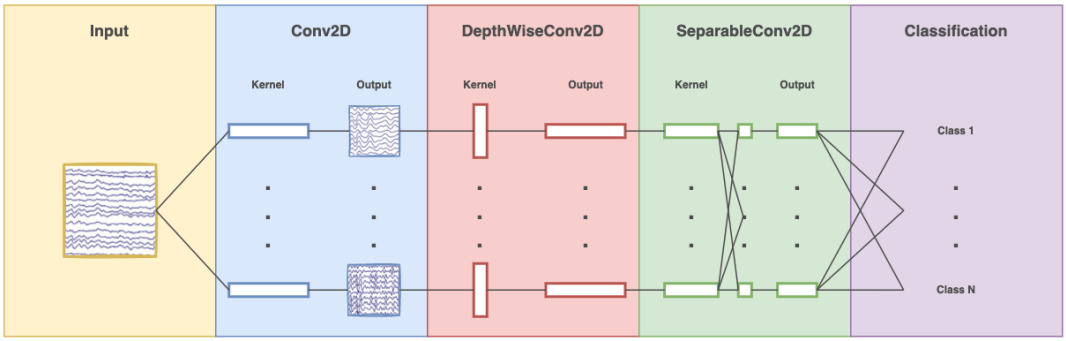


Figure 4.5: A conceptual overview of EEGNet after the overview by Lawhern et al. (2016).

The fourth and final dataset revolves MI, the most challenging paradigm according to Lawhern et al. (2016). Lawhern et al. (2016) have shown that EEGNet has comparable performance across all these tested paradigms compared to the state-of-the-art of that specific paradigm when few data samples are available.

Third, Lawhern et al. (2016) proposed a method of visualising the learned features of the model. Together with DeepConvNet and ShallowConvNet that is discussed later in this chapter, EEGNet is considered a state-of-the-art method in the BCI field.

The most important components of EEGNet are the different convolutional layers as shown in the conceptual overview of 4.5. The first convolutional layer is a regular two-dimensional convolutional layer and is used to learn frequency-like filters. The second convolutional layer is a depthwise convolution layer and is used to learn spatial filters specific to the frequency-converted signal given as output by the previous layer. The third and final convolutional layer is a separable convolution layer. As discussed in Section 3.3.4, this is a combination of a depthwise convolution followed by a pointwise convolution. Lawhern et al. (2016) state this depthwise convolution learns a temporal summary for each feature map individually and the pointwise convolution layer learns how to optimally mix the feature maps together.

When comparing this design with the FBCSP method visualised in Figure 4.3 it is clear that they share a lot of similarities. The first convolutional layer which tries to learn frequency-like filters relates to the first step of FBCSP performing different band-pass filters. The second convolutional layer which tries to learn a spatial filter for the frequency filtered signal is comparable to the use of CSP in the second phase of FBCSP. The third and final convolutional layer which aims to extract temporal features from the spatial filter and connect all layers back together is comparable to the feature extraction from the CSP threads in FBCSP and the traditional two-step ML classifier used in FBCSP.

An overview of all layers effectively used is given in Figure 4.6. This combines the use of dropout and batch normalization to combat overfitting tendencies and the internal covariate shift as discussed in Section 3.3.4 among other technicalities. Compared to the implementation of EEGNet provided by Schirrmeister et al. (2022) many more parameters of the model were made tunable. In particular, the following parameters are tunable and their defaults for the experiments are given in brackets: channel amount (21), amount of timepoints (100), amount of classes (3), the dropout rate (0.5), the kernel length (50), F1 (8), F2 (16), D (2), the normalisation rate (0.25) and the dropout type (2D spatial dropout). Most of these parameters are based on

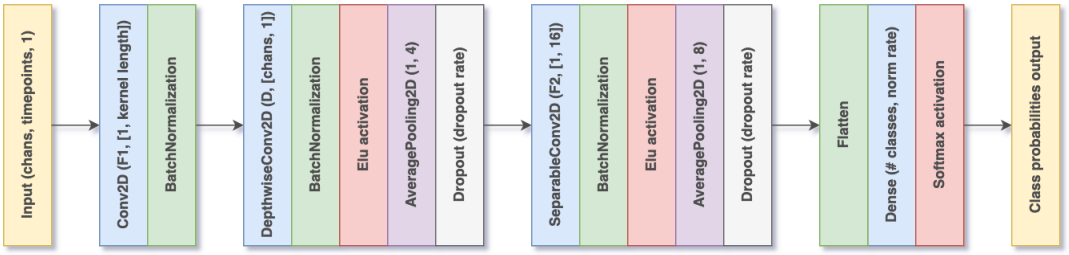


Figure 4.6: All layers of the EEGNet model.

the recommnedation provided by Lawhern et al. (2016) and Schirrmeister et al. (2022).

4.3.2 DeepConvNet

DeepConvNet is another CNN based EEG classification model. It is proposed by Schirrmeister et al. (2017) and was made for the more specific use of decoding MI and effective motory tasks. Although DeepConvNet took a relatively general DL approach, it is shown by by Lawhern et al. (2016) that it generalizes worse to BCI paradigms beside MI. In particular, Lawhern et al. (2016) found DeepConvNet to perform poorly on the MRCP paradigm dataset where EEGNet got almost double the AUC. Lawhern et al. (2016) found the performance of EEGNet and DeepConvNet to be comparable for the MI dataset. To further validate these results, DeepConvNet was also used in the experiments of this master thesis.

An additional motivation for using the DeepConvNet model proposed by Lawhern et al. (2016) is the fact that it also has a visualisation technique of the learned features just like EEGNet. Another motivating aspect of including DeepConvNet in the experiments of this master thesis follows from the authors also proposing a more shallow alternative, ShallowConvNet. ShallowConvNet is also used in the experiments of this master thesis and discussed in Section 4.3.3. Having both an explicitly deep and shallow network allows for comparing the two besides the theoretical discussion from Section 3.3.4. Finally, since DeepConvNet is a relatively general DL model that is inspired by successful CNN from the computer vision field such as ImageNet by Deng et al. (2009), it doesn't strictly represent feature extraction in terms of knowing good strategies. This is sharp contrast with EEGNet that took clear inspiration of FBCSP as discussed in Section 4.3.1.

An overview of all layers effectively used is given in Figure 4.7. One noteworthy aspect of this architecture is that there is no activation function between the second and third layers, which are both convolutional layers. The first convolutional layer goes over the timepoints while the second goes over the channels. Since these are two separate dimensions and there is no activation function in between, they could have been combined into a singular convolutional layer. Whilst the base implementation of DeepConvNet was provided by Schirrmeister et al. (2022), many more parameters of the model were made tunable once again. In particular, the following parameters are tunable and their defaults for the experiments are given in brackets: number of classes (3), number of channels (21), amount of timepoints (100), dropout rate (0.6), the kernel length of the first convolutional layer (4), striding of 2D convolution layers (1, 2), pool size (1, 2). Most of these parameters are lowered from their recommended settings by Schirrmeister et al. (2017) due to the low amount of training data present for the experiments of this master thesis.

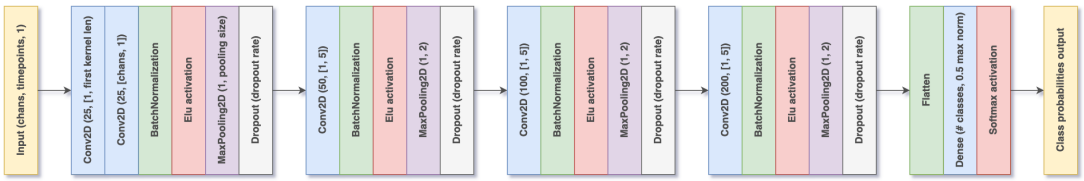


Figure 4.7: All layers of the DeepConvNet model.

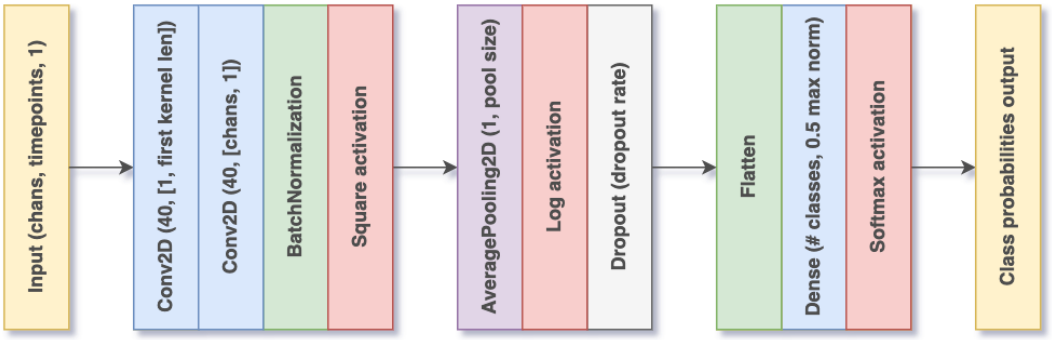


Figure 4.8: All layers of the ShallowConvNet model.

4.3.3 ShallowConvNet

As discussed in Section 4.3.2, ShallowConvNet is a more shallow CNN based EEG classificiation model proposed in the same paper of the DeepConvNet proposal (Schirrmeister et al., 2017). As opposed to DeepConvNet which took inspiration from complex CNN from the computer vision field, ShallowConvNet takes inspiration from FBCSP in their design. Schirrmeister et al. (2017) argues the first convolutional layer corresponds to the band pass filtering of FBCSP and the second to the CSP filter in FBCSP. Schirrmeister et al. (2017) argues the rest of the model relates to the FBCSP phase of finding features in the CSP streams, combining them and training a traditional ML classifier.

An overview of all layers effectively used is given in Figure 4.7. Again, there is no activation function between the second and third layers, which are both convolutional layers. As discussed for DeepConvNet, this means they could have been combined into a singular convolutional layer. Whilst the base implementation of ShallowConvNet was provided by Schirrmeister et al. (2022), some more parameters of the model were made tunable once again. In particular, the following parameters are tunable and their defaults for the experiments are given in brackets: number of classes (3), number of channels (21), amount of timepoints (100), dropout rate (0.5), the kernel length of the first convolutional layer (10), striding of 2D convolution layers (1, 6), pool size (1, 30). Most of these parameters correspond to the recommended settings by Schirrmeister et al. (2017).

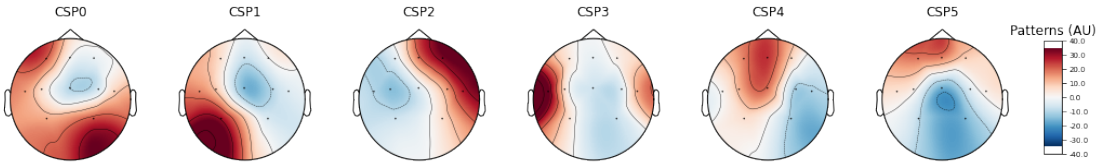


Figure 4.9: topomaps created from multiple CSP components using subsampled channels from the international 10-20 system.

4.3.4 Interpreting these black box models

It was discussed for both the EEGNet model as well as the DeepConvNet and ShallowConvNet model that the visualisation methods are described in the papers proposing the models (Lawhern et al., 2016; Schirrmeister et al., 2017). The intuitive reasoning behind these visualisations is that, as discussed, the first two convolutional layers of these models relate to the first two steps of the FBCSP feature extraction method. This means that the transformation offered by the first convolutional layer corresponds to frequency-like filters. As such, the trained model can be shortened to include only this first convolutional layer and when passing through the source EEG data, each output channel (filter) of the convolutional layer represents a frequency-like filtered EEG signal when plotted. Both through frequency analysis of these output channels and by analysis of the weights of the learned kernel, insight into what exact modifications have been made can be determined. Intuitively, this is how both papers visualise or determine the learned frequency band-pass filters.

Likewise, the second convolutional layer of these models corresponds to the temporal filtering done in FBCSP. Looking at the learned kernel weights, it can be seen how much each channel of the input EEG contributes to the output of the second convolutional layers. This allows for creating topographic maps (topomaps) similar to those that can be created from a CSP feature extraction map. An example of such topomaps created from multiple CSP components is given in Figure 4.9.

For more details on how these visualisation work for the discussed CNN models, the reader is referred to the original paper proposing them (Lawhern et al., 2016; Schirrmeister et al., 2017). Neither Lawhern et al. (2016, Authors of EEGNet) or Schirrmeister et al. (2017, Authors of DeepConvNet and ShallowConvNet) provide their source code to recreate these visualisations. Adding to this, the proposed visualisation method in the EEGNet paper makes use of the Deep-Explain Python library by (Ancona et al., 2020) for visualising Keras models. However, this library only has official support for TensorFlow V1, whilst the Keras and the TensorFlow version used for the implementations in this master thesis are V2. Because of this, providing further details on this visualisation through means of reimplementation of these methods falls outside the scope of this master thesis. More details on visualisation techniques for CNN in general is given by Qin et al. (2018).

4.4 Adding memory to one-step DL approaches

Section 3.3.4 discussed that convolutional layers used on time series data, or on the time axis of a 2D signal such as EEG, can learn kernels which detect local patterns in the time series. In that same section, it is also discussed that LSTM layers have explicit memory built into their design. This makes processing time series where temporal features are of most importance an ideal application for LSTM. For these reasons, studies across many fields have been done to

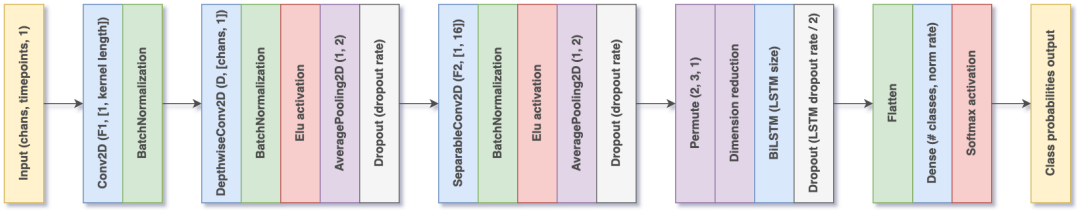


Figure 4.10: All layers of the CNN-LSTM model using all EEGNet layers for feature extraction and the addition of LSTM functionality for time series processing.

encorperate both convolutional layers for feature extraction and an LSTM layers for working with the time-series features into a singular CNN-LSTM model (Garcia-Moreno et al., 2020; Jeong et al., 2020).

Given that the EEGNet model works well on multiple BCI paradigms, it is assumed that the features extracted by EEGNet carry a lot of information. Following this reasoning, this master thesis proposes two CNN-LSTM models where the CNN based feature extraction is heavily inspired by the EEGNet model. Having the EEGNet architecture first, the EEGNet-specific visualisation could still be used. Likewise, the visualisation of LSTM layers to determine which areas in a temporal domain are of interest has also already been widely studied. Most commonly in the processing of ECG, this technique has seen great success, with Vijayarangan et al. (2020) using LSTM to visualise which regions of an ECG was used most for determining the classification of possible arrhythmia. As such, it is believed that a CNN-LSTM model can give great insight into both the spatial features and temporal features learned when using appropriate visualisation techniques. The study of which is an interesting future work for this master thesis.

4.4.1 Adding an additional LSTM layer to EEGNet

The first proposed CNN-LSTM model consists of the combination of all EEGNet layers discussed in Section 4.3.1 with a traditional bidirectional LSTM layer. This layer type was discussed in Section 3.3.4. An overview of all layers used is given in Figure 4.10. Note some subtle changes in the EEGNet feature extraction part of the model, namely a reduction in the downsampling performed by the pooling layers. This was done to retain a reasonable amount of temporal resolution. Besides this, some dimensionality transformations and permutations are in place to ensure a correct data representation for the LSTM layer. It is important to note that the LSTM layer also makes use of internal dropout functionality provided by Keras (Chollet et al., 2015). The dropout layer shown after the LSTM layer is an additional regular dropout layer using half of the internal dropout rate provided for the LSTM layer. Such an aggressive dropout strategy is needed due to the many additional trainable parameters introduced by the LSTM increasing the risk of overfitting.

The input of the LSTM layer using the default parameters given below corresponds to a 2D matrix of 25 timepoints, a reduction of four compared to the input signal due to the two pooling layers, having 16 features each. The following parameters are tunable for the EEGNet part of the model and their defaults for the experiments are given in brackets: channel amount (21), amount of timepoints (100), amount of classes (3), the dropout rate (0.5), the kernel length (50), F1 (8), F2 (16), D (2), the normalisation rate (0.30) and the dropout type (2D spatial dropout). For the added layers, the parameters and their defaults for the experiments given in brackets

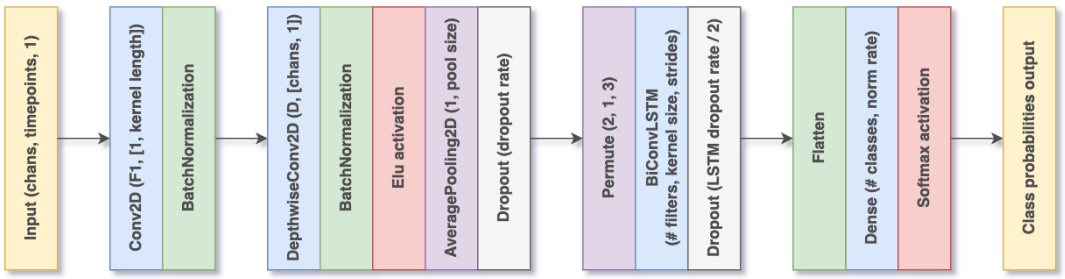


Figure 4.11: All layers of the CNN-LSTM model using a reduced version of EEGNet for feature extraction and the addition of convolutional LSTM functionality for time series processing.

are: LSTM units (64), LSTM specific dropout (0.6) and LSTM specific kernel legularizer (L1= L2 = 0.0001).

4.4.2 EEGNet with convolutional LSTM layer

The second CNN-LSTM model that this master thesis proposes consists of the combination of only the first two convolutional layers from EEGNet discussed in Section 4.3.1 with a bidirectional one dimensional convolutional LSTM layer. This layer type was discussed in Section 3.3.4. An overview of all layers used is given in Figure 4.11. In essence, the last convolutional layer that was present in EEGNet is replaced by a LSTM layer that uses a comparable convolution.

For the same reasons as before, a reduction in the downsampling performed by the pooling layer is performed, with the default setting used for this experiment not using this pooling layer at all. Again, the LSTM layer makes use of internal dropout functionality provided by Keras (Chollet et al., 2015) and the dropout layer shown after the LSTM layer is an additional regular dropout layer using half of the internal dropout rate provided for the LSTM layer. It is noted that the Keras implementation of the one-dimensional convolutional LSTM layer has no CUDA support and as such is incredibly slow to train.

The input of the LSTM layer using the default parameters given below corresponds to 100 timepoints having 16 features each. These 16 features are over 1 channel due to the depthwise convolutional layer present in the model. The following parameters are tunable for the EEGNet part of the model and their defaults for the experiments are given in brackets: channel amount (21), amount of timepoints (100), amount of classes (3), the dropout rate (0.5), the kernel length (50), F1 (8), D (2), the normalisation rate (0.25), average pooling before LSTM (None) and the dropout type (dropout). For the added layers, the parameters and their defaults for the experiments given in brackets are: LSTM filters (64), LSTM kernel size (4), LSTM specific dropout (0.6), strides used for convolution (1) and LSTM specific kernel legularizer (L1= L2 = 0.0001).

4.5 Chapter conclusions

This chapter discussed the seven different MI EEG classification pipelines used for the experiments in this master thesis. Two traditional two-step ML approaches were discussed, both relying on the CSP spatial filter for feature extraction. One used a hyperparameter tuned frequency

band-pass filter for preprocessing the signal and the other a fixed one. This led to the discussion of FBCSP, a pipeline that is not implemented but the working of which is addressed. The main focus of this master thesis revolves around one-step DL methods and five different such pipelines were discussed. Three implemented state-of-the-art literature proposed CNN based models were discussed: EEGNet, DeepConvNet and ShallowConvNet. Besides their architectural design, their relation to the FBCSP method was also discussed as well as the possibilities of visualising them. Finally, two CNN-LSTM models were proposed that extend upon EEGNet. One makes use of an additional bidirectional LSTM layer on top of the EEGNet-derived features whilst the other aims to directly incorporate LSTM functionality into the EEGNet architecture by using a bidirectional convolutional LSTM layer. For these models, it was also discussed how potential visualisation of the learned behaviour could be done.

Chapter 5

Moving towards an online BCI system

5.1 Introduction

Whilst the focus of this master thesis is on the evaluation of offline MI EEG classification pipelines, this chapter aims to provide a link of these pipelines to a complete BCI system. Some of the key differences in what is the wanted goal of a BCI system compared to that of a MI EEG classifier are discussed. This includes the difference between classification and regression in the pipeline of a BCI system, working with relatively limited data and computational power and focusing on limiting risk rather than the best overall accuracy. Next, some common steps in going from these proposed offline MI EEG classification pipelines to an online BCI are discussed. Since these are provided as pointers rather than effective guidelines, most explanations in this chapter remain high-level. This includes providing some details on working with continuous data, tricks to improve classification performance, the potential of spatial subsampling and a revision of smart mapping between predicted classes and the actions taken on the external device.

5.2 BCI system try to solve another problem

Whilst the general pipeline for classifying brain signals discussed in Section 3.2 forms the majority of the BCI system pipeline, the components to map predicted class labels to effective actions taken on the external device are missing. Figure 5.1 shows the complete BCI pipeline where a high-level controller for mapping classification labels to specific actions on the external device is present. Whilst rather uncommon in BCI research, it is also possible for a regression model to be used to create a more low-level mapping to the external device (Dillen et al., 2022). The difference between both approaches is discussed in the following section. Some other differences are present, such as a far more complex dataset, relatively low-powered computational hardware and a focus on minimizing risk rather than maximising accuracy. These things are also briefly discussed in the following sections.

5.2.1 Using classification vs regression in a BCI system

As discussed in Section 1.3.2, one of the use-cases for BCIs is their use as a novel interaction method for prosthesis and exoskeleton control. Ideally, the control over such an exoskeleton would

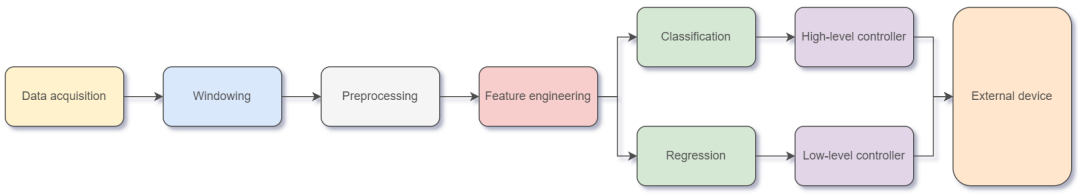


Figure 5.1: Overview of a complete BCI pipeline.

require some sort of continuous mapping rather than a discrete one. A traditional exoskeleton might rely on muscle movement to create a near 1:1 mapping where continuous values are passed to the low-level controller which denotes the angles of the arm and as such the desired angles of the exoskeleton. Regression algorithms can be used to map a feature vector, i.e. from EMG, to a continuous output that denotes such angles.

Since classification is discrete by design, such a near 1:1 mapping is not feasible. Instead, an often limited amount of classes exist to keep learning possible and a high-level controller is used to map one class label with a relatively complex set of actions rather than just one minuscule one. The design of such a system was already shown in Figure 1.8 for a robot arm. Intuitively, the mapping possible between a regressor and a low-level controller could be compared with the control of a joystick allowing fine control of movement in a game. In that setting, classification linked to a high-level controller could be compared with the regular buttons on the game controller which are mapped to an action that performs multiple complex steps upon the press of one button.

An argument could be made that the regressor approach is more intuitive and allows for better control over the external device. However, due to the discussed complexity of brain signals, creating an effective mapping from brain signals to a continuous value is a far harder problem than the classification of only a select few classes. This is further illustrated by a lack of successful literature on proposed regression pipelines for EEG systems as discussed by Dillen et al. (2022).

5.2.2 Complexer data and less powerful systems

The data used for the experiments of this master thesis, and most of the open-source MI EEG datasets available, contain data recorded in near lab-like environments. Section 6.2 discusses the used dataset for the experiments in this master thesis where it will become clear that these recordings are also relatively lab-like. Whilst this makes for clean data, it is by no means representative of the real world. As discussed in Section 2.4.6, many artefacts such as electrode popping and EMG artefacts are expected when the BCI user is walking, talking or even just looking around. This means the test metrics obtained from these-open source datasets are likely to be non-representative of their performance when used in a real-world setting.

Adding to this, the portability and affordability desired for a BCI system mean that the available hardware is often as cheap as possible, especially for the processing unit. As such, complex deep learning models requiring specific GPUs or drivers to be present are non-feasible for real-world use. This causes a problem for the convolutional LSTM extension to EEGNet proposed in this master thesis, as is further discussed in Section 6.9 where some additional pilot studies are performed.

5.2.3 Limiting risky errors

As discussed in Section 6.4, reporting on test accuracy alone has little meaning for a BCI setting, where additional metrics such as the PPV are desired to obtain a complete insight on the model's performance. This is particularly important when moving to an online system. A system having 80% accuracy but where all misclassification results in the passive case being predicted, where no action is performed, is far more desirable than one with an accuracy of 90% but where it performs a risky action 10% of the time during misclassification. Traditional optimisation of a model will typically only look at accuracy or loss metrics and as such doesn't take into account this information. There exist multiple ways of calculating the risk-related metrics of a classifier. For the evaluation of the pipelines in this master thesis, Section 6.4 details three such metrics that give insight into the risky behaviour of the classifier.

5.3 Common steps towards an online BCI system

With the general difference between creating, training and optimizing a classifier and a complete BCI system discussed, some proposals are made on how to adopt the classification pipeline to consider these aspects.

5.3.1 Working with continuous data

Due to the use of ERS and ERD potentials by using MI, which can be induced without an external stimuli, there is no specific event onset (see section 2.4.4). As such, the windowing strategy used in the experiments of this master thesis is not applicable. Whilst it could be made to work, by providing an event onset indication on regular time intervals when the user is allowed to perform a MI task for classification, it is not the most attractive solution. As such, sliding windows, preferably overlapping, should be used. However, such a window design makes the problem complex to learn and classify. As such, some of the evaluations discussed in Chapter 6 should ideally be reproduced using such sliding windows to finetune parameters and gain insight into the model performance when using sliding windows.

5.3.2 Improving classification performance and adding more classes

Even when the classification results obtained are satisfactory, small risky errors can already be troublesome for some risky BCI systems. As such, additional tricks should be put into place to increase the performance of these models w.r.t. risky misclassification. For example, if using sliding windows and assuming that misclassification is independent of the previous sliding window, requiring three consecutive classifications of the same class before performing a specific action can reduce the original misclassification of that class from 20% to just 0.8% ($0.2 * 0.2 * 0.2$). However, the assumption that misclassification is independent of the previous sliding window does not hold in real life, so the actual misclassification decrease will be less. Alternatively, before performing a potentially risky action this action could be communicated to the user in such a way that the user can still cancel it. This cancelling can then be done by some of the easier to detect EMG artefacts discussed in section 2.4.6 as these have high classification accuracy. Another interesting technique is using the ROC curve curve for determining an ideal prediction threshold such that the FPR of risky actions is minimal as was discussed in Section 3.4.

5.3.3 Calibrating the classification pipeline before use

It was discussed that brain signals are non-stationary oscillatory signals that can vary greatly between subjects and even between sessions of the same subject. Besides using techniques to reduce risky misclassification, it is often the case that calibration of a DL model is performed. The main idea behind calibration is using a data acquisition method for obtaining labelled samples of the subject wanting to use the BCI before actually trying to predict for that session of the subject. These additional labelled samples are often obtained in less than 20 minutes (Lawhern et al., 2016). After being obtained, a previously trained model, either intersession or intersubject, is loaded and the additional data is fed to the model for further training. The idea is that this reduces an intersubject or intersession problem to an intrasession problem, as the classifier has access to the data from the same session it asked to predict samples for after calibration. Since this calibration data is often limited, a model can sometimes overfit the training data in such a small amount of time that the validation scores won't increase. This often requires stopping certain parts of a model from learning, *freezing* them, and only training the last part(s) of the model further on the additional data using a very low learning rate.

Calibration can have varying results, if too few samples are present and the model overfits right away neither validation nor test accuracy will increase and they often even decrease. If too many new samples are present, it might be better to just train a completely new shallow model using an intrasession approach. Finding the sweet spot is not an easy task and often it is too long in time for a user to enjoy the process of calibration each time they want to use the BCI system. For these reasons, data augmentation techniques for EEG are being proposed which aim to artificially increase the number of samples obtained from a short calibration period. Fahimi et al. (2021) has shown the use of a generative adversarial network (GAN) to augment such EEG data with satisfactory results.

5.3.4 Spatial subsampling

As discussed in section 2.4.2, when it is known in which region of the brain the signals of interest are expected to originate, the channels can be subsampled to only include those that are of interest. This decrease the complexity, making the prediction computationally easier to perform and thus easier to execute on the affordable and portable hardware desired from a general BCI system. Additionally, removing the frontal electrodes will remove the channels most related to an artefact from facial muscle movement and eye movement. This could help in reducing the added data complexity by removing some of the artefact-filled channels.

5.3.5 Configuring the high-level controller

Once the MI EEG pipelines are reconfigured to work better for the use of a sliding window in an online BCI setting, the mapping from classification labels to actions may be done. Not only should this mapping take into account the PPV of a class to determine which is most fit for taking risky actions, but the mapping and interaction with the external device itself should also be carefully designed to require the least possible steps to achieve the desired result. This was already explained with the robotic arm example in Section 1.4.2.

Once an appropriate mapping is determined, implementing it such that the external device does the desired action can be a difficult task since many devices often have a different way of being programmatically controlled (interfacing method). As discussed by Dillen et al. (2022), there are currently proposals being made for standardized platforms such that multiple external devices such as exoskeleton and prosthesis share an equal interfacing method. These standardized

platforms can also allow for simulated interaction, which can be helpful when combined with VR to let a new user test if a certain external device would work as desired.

5.4 Chapter conclusions

This chapter gave a brief overview of the challenges present when going from an offline MI EEG classification pipeline to an online BCI system. Some broad steps were discussed that could be performed to facilitate this move to an online BCI system. In the next chapter, some pilot studies are also discussed that provide initial results of some of the changes proposed in this section.

Part III

Reflection on the results of this master thesis

Chapter 6

Evaluation of the proposed pipelines

6.1 Introduction

The previous chapters discussed in great detail which MI EEG classification pipelines were considered for this master thesis, how they work and how they can be evaluated. This chapter details the performed evaluations and the results obtained. In particular, three main test settings were used: intrasession, intersession and intersubject evaluation. Besides providing an overview of the obtained evaluation metrics, some reasoning into why certain experiments were performed and why certain results were expected is also given. The used open-source MI EEG dataset is also briefly discussed. Most experiments revolve around offline MI EEG classification performance, as this is the focus of this master thesis. Some pilot experiments regarding recommended changes for going to an online setting discussed in Chapter 5 were also performed and are also discussed in this section.

6.2 Three class MI EEG data source

Following this master thesis' proposal of splitting BCI development in at least four distinct papers (Section 1.4.4), this master thesis focusing on the classification pipeline should provide a summary of the data used. The MI EEG data used for this master thesis is from an open-source dataset by Kaya et al. (2018). Kaya et al. (2018) provide four different types of MI interaction tasks, of which this master thesis works with the three class MI EEG dataset provided as the "CLA" dataset.

The hardware used for the data acquisition of this dataset uses the international 10-20 system discussed in 2.4.1. The used EEG equipment makes use of wet electrodes and a sampling rate of 200Hz. The participants were seated in a comfortable position throughout the experiment and remained motionless throughout the recordings, with a fixed gaze-fixation point to limit the presence of muscle artefacts which were discussed in Section 2.4.6. The provided data is band-pass filtered to include the frequencies between 0.53Hz and 70Hz. A band-stop filter was also in place to remove AC artefacts as discussed in Section 2.4.6. All of these frequency filters are hardware filters directly integrated into the hardware of the EEG-1200 hardware used for data acquisition.

The recordings of the CLA dataset provides balanced sampels of three MI tasks reffered to

as left-hand MI, right-hand MI and neutral MI. For the left-hand and right-hand MI tasks, the participants were asked to imagine closing and opening the respective fist once. For the neutral MI task, the patient was asked to perform no MI. The communication of which task to perform was done via a simple graphical interface. Upon showing the icon for which task to perform (event onset), the subject should perform the specific MI task. The icon is visible on the screen for one second past the event onset. A random resting period of 1.5 to 2.5 seconds was present between the event offset and the event onset of a new task. This process was repeated for 15 minutes, after which a longer resting period was present where it was ensured the EEG equipment remained properly seated. In total, three 15-minute trials were performed in one session. This resulted in roughly 300 samples for each MI task for each session. Some subjects only had one recorded session whilst others had up to three. All subjects were healthy, between 20 and 35 years old and living in Turkey. No specific survey was performed to test the MI capabilities beforehand nor does Kaya et al. (2018) describe that any specific MI training happened.

For the experiments of this master thesis, only subjects B, C and E of the CLA dataset were considered. The motivation for this is them being the only ones with three different sessions available. Given that there is no mention of MI training by Kaya et al. (2018), it was reasoned that those who did three sessions would likely be better at MI, especially in their last session. Given the experience and feedback from two previous trials, it was also reasoned these subjects would create fewer muscle artefacts that could be unwillingly used for learning, such as those discussed in Section 2.4.6. As such, the test sessions for these subjects were always the last recorded session for the experiments of this master thesis.

The data from the CLA dataset is provided as MatLab files originally. To make them usable in Python, the experimental notebooks provided on the GitHub repository of this project provide conversion methods to convert these MatLab files to MNE Raw objects (Bontinck, 2021; Gramfort, 2013). To facilitate the use of this dataset in Python for other researchers, the GitHub repository also provides the CLA dataset as functional image file format (FIF) files. These FIF files can be easily opened with MNE Python and include all provided details by Kaya et al. (2018) stored in the associated info object. The utility file `CLA_dataset.py` provides many functions for working with this CLA dataset in Python.

6.3 General remarks on experiment setups

To provide the best possible reproducibility all resources needed to recreate the empirical evaluation results are available on the documented GitHub repository of this master thesis (Bontinck, 2021). This includes saved weights of the exact models used for the results discussed in this chapter and many utility files to facilitate loading and working with these models.

All training and testing was done on a 64-bit Windows 10 Pro machine with an Intel Core i5-4690K overclocked at 4.2GHz. An MSI NVIDIA GeForce GTX970 GPU was used in the training of the DL models with CUDA support through the CuDNN 8.2.2 driver. The system in question had 16GB of ram, of which an approximate 10GB was usable for the training procedure. These specifications allowed for sufficient training of the two-step ML and state-of-the-art CNN models in a reasonable time. The longest cross-validated grid search hyperparameter tuning experiment for the two-step ML approaches took around 30 hours to finish. The longest experiment for the literature proposed CNN models was finished in less than 6 hours. The LSTM models were computationally heavy, and the convolutional LSTM extension to EEGNet does not have CUDA support. This resulted in longer training times, with manageable times of under 6 hours for the regular LSTM extended EEGNet model. However, for the convolutional LSTM extend EEGNet model, certain configurations were not possible due to limited video RAM and fewer epochs

were performed for training. The latter didn't seem to influence the results as the evolution of both training and validation accuracy and loss converged in all experiments. These metrics were closely monitored during all training of the DL methods to ensure no overfitting tendencies are present. En example of validation loss and accuracy evolution is given in Figure 6.1.

The used representation of the EEG signal provided through MNE makes use of extremely small values. This causes issues for the DL approaches and as such the data was first multiplied by 1000000 before using it in the DL classification pipelines. Each experiment made use of a fixed length window around a known event as described in Section 3.2.2. Two window lengths were considered, a 0.5-seconds window used for all experiments and a longer 1.5-second window additionally tested for most experiments. The 0.5 seconds window started 0.1 seconds after the event onset and ended 0.6 seconds after the event onset. The 1.5 seconds window started 0.25 seconds before the event onset and ended 1.25 seconds after the event onset. This chapter discusses the results of the 0.5-seconds unless stated otherwise.

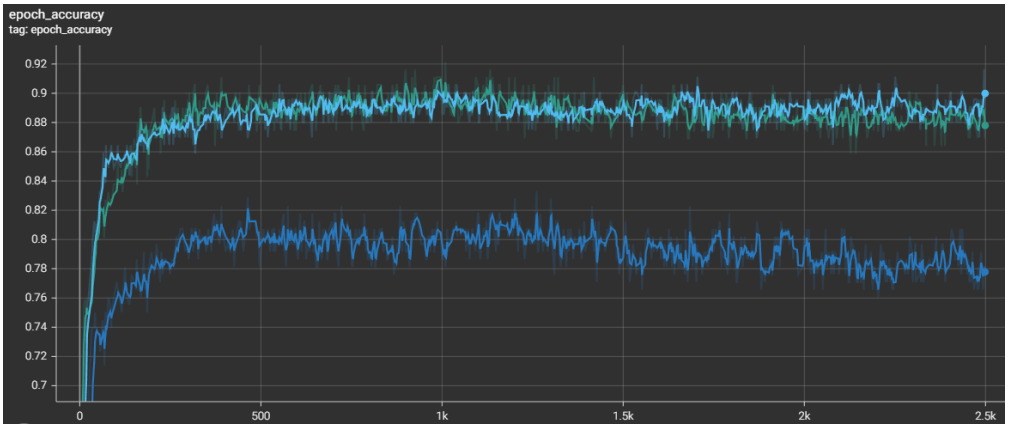
The traditional two-step ML approaches made use of cross validated gridsearch hyperparameter tuning to find the best possible combination of parameter settings, as discussed in Section 4.2.2. Due to the computational power that would be needed to perform such a cross-validated grid search for hyperparameter tuning of the DL methods, parameters were chosen based on the literature proposed configurations and finetuned based on pilot studies performed in the experimental notebooks. This also means that whilst the traditional two-step ML approaches mostly have 4-fold validation accuracy reported together with their standard deviation, the DL models only have a singular validation accuracy available obtained from a typical 20% split during Keras training.

The DL models were trained for sufficient epochs, 250 to 2500 dependent on the data and pipeline, where both training scores and validation scores on basis of accuracy and loss were monitored to ensure stable training. For each training procedure of the DL models, two variants of the model were saved. One where the validation loss was lowest and one where the validation accuracy was the highest. These variants are used for obtaining the final results on the test sets, which were equal for all models and 20% of the total available data. The best variant of these two is the one reported in this chapter. Due to the use of balanced data, the accuracy of a random model would be $\frac{1}{3}$ in all experiments.

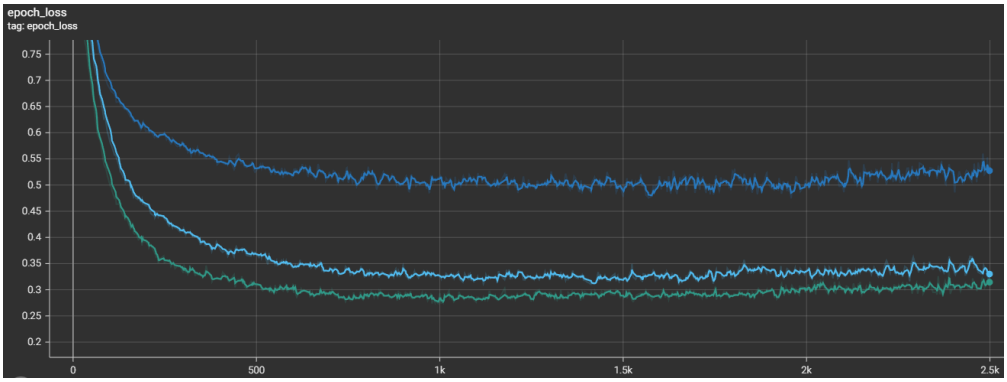
These general remarks should suffice for a precise understanding of the results discussed in the remainder of this chapter. As discussed, all learned models and the code for all performed experiments is available on the GitHub repository of this master thesis to ensure complete reproducibility (Bontinck, 2021).

6.4 Considered evaluation metrics

Considering the goal of these MI EEG classification pipelines is to ultimately be used in a BCI application, reporting on the validation and/or testing accuracy alone doesn't tell that much. From the alternatives provided in Section 3.4, three additional metrics derived from the test set's CM are discussed for the main experiments of this master thesis. The first two provide the PPV for both left-hand and right-hand MI class labels. These metrics are interesting, as they provide insight into how many times an action linked to this label would be performed with the user's intent. By doing so, it also directly provides how many times an action was taken without the user's intent, as this is just 100% - the metric. The higher this metric, the better the model is in this regard. As a more general idea of how many times the system would make a *risky error*, another metric is also proposed. This metric represents how many times either the left-hand or right-hand MI class label is wrongly predicted out of all predictions. A lower value thus denotes



(a) Evolution of validation accuracy over time.



(b) Evolution of validation loss over time.

Figure 6.1: Evolution of both validation accuracy and validation loss over time. Graphs shown are from the intrasession experiment of EEGNet for subjects B (dark blue), C (light blue) and E (green). The time axis denotes the epoch in which the result is obtained.

a better model in this regard. This metric is included as it is of great important when moving the system to an effective online BCI as discussed in Section 5.2.3. The best performing model in each of these metrics has its score for the metric in bold. All of the CMs used to obtain these results can be found on the GitHub repository of this master thesis (Bontinck, 2021).

6.5 Intrasession evaluation

The first experimental setup consisted of an intrasession evaluation. For such an experiment, each classification pipeline is trained and tested on MI EEG data from the same subject and the same session. This experimental setup is the easiest to learn as it is least influenced by the generalisability issues of MI EEG data discussed in Section 2.4.5. Whilst in other fields this approach would be considered data leakage, a phenomenon discussed in Section 3.4, it is a relatively common setup for this kind of research. The main defence for this is that most real-world applications of MI require up to 20 minutes of calibration anyway (Lawhern et al., 2016). The models are trained on 60% of the data, validated on 20% of the data for finding the

Model	Test subject	validation accuracy	Test accuracy	Left MI PPV	Right MI PPV	Risky error	Non-default parameter settings
CSP + RF	B	65,88 +- 3,16	59.38	67.19	64.71	20.31	4 CSP components | RF max depth 10, max features 0.4, min sample split 10, 50 estimators
AutoFreq + CSP + LDA	B	66.15 +- 3.01	60.42	66.15	61.02	23.44	2 - 30 Hz filter | 4 CSP components | LDA SVD solver with 0.0001 tol
CSP + LDA	B	66.15 +- 5.04	60.94	66.15	62.71	22.92	6 CSP components | LDA SVD solver with 0.0001 tol
CSP + SVM	B	66.93 +- 2.98	61.46	68.85	63.16	20.83	4 CSP components | SVM RBF with C 0.1 and Gamma auto
BiConvLSTM EEGNet	B	81.77 @ epoch 448	70.83	81.36	68.25	16.15	40 LSTM filters | LSTM kernel size 9, LSTM dropout 0.7
BiLSTM EEGNet	B	80.21 @ epoch 75	71.35	78.46	69.84	17.19	/
ShallowConvNet	B	72.92 @ epoch 1385	72.40	76.67	71.88	16.67	0.8 dropout
EEGNet	B	83.33 @ epoch 1262	77.08	84.75	71.01	15.10	/
DeepConvNet	B	80.73 @ epoch 2243	**80.73**	**89.29**	**85.19**	**7.29**	/
CSP + RF	C	71.19 +- 3.16	70.83	75.00	75.81	15.63	6 CSP components | RF max depth 3, max features 0.4, min sample split 5, 250 estimators
AutoFreq + CSP + LDA	C	71.71 +- 3.49	71.88	77.59	72.73	16.15	2 - 30 Hz filter | 10 CSP components | LDA SVD solver with 0.0001 tol
CSP + LDA	C	71.44 +- 3.41	72.40	77.59	73.85	15.63	10 CSP components | LDA SVD solver with 0.0001 tol
CSP + SVM	C	72.62 +- 2.98	74.48	80.33	76.19	14.06	6 CSP components | SVM RBF with C 100 and Gamma 0.001
ShallowConvNet	C	89.06 @ epoch 187	86.46	88.24	85.94	8.85	40 LSTM filters | LSTM kernel size 9, LSTM dropout 0.7
BiConvLSTM EEGNet	C	88.54 @ epoch 78	88.54	**93.75**	86.36	6.77	0.8 dropout
DeepConvNet	C	90.10 @ epoch 301	88.54	88.24	87.69	8.33	/
EEGNet	C	91.67 @ epoch 2497	89.06	87.32	91.80	7.29	/
BiLSTM EEGNet	C	89.06 @ epoch 1143	**90.63**	88.41	**93.44**	**6.25**	/
CSP + SVM	E	72.51 +- 1.76	70.16	75.00	66.67	19.90	10 CSP components | RF max depth None, max features 0.2, min sample split 2, 250 estimators
CSP + LDA	E	75.92 +- 0.98	72.77	76.92	72.88	16.23	1 - 34 Hz filter | 10 CSP components | LDA SVD solver with 0.0001 tol
CSP + RF	E	73.56 +- 1.59	73.30	80.00	71.43	15.71	6 CSP components | SVM sigmoid with C 100 and Gamma 0.01
AutoFreq + CSP + LDA	E	73.42 +- 1.71	73.82	74.24	76.36	15.71	10 CSP components | LDA SVD solver with 0.0001 tol
BiLSTM EEGNet	E	92.15 @ epoch 339	82.72	83.33	87.10	9.95	40 LSTM filters LSTM kernel size 9, LSTM dropout 0.7
BiConvLSTM EEGNet	E	90.58 @ epoch 478	86.39	**91.80**	90.00	5.76	/
EEGNet	E	91.62 @ epoch 338	86.91	90.16	88.89	6.81	0.8 dropout
ShallowConvNet	E	92.15 @ epoch 1010	88.48	88.89	**93.55**	5.76	/
DeepConvNet	E	92.15 @ epoch 221	**88.48**	90.63	93.10	**5.24**	/

Figure 6.2: Intrasession results for all classification approaches. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches.

best parameter configurations and tested on the test set consisting of 20% of the data.

Figure 6.2 summarizes all obtained results in terms of validation and test accuracy along with the three other CM derived metrics discussed in Section 6.4. Given the limited amount of training data, around 200 samples per class, a reasonable assumption would be to think that a deep DL model such as DeepConvNet would overfit and perform poorly on the test set and that a shallow DL model such as ShallowConvNet would perform better. However, given the test set is from the same subject in the same session, the test data matches the training data especially well and DeepConvNet is the most consistent approach in terms of accuracy. It has a mean test accuracy of 85.92 with a standard deviation of $\pm$ 3.67 compared to ShallowConvNet's 82.45 $\pm$ 7.15. It is noteworthy that subject B seems to be more difficult to learn compared to the other two subjects. This finding is consistent between all experiments and could potentially be related to the subject's MI capabilities. However, since no survey to test these capabilities was performed by Kaya et al. (2018), it can't be validated.

All the DL models outperform the two-step CSP approaches. However, given the simplicity of CSP, the obtained results are still pleasant and extensions such as FBCSP are expected to perform even better. With a risky error below 10% for all patients when using DeepConvNet, and having a PPV for both left and right-hand classes, the model could have great potential for use in a comparable BCI setting. Multiple tricks, such as requiring multiple successive classifications of the same class before acting, can help in reducing the risky error even more but fall outside the scope of this master thesis.

6.6 Intersession evaluation

The second experimental setup is more representative of a real-world application. By training on the first two sessions of a subject and testing on the third, an intersession evaluation is performed. It is noted that both sessions were used as training data, with a subsample of both being used as validation data. A more representative validation strategy would be to use one session as training data and one as validation data. However, this would limit the training set to one session, for which it is reasoned that the generalisability of the model will suffer due to it not having access to different sessions and thus a greater variety of data. Following this reasoning, which was further supported by a pilot study, it was chosen to use 60% of both sessions as training data, 20% as validation data and 20% as test data.

Figure 6.3 summarizes all obtained results similar to the summary from the intrasubject experiment. Both DeepConvNet and EEGNet perform equally well, with a mean test accuracy of 70.15 $\pm$ 3.97 and 69.84 $\pm$ 2.94 respectively. The difference between EEGNet and the extension proposed is slim to none when looking at the metrics provided. Whilst both EEGNet and ShallowConvNet claimed to be inspired by the same FBCSP approach that is known to perform well for these tasks, they differ significantly in some of the reported evaluation metrics. This is most notable for subject E, where EEGNet has a risky error of less than half that of ShallowConvNet. Likewise, the PPV for left-hand MI is significantly higher in favour of EEGNet. This suggests that EEGNet makes most of its mistakes on the false passive MI predictions, whereas ShallowConvNet does so on false left-hand MI predictions. This suggests that EEGNet and ShallowConvNet are learning different features.

The best-achieved test accuracy of 65.94 for subject B, obtained by the bidirectional convolutional LSTM extension to EEGNet, is likely too low for most BCI applications, even with metrics in place to limit risky errors. The test accuracies hovering around 73% for the other two subjects, obtained by both DeepConvNet and EEGNet, are more promising in this regard and given some error-reducing metrics are in place, it would allow for use in BCI applications with

Model	Test subject	validation accuracy	Test accuracy	Left MI PPV	Right MI PPV	Risky error	Non-default parameter settings
AutoFreq + CSP + LDA	B	48.07 +- 1.66	34.90	34.17	32.56	57.81	4 - 30 Hz filter | 10 CSP components | LDA SVD solver with 0.0001 tol
CSP + RF	B	44.89 +- 3.51	43.13	61.15	49.11	12.29	10 CSP components | RF with max depth 3, 0.4 features, 10 min sample split, 500 estimators
CSP + LDA	B	45.00 +- 2.57	46.77	47.92	53.42	30.00	CSP 10 components | SVD LDA with 0.0001 tol
CSP + SVM	B	46.25 +- 2.76	46.77	48.84	58.47	23.44	10 CSP components | rbf SVM with C 10 and gamma 0.01
ShallowConvNet	B	0.651 @ epoch 2151	62.08	61.79	65.75	24.58	/
BiLSTM EEGNet	B	0.6458 @ epoch 177	63.02	68.59	70.54	18.13	/
DeepConvNet	B	0.6753 @ epoch 1220	64.58	71.62	**72.96**	**15.31**	/
EEGNet	B	0.6927 @ epoch 1473	65.73	71.52	71.81	16.77	/
BiConvLSTM EEGNet	B	0.6545 @ epoch 300	**65.94**	**71.55**	68.15	19.06	40 LSTM filter, 0.7 LSTM dropout
CSP + RF	C	81.98 +- 1.98	32.33	23.53	33.00	67.67	10 CSP components | RF with max depth None, 0.2 features, 2 min sample split, 50 estimators
AutoFreq + CSP + LDA	C	82.97 +- 2.96	35.66	43.87	34.08	64.34	2 - 28 Hz filter | 10 CSP components | LDA lsqr solver
CSP + LDA	C	81.77 +- 1.94	35.87	43.67	34.33	64.13	CSP 10 components | SVD LDA with 0.0001 tol
CSP + SVM	C	83.38 +- 2.13	37.54	42.75	35.65	62.46	10 CSP components | rbf SVM with C 1 and gamma auto
ShallowConvNet	C	0.8247 @ epoch 1115	67.26	78.71	64.18	18.87	/
BiLSTM EEGNet	C	0.8681 @ epoch 947	68.58	77.12	60.00	23.80	/
BiConvLSTM EEGNet	C	0.8663 @ epoch 300	68.72	77.97	64.31	20.44	40 LSTM filter, 0.7 LSTM dropout
EEGNet	C	0.8837 @ epoch 1159	71.32	76.26	66.84	21.90	/
DeepConvNet	C	0.8663 @ epoch 1266	**72.37**	**79.40**	**68.96**	**18.25**	/
CSP + SVM	E	58.16 +- 2.55	38.95	40.59	56.94	18.12	10 CSP components | rbf SVM with C 1 and gamma scale
AutoFreq + CSP + LDA	E	56.86 +- 1.56	47.85	45.32	69.40	39.79	1 - 28 Hz filter | 10 CSP components | LDA SVD solver with 0.0001 tol
CSP + RF	E	57.70 +- 1.98	49.01	46.46	66.67	35.92	10 CSP components | RF with max depth 10, 0.4 features, 10 min sample split, 250 estimators
CSP + LDA	E	55.25 +- 3.67	55.18	54.04	61.60	32.67	CSP 10 components | SVD LDA with 0.0001 tol
BiConvLSTM EEGNet	E	0.7431 @ epoch 191	64.92	58.66	92.16	23.25	40 LSTM filter, 0.7 LSTM dropout
ShallowConvNet	E	0.7292 @ epoch 138	65.97	55.39	93.22	26.81	/
BiLSTM EEGNet	E	0.7361 @ epoch 166	67.02	66.59	91.16	16.65	/
EEGNet	E	0.7674 @ epoch 782	72.46	**73.18**	92.35	**12.57**	/
DeepConvNet	E	0.7483 @ epoch 79	**73.51**	69.53	**95.96**	13.82	/

Figure 6.3: Intersession results for all classification approaches. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches.

no risks.

The traditional two-step ML approaches using CSP-based feature extraction struggle at learning a generalizable model. When looking at the difference between validation accuracy and test accuracy, especially notable for subject C, it is clear the models are overfitting. Whilst this is a trend for all models on subject C, suggesting its last recording is less in line with its first two, it is extra apparent for the CSP approaches. Their test accuracies obtained by the CSP methods for subject C are similar to that of random prediction (33%).

6.7 Intersubject evaluation

The final and most difficult-to-learn experimental setup is an intersubject one, where the classification pipelines are tested on the last session of one subject and trained on all sessions of the other subjects. Having a model that works even reasonably well in this experimental setup is considered a pleasing results as the common generalisability issues discussed in Section 2.4.5 and the more difficult nature of MI classificiation discussed in Section 2.4.4 make this a complex task.

Like the reasoning provided for not keeping one of the two training sessions as a validation session, it is chosen to use both subjects in training and have a random sample split from this data as a validation split. This is done since it was thought that opting to have a better validation strategy was expected to result in poorer performance due to the reduction in data. Figure 6.4 summarizes all obtained results similar to the summaries before. With a mean test accuracy of 63.61 ± 1.83, the bidirectional convolutional LSTM extension to EEGNet performs best. Whilst this is promising for the architecture proposed but this master thesis, it is without any statistical significance as the other LSTM extension, EEGNet and DeepConvNet hover around the same mean test accuracy with slightly larger standard deviations.

More interestingly, for neither of the three subjects is the classification pipeline with the best test set accuracy the one with the best score for any of the other CM based metrics. This further illustrates how accuracy alone doesn't provide much information. This is especially noteworthy for the DeepConvNet model of subject B, where the misclassification rate for risky actions is almost twice as low compared to the CNN-LSTM model which had the best test accuracy. Whilst none of the pipelines are trained to favour false passive MI misclassification, it is interesting to see that some models do this significantly more than others.

Another interesting finding is the EEGNet extended test accuracy obtained for subject B. Not only is it the best for that subject, but it also outperforms the intersession results for that subject. This illustrates that intersession variability for one subject can be incredibly high as it is more beneficial to learn from two distinct others to have more variable features extracted. In summary, the intersubject accuracy obtained and the PPV and risky errors provided are pleasing for most DL models, but they are most likely too poor for use in an online BCI system.

6.8 Longer window lengths

It was discussed in Section 3.3.4 that networks using LSTM functionality are an interesting choice for time series such as EEG and it was deemed possible an improvement could be seen. However, none of the discussed main experiments provided statistically significant improvements for the extended EEGNet models over the base one. When reflecting on the experimental setups, the use of a relatively short fixed window around the known event is likely the reason that added memory and sequence understanding doesn't have a significant benefit for the experiments. With the windows starting 0.1 seconds after the event onset and being only 0.5 seconds long, it is

Model	Test subject	validation accuracy	Test accuracy	Left MI PPV	Right MI PPV	Risky error	Non-default parameter settings
CSP + SVM	B	60.39 +- 1.04	38.57	40.77	38.94	36.55	10 CSP components | RBF SVM with C 10 and gamma auto
AutoFreq + CSP + LDA	B	58.53 +- 1.66	38.99	41.94	36.26	50.52	1 - 26 Hz filter | 10 CSP components | LDA LSQR solver
CSP + RF	B	59.83 +- 0.68	39.09	40.06	38.66	42.63	10 CSP components | RF with None max depth, 0.2 max features, 2 min samples split and 500 estimators
CSP + LDA	B	56.62 +- 1.29	39.61	42.24	41.22	32.66	10 CSP components | SVD LDA with 0.0001 tol
ShallowConvNet	B	0.7625 @ epoch 545	59.69	60.94	60.44	27.92	0.4 dropout
DeepConvNet	B	0.7543 @ epoch 1554	62.50	**77.68**	**69.55**	**13.13**	/
BiConvLSTM EEGNet	B	0.7497 @ epoch 82	62.60	61.94	64.57	26.25	40 LSTM filters, kernel size 9, 0.7 LSTM dropout
EEGNet	B	0.7654 @ epoch 92	64.79	67.09	65.96	22.60	/
BiLSTM EEGNet	B	0.7648 @ epoch 128	**66.35**	71.24	64.02	22.40	/
CSP + SVM	C	51.69 +- 44.11	44.11	51.88	41.49	37.93	10 CSP components | RBF SVM with C 10 and gamma scale
AutoFreq + CSP + LDA	C	47.91 +- 0.19	44.32	67.28	41.33	29.21	2 - 34 Hz filter | 10 CSP components | LDA SVD solver with 0.0001 tol
CSP + RF	C	50.40 +- 1.97	44.53	52.17	40.94	40.95	10 CSP components | RF with None max depth, sqrt max features, 10 min samples split and 250 estimators
CSP + LDA	C	47.81 +- 1.85	47.31	61.58	42.47	35.39	10 CSP components | lsqr LDA
BiLSTM EEGNet	C	0.7486 @ epoch 6	59.85	72.79	**65.52**	**16.37**	/
EEGNet	C	0.6993 @ epoch 458	59.96	**77.65**	51.13	28.57	/
ShallowConvNet	C	0.7196 @ epoch 57	60.58	72.89	51.98	27.95	0.4 dropout
DeepConvNet	C	0.7132 @ epoch 990	61.94	75.51	58.40	23.04	/
BiConvLSTM EEGNet	C	0.7329 @ epoch 50	**62.04**	77.29	54.74	24.82	40 LSTM filters, kernel size 9, 0.7 LSTM dropout
CSP + SVM	E	57.36 +- 2.20	33.81	34.25	30.74	65.95	10 CSP components | RBF SVM with C 1 and gamma scale
CSP + RF	E	57.20 +- 2.74	37.08	38.08	35.82	62.47	10 CSP components | RF with 10 max depth, log2 max features, 10 min samples split and 250 estimators
AutoFreq + CSP + LDA	E	55.03 +- 1.68	40.77	42.47	38.91	58.15	2 - 28 Hz filter | 10 CSP components | LDA SVD solver with 0.0001 tol
CSP + LDA	E	55.67 +- 2.87	40.98	44.31	38.72	58.84	10 CSP components | SVD LDA with 0.0001 tol
ShallowConvNet	E	0.7841 @ epoch 22	52.67	51.23	52.10	43.46	0.4 dropout
EEGNet	E	0.7454 @ epoch 889	64.19	68.45	73.95	17.59	/
BiLSTM EEGNet	E	0.7182 @ epoch 72	65.34	68.03	75.24	17.70	/
BiConvLSTM EEGNet	E	0.724 @ epoch 240	66.18	**74.91**	**84.36**	**10.68**	40 LSTM filters, kernel size 9, 0.7 LSTM dropout
DeepConvNet	E	0.7656 @ epoch 835	**66.81**	67.29	83.87	16.54	/

Figure 6.4: Intersubject results for all classification approaches. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches.

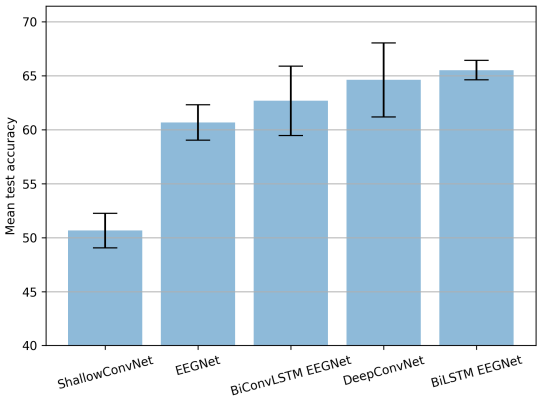


Figure 6.5: Mean test accuracies and their standard deviation from the longer window experiments in an intersubject evaluation setting.

unlikely for there to be unneeded segments of time that can be recognized by the LSTM. Due to the fixed window location around the event and the relatively constant reaction speed expected from the participants, there are also no major timeshifts expected in the signal. As such, the local patterns learned by the CNN models whose relatively wide kernel goes over the temporal axis, are likely to capture all meaningful information. Adding to this, the use of two convolutions over the time axis by EEGNet, the last convolutional layer using a downsampled and temporal feature extracted output of the first and second convolutional layer has access to a broader time range. It could be argued that this also provides some forms of memory as it is looking for local patterns in data that is already feature extracted on a smaller time window for local patterns.

However, considering that the difference between the CNN models and the LSTM extended EEGNet models became more favourable for the latter when complexity increased for the intersubject setting suggests that LSTM might be learning meaningful additional information. As a first step to test this theory, a three times longer windowing technique was considered, starting 0.25 seconds before the onset and ending 1.25 seconds after the onset. The summarized results of these longer window experiments are added to the extra figures list at the end of this master thesis. An overall increase in performance for most models and metrics is seen throughout all experiments. This suggests the half-second window doesn't capture all useful information.

Most interestingly are the intersubject experiments using these longer windows, the accuracy results of which are summarized in Figure 6.5. With a mean test accuracy of 65.52 ± 0.89, the LSTM extended EEGNet model performed better than the regular EEGNet model which had a mean test accuracy of 60.68 ± 1.64. although this is only a first step in showing that the additional functionality of LSTM can be helpful, it is a promising one. Further research into this classification approach, especially using sliding windows where temporal insight is even more important, could point to an even more significant finding than this initial one.

6.9 Additional pilot studies

Taking into consideration the steps of moving from an offline MI MI classification approach to an online BCI discussed in Chapter 5, some additional pilot studies were done. Since these are pilot studies, they only aim to provide a deeper insight into the discussed proposals and an indication of what can be expected from future research.

6.9.1 Improving intersubject performance by providing more data

Model	Test subject	validation accuracy	Test accuracy	Left MI PPV	Right MI PPV	Risky error	Non-default parameter settings
BiLSTM EEGNet	B	0.6409 @ epoch 55	71.67	70.51	73.10	20.83	/
EEGNet	B	0.643 @ epoch 13	72.81	72.19	74.43	18.96	F1 12, F2 24
BiConvLSTM EEGNet	B	nknown due to premature sto	74.48	72.78	76.51	18.23	40 LSTM filters, LSTM kernel size 9, 0.7 dropout
DeepConvNet	B	0.619 @ epoch 437	**77.50**	**79.47**	**79.52**	**13.54**	/
EEGNet	C	0.642 @ epoch 1410	59.85	73.05	51.58	32.64	F1 12, F2 24
BiLSTM EEGNet	C	0.6482 @ epoch 56	60.79	73.82	51.05	32.95	/
DeepConvNet	C	0.6169 @ epoch 653	**68.40**	**80.48**	**58.47**	**24.61**	/
BiLSTM EEGNet	E	0.6493 @ epoch 144	83.87	81.67	87.10	11.31	/
EEGNet	E	0.6461 @ epoch 102	86.60	87.10	88.31	8.38	F1 12, F2 24
DeepConvNet	E	0.6138 @ epoch 1085	**91.31**	**91.87**	**93.85**	**4.82**	/

Figure 6.6: Intersubject results for some of the one-step DL classification approaches using more training data. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches.

The main experiments of this master thesis only made use of data from three of the provided subjects in the CLA dataset, as discussed in Section 6.2. An additional pilot experiment was performed to test if adding more training data would increase intersubject performance. Due to the increased data availability, this experiment setup did make use of a distinct subject for the validation split. Figure 6.4 summarizes all obtained results like before. The added data gave significant improvement for all models without having performed additional hyperparameter tuning beyond logical reasoning. The deepest CNN, DeepConvNet, which now has access to sufficient data for training far outperforms all other DL classification pipelines for all subjects. This is not unreasonable given the other models were developed with a small dataset in mind (Lawhern et al., 2016). The results obtained from this intersubject experiment improve upon even the Intersession experiment results discussed in Section 6.6.

With a mean accuracy of 79.09 $\pm$ 9.42 DeepConvNet has a 77.50% accuracy for subject B and 91.31 accuracy for subject E. The latter could directly be used in an online bci setting where MI tasks are expected on a known time interval. The sharp contrast between subject C's test accuracy of EEGnet (68.40%) and that of subject B also further illustrates how different brain signals can be between subject and session. This pilot experiment provided very favourable results and it is clear that given enough data, deep networks such as DeepConvNet can generalise well. The worse but still comparable results of EEGNet and its extensions proposed by this master thesis can be expected, as the shallow structure of EEGNet was made with the idea of having access to few samples in mind. Further research building upon this work or dataset should therefore consider using all data available, perhaps even samples from the other interaction methods provided by Kaya et al. (2018), as these overlap in part with the CLA dataset and thus can provide even more information to classification pipelines.

6.9.2 Comparison of prediction times

When moving towards an online system where multiple overlapping sliding windows are likely to be used per second, the time it takes to make a prediction mustn't be too long as to cause issues for the often low-powered chip available in the processing hardware. Since the one-step DL models were most promising when looking at the experiment results, a small experiment was performed to test the prediction time of each proposed one-step DL pipeline. This was done by letting a trained model make the same 100 predictions 10 times. Dividing this result by 100 gives the time needed per prediction and is less influenced by computational overhead caused by the timing functionality. This experiment was then repeated a further ten times to obtain an idea of the variance of these estimates. The results of this experiment are summarized in Figure 6.7, where the time in seconds needed to predict 1 sample of a 0.5-second long window is shown.

From this experiment, it becomes clear that besides slow training the bidirectional convolutional LSTM EEGNet extension is also more than 20 times slower in prediction time. With a prediction time of over 0.1 seconds on a relatively capable system, it is unlikely that this model can be used for predicting sliding windows on a low-powered chip. As such, a reconsideration of this model's complexity might be required when moving to an online system. The other LSTM extended EEGNet model that was proposed in this master thesis did not suffer from this problem.

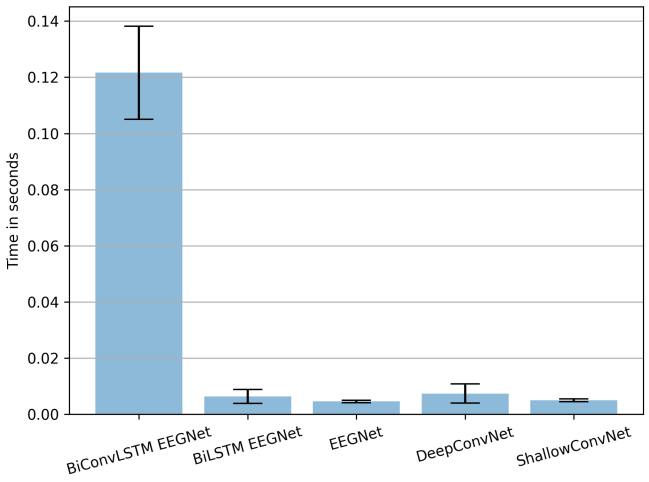


Figure 6.7: Average prediction time in seconds for one 0.5 seconds long window with its standard deviation.

6.9.3 Pilot studies with less promising results

Besides the pilot studies already mentioned, the paper notebook 8 available on the GitHub of the master thesis also provides an initial pilot study on using CSP + LDA on MI EEG data where the electrodes were subsampled based on the proposals of Section 2.4.2. Whilst it has shown to be capable of producing similar results for subject B as those reported in this chapter, the generalisability to intersession or intersubject use of these traditional two-step ML pipelines is not improved. However, further studies to see if the one-step DL methods can achieve comparable results using this subsampled channel selection should be made. If so, this limits the dimensionality of the data considerably and will facilitate the learning process from a technical

perspective and lower the prediction times reported in Section 6.9.2.

Another pilot study involved calibrating the DL models on a new session by freezing the feature extraction layers and letting the fully connected layers learn further at a low learning rate using 5 minutes' worth of data. Other variants of this freezing scheme were also used. This produces mixed results as even with high dropout rates and regulations, the models seem to directly overfit on the new data samples and they do not boost the accuracy of the validation or test set by any significant means. As discussed by Lawhern et al. (2016), most calibration done for MI modelling has data available in the order of 20 minutes rather than 5. Thus, using augmentation methods on the EEG from 5 minutes might be a good approach to calibration using only a small amount of data. Fahimi et al. (2021) proposed the use of a GAN for this with promosing results.

6.10 Chapter conclusions

This chapter discussed the different experiments performed on the seven classification pipelines. Three main evaluation paradigms were considered: intrasession, intersession and intersubject classification. Besides accuracy, two other CM derived metrics were used for reporting the classification results. Thes additional metrics give insight into the models' performance for predicting labels that could be associated with risky actions in an online BCI.

When using a half-second window no statistically significant results were found that suggest the added memory provided by the LSTM layer benefits the classifier. After discussing how this might be due to the data not being as prone to time shifts due to the use of a fixed windowing technique and a small window, a larger window was evaluated for the same three evaluation paradigms. From this, it was found that the LSTM extended EEGNet model with a mean test accuracy of 65.52 ± 0.89, performed better than the regular EEGNet model which had a mean test accuracy of 60.68 ± 1.64 over all three subjects. This suggests that the LSTM layer may still add increased capabilities to EEGNet, but that they require a different window strategy and further research to become better apparent.

Besides these main experiments, some pilot studies were also discussed. One of these pilot studies showed that given enough data, deep DL methods such as DeepConvNet can achieve more than 90% classification accuracy for a three-class MI problem in an intersubject evaluation paradigm. These are great findings pushing towards further research to be done. Some other potential promising extensions, such as reducing the complexity of the training by reducing the number of channels used and the use of augmented EEG data for calibration, were also discussed but not favourably evaluated by the pilot experiments.

Chapter 7

Conclusions, future work and self-reflection

7.1 Introduction

This master thesis set out two major goals, to provide a document that facilitates the entry into the BCI field for computer scientists and to evaluate different types of offline MI EEG classification pipelines, especially CNN based approaches and the proposed LSTM extensions. The latter aimed to answer the research question if adding LSTM functionality to a state-of-the-art CNN MI EEG classification pipeline can further improve its performance. This chapter reviews to what degree these goals were accomplished and what the found answer to this research question is. Some future work that emerged from the work done in this master thesis is also addressed and the master thesis ends with a personal endnote.

7.2 Are brain-computer interfaces made easy?

One of the two major goals of this master thesis was to facilitate the entry into the BCI field for computer scientists as a steep learning curve deters many that are interested. In this regard, the first chapter of this master thesis is a rather unique approach to accomplishing this goal. Realising that many excellent systematic review articles on the BCI field and related topics already exist, it was chosen to provide a more intuitive introductory chapter into BCI by focusing on some of the field's history, present and future whilst referencing to resources providing a more formal and systematic overview.

In this regard, research from the late 1990s and onwards was studied to derive the most important components in the cycle of increasing popularity for BCI research. A clear upwards trend in publications started around 2000, with researchers primarily focused on using traditional two-step ML in combination with CSP or one of its many extensions to classify many different types of brain signals. As DL saw a spike in popularity, so did its application to brain signals. This is one of many discussed reasons that the 2010s saw an explosion in both research and commercial interest of BCIs. Taking into account the many use cases for BCIs discussed in this master thesis, the appeal of a general BCI system is easy to see. However, some obstacles and open issues are still in the way for BCI usage to see true widespread use in the real world. These challenges and some of the opportunities were therefore also discussed, together with a note on some ethical questions.

From these rather intuitive explanations, a shift was made to a more theoretical and technical description of brain signals, measuring modalities and some of the standards and guidelines used when working with brain signals and EEG in particular. The focus then shifted from a neurological point of view on brain signals and data acquisition to a computer scientist point of view on the classification pipelines BCI systems use. A general pipeline for classifying brain signals with traditional two-step ML and one-step DL was discussed in detail, with provided examples for each component. This included high-level descriptions visualized in easy-to-understand figures but also mathematical concepts needed to understand the working of multiple preprocessing steps and ML and DL concepts. Some practical points for a computer scientist working on such a brain signal classification pipeline were also given. This included how to perform correct model evaluation and some of the many potential pitfalls when doing such research.

The remainder of this master thesis was then focused primarily on the evaluation of multiple proposed classification pipelines, which also helps in further understanding how the previously mentioned theoretical and conceptual components translate to an effective pipeline. Taking all of these things into account, the author of this master thesis believes the goal of providing a low-entry document to the BCI field from a computer scientist's point of view is well achieved. Being made by a computer science student that had to go through the many ups and downs in the steep learning curve of the field, it was easier to see the gaps in more intuitive and fundamental background knowledge that was missing from scientifically far superior systematic review articles. therefore calling this first part of the thesis equal to those articles would be unfair to both the articles and this master thesis. It is a rather unique document where initially a more intuitive approach to explaining some of the difficult components was taken but later on the document switched back to its scientific roots to form what can be considered a scoping review for the BCI field.

7.3 Is providing LSTM functionalities to a CNN model beneficial?

The second major goal of this master thesis was to evaluate different types of offline MI EEG classification pipelines, especially CNN based approaches and to propose a useful extension to one of these state-of-the-art models. In practice, this meant two traditional two-step ML approaches using CSP and three state-of-the-art one-step DL models based on a CNN architecture. In particular, the literature proposed DL architectures were EEGNet (as proposed by Lawhern et al., 2016), ShallowConvNet and DeepConvNet (both proposed by Schirrmeister et al., 2017). After a detailed discussion of these architectures, it was discussed how the time series nature of EEG could potentially be an ideal application for LSTM.

Rather than proposing an ANN based on LSTM alone, it was explored if adding this functionality to an existing proven to be successful CNN model can further boost its performance. One extension used a regular LSTM layer after an otherwise almost equal to EEGNet architecture, whilst another replaced the second half of EEGNet with a convolutional LSTM arrangement. initially, the latter was thought to be a more interesting design as it directly integrated into the EEGNet proposed architecture. However, the Keras-provided implementation of this convolutional LSTM layer does not support CUDA optimisation and as such proved to be computationally heavy to train. This made the already computationally hard process of finding a good network architecture and parameters even harder with waiting times in the hours to get any insight on how the changes affect validation performance. Adding to this, a prediction time test showed that the prediction using this convolutional LSTM based model was more than 20 times slower than EEGNet and the regular LSTM extended EEGNet.

This rather practical issue does make it that it is unlikely this network architecture will be studied further, as it got comparable to or negligible better results than the EEGNet model it was based on. Luckily, intersubject testing using a 1.5-second window centred around a known event showed promising results for the regular LSTM extension to EEGNet. A mean test accuracy of 65.52 ± 0.89, over the 60.68 ± 1.64 obtained by the EEGNet model was found. Whilst the standard deviations don't overlap, it is hard to call this definitive proof that adding LSTM functionality boosts the performance of an EEGNet. As such, the research question remains an open one, with data that is suggestive of a positive answer.

7.4 Future work

Multiple future extensions on this master thesis are possible, some of which have already been discussed throughout this master thesis. Perhaps the most suitable future work would be to use the proposed regular bidirectional LSTM extended EEGNet model in an equal test environment but using overlapping sliding windows. Not being centred around the event, and thus not having the ERS and/or ERD following the MI task at a somewhat fixed time, it is expected that EEGNet would perform worse compared to the results noted here. In this regard, the added LSTM layer is expected to work better with processing the time series as it has been built around the idea of being used in sequential data where some time points might be more important than others. This makes it rather likely that a more significant difference will be noticed and that a formal answer to the research question of this master thesis can be formed.

The sliding window approach would also make it more suitable for use in an online BCI system. As such, the logical next step after a positive finding would be to train the model on all available data of the CLA dataset and perhaps even some of the other MI datasets provided by Kaya et al. (2018) that contain left-hand, right-hand and passive MI tasks. A pilot study already showed that this can increase intersubject performance significantly. The model could then be compared to existing LSTM and CNN-LSTM models proposed in the literature as well as the EEGNet and DeepConvNet model from this paper. The ShallowConvNet model may be considered as well, although the author of this paper believes EEGNet to be a more attractive variant of a DL model that is inspired on FBCSP. A different dataset containing more MI classes may also be considered, although this will likely degrade the accuracy and risky errors, potentially making it unsuitable for online BCI use. Testing different thresholds for classifying the samples as being of passive MI is also an option that is worthwhile to look into. As discussed, having a low accuracy but better risky misclassification rate is desired for BCI usage and is possible to obtain by changing this threshold.

For those interested in the visualisation of these models, and given the discussed law proposals surrounding explainable AI perhaps everyone should be interested, looking into the discussed potentials of visualising the model is also a worthwhile route. This should then be combined by subsampling the channels as proposed in this master thesis to only include those channels that have valuable information surrounding the MI tasks. Besides being more compliant with explainable AI principles, the discussed medical applications of BCI surrounding neuroplasticity and rehabilitation can directly use such visualisation in their day-to-day life. There are many other potential future works, but if there is one thing this master thesis aimed to do it is to illustrate how fascinating and full of possibilities the BCI field is and as such, the interested reader most likely has obtained multiple ideas themself from simply reading this master thesis.

7.5 Final remarks and a personal endnote

Having looked back at this master thesis that passed the three-digit page numeration, this final section addresses some self-reflection on this work. Three major things became apparent whilst working on this master thesis. First, to say it with a nice quote, *the more you know, the more you know you don't know*. The steep learning curve of the BCI field combined with a high interest in the field resulted in four papers being added to the to-read list whilst only one was removed. This resulted in an over-the-top literature review, with more than 300 consulted sources. Whilst the page count of this master thesis was originally even higher, it should have probably been more compact but an unlikely combination of events caused a significant shortage in time to downsize this work qualitatively and it was chosen to let Chapter 1 as is. Second, whilst a stressful process at times, it made me discover a passion for performing research I didn't know I had. If it were not for time constraints, I would've happily worked on this master thesis for another few months. Third, I want to end this master thesis similarly to how it started. I want to end it by thanking my supervisor Professor Doctor Geraint Wiggins and my advisor Arnau Dillen for not only allowing me to perform this research but for being understanding of some of the unfortunate events that occurred along the way.

Extra figures

Model	Test subject	validation accuracy	Test accuracy	Left MI PPV	Right MI PPV	Risky error	Non-default parameter settings
BiConvLSTM EEGNet	B	0.8281 @ epoch 494	72.92	70.51	77.05	19.27	300 time points, 150 kernel length, F1 6, D 2 | LSTM kernel size 16, AVG pooling 4
BiLSTM EEGNet	B	0.8385 @ epoch 350	75.00	81.67	70.00	16.67	300 time points, 150 kernel length, F1 16, F2 32, D4 | LSTM size 128
ShallowConvNet	B	0.7812 @ epoch 1033	75.52	**84.75**	67.57	17.19	300 time points, 25 convolutional filters, strides 15, pool size 75
DeepConvNet	B	0.8438 @ epoch 177	75.52	76.81	75.86	**15.63**	300 time points, dropout 0.5, 12 first layer filters, strides 4, pool size 4
EEGNet	B	0.8802 @ epoch 383	**78.13**	83.33	72.46	**15.63**	300 time points, kernel length 150, F1 16, F2 32, D 4
BiConvLSTM EEGNet	C	0.8854 @ epoch 424	87.50	85.51	87.69	9.38	300 time points, 150 kernel length, F1 6, D 2 | LSTM kernel size 16, AVG pooling 4
BiLSTM EEGNet	C	0.8958 @ epoch 920	88.02	89.71	86.36	8.33	300 time points, 150 kernel length, F1 16, F2 32, D4 | LSTM size 128
EEGNet	C	0.9479 @ epoch 2290	95.31	92.54	96.88	3.65	300 time points, kernel length 150, F1 16, F2 32, D 4
ShallowConvNet	C	0.9427 @ epoch 2167	95.83	**95.38**	96.77	2.60	300 time points, 25 convolutional filters, strides 15, pool size 75
DeepConvNet	C	0.9271 @ epoch 1430	**96.88**	**95.38**	**98.41**	**2.08**	300 time points, dropout 0.5, 12 first layer filters, strides 4, pool size 4
BiLSTM EEGNet	E	0.8958 @ epoch 920	82.72	85.07	87.10	9.42	300 time points, 150 kernel length, F1 16, F2 32, D4 | LSTM size 128
BiConvLSTM EEGNet	E	0.9267 @ epoch 389	86.39	84.29	90.16	8.90	300 time points, 150 kernel length, F1 6, D 2 | LSTM kernel size 16, AVG pooling 4
DeepConvNet	E	0.9686 @ epoch 717	89.01	88.73	87.69	8.38	300 time points, dropout 0.5, 12 first layer filters, strides 4, pool size 4
EEGNet	E	0.9529 @ epoch 896	**93.19**	**93.85**	93.65	**4.19**	300 time points, kernel length 150, F1 16, F2 32, D 4
ShallowConvNet	E	0.9686 @ epoch 973	**93.19**	93.75	**93.75**	**4.19**	300 time points, 25 convolutional filters, strides 15, pool size 75

Figure 7.1: Intrasession results for all classification approaches using the longer window. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches.

Model	Test subject	validation accuracy	Test accuracy	Left MI PPV	Right MI PPV	Risky error	Non-default parameter settings
BiConvLSTM EEGNet	B	0.8281 @ epoch 494	54.48	67.58	50.32	22.40	300 time points, 150 kernel length, F1 6, D 2 | LSTM kernel size 16, AVG pooling 4
BiLSTM EEGNet	B	0.6892 @ epoch 246	65.73	66.67	72.69	19.17	300 time points, 150 kernel length, F1 16, F2 32, D4 | LSTM size 192
ShallowConvNet	B	0.7483 @ epoch 591	63.11	58.85	74.18	22.88	300 time points, 25 convolutional filters, strides 15, pool size 75
EEGNet	B	0.7378 @ epoch 2483	67.40	**72.54**	73.78	**15.73**	300 time points, kernel length 150, F1 16, F2 32, D 4
DeepConvNet	B	0.7569 @ epoch 2496	**67.92**	70.06	**76.19**	16.04	300 time points, dropout 0.5, 12 first layer filters, strides 4, pool size 4
ShallowConvNet	C	0.901 @ epoch 1701	56.00	84.93	44.08	41.71	300 time points, 25 convolutional filters, strides 15, pool size 75
BiConvLSTM EEGNet	C	0.8819 @ epoch 164	61.84	76.52	63.69	17.52	300 time points, 150 kernel length, F1 6, D 2 | LSTM kernel size 16, AVG pooling 4
BiLSTM EEGNet	C	0.875 @ epoch 53	65.48	75.09	64.97	19.71	300 time points, 150 kernel length, F1 16, F2 32, D4 | LSTM size 192
EEGNet	C	0.9149 @ epoch 885	63.92	70.35	53.23	32.43	300 time points, kernel length 150, F1 16, F2 32, D 4
DeepConvNet	C	0.9184 @ epoch 1316	**75.34**	**89.92**	**72.55**	**11.49**	300 time points, dropout 0.5, 12 first layer filters, strides 4, pool size 4
ShallowConvNet	E	0.8038 @ epoch 103	60.31	51.24	94.12	31.52	300 time points, 25 convolutional filters, strides 15, pool size 75
BiLSTM EEGNet	E	0.7934 @ epoch 197	69.32	62.89	92.35	18.85	300 time points, 150 kernel length, F1 16, F2 32, D4 | LSTM size 192
BiConvLSTM EEGNet	E	0.8108 @ epoch 198	69.42	68.72	93.33	14.14	300 time points, 150 kernel length, F1 6, D 2 | LSTM kernel size 16, AVG pooling 4
DeepConvNet	E	0.8003 @ epoch 54	74.03	69.11	93.41	15.29	300 time points, dropout 0.5, 12 first layer filters, strides 4, pool size 4
EEGNet	E	0.8333 @ epoch 233	**78.12**	**74.68**	**94.97**	**11.41**	300 time points, kernel length 150, F1 16, F2 32, D 4

Figure 7.2: Intersession results for all classification approaches using the longer window. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches.

Model	Test subject	validation accuracy	Test accuracy	Left MI PPV	Right MI PPV	Risky error	Non-default parameter settings
ShallowConvNet	B	0.7949 @ epoch 2488	52.71	50.99	51.99	45.83	300 time points, 25 convolutional filters, strides 15, pool size 75
DeepConvNet	B	0.8007 @ epoch 49	60.73	60.63	57.86	35.42	300 time points, dropout 0.5, 12 first layer filters, strides 4, pool size 4
EEGNet	B	0.8001 @ epoch 438	62.19	63.09	56.81	34.17	300 time points, kernel length 150, F1 16, F2 32, D 4
BiConvLSTM EEGNet	B	0.7908 @ epoch 44	62.19	60.35	67.03	**25.73**	300 time points, 150 kernel length, F1 6, D 2 | LSTM kernel size 16, AVG pooling 4
BiLSTM EEGNet	B	0.7955 @ epoch 870	**65.73**	**65.17**	**63.59**	26.88	300 time points, 150 kernel length, F1 16, F2 32, D4 | LSTM size 128
ShallowConvNet	C	0.7827 @ epoch 691	50.47	90.91	40.89	44.84	300 time points, 25 convolutional filters, strides 15, pool size 75
EEGNet	C	0.7578 @ epoch 292	58.39	78.73	50.00	28.36	300 time points, kernel length 150, F1 16, F2 32, D 4
BiConvLSTM EEGNet	C	0.7474 @ epoch 119	59.23	77.08	53.35	24.61	300 time points, 150 kernel length, F1 6, D 2 | LSTM kernel size 16, AVG pooling 4
DeepConvNet	C	0.7561 @ epoch 271	63.50	**82.87**	54.27	23.36	300 time points, dropout 0.5, 12 first layer filters, strides 4, pool size 4
BiLSTM EEGNet	C	0.7735 @ epoch 138	**64.34**	80.19	**62.39**	**18.14**	300 time points, 150 kernel length, F1 16, F2 32, D4 | LSTM size 128
ShallowConvNet	E	0.7222 @ epoch 8	48.80	57.54	46.11	42.72	300 time points, 25 convolutional filters, strides 15, pool size 75
EEGNet	E	0.798 @ epoch 2206	61.47	62.57	68.36	23.14	300 time points, kernel length 150, F1 16, F2 32, D 4
BiLSTM EEGNet	E	0.7332 @ epoch 140	66.49	65.23	**79.74**	17.59	300 time points, 150 kernel length, F1 16, F2 32, D4 | LSTM size 128
BiConvLSTM EEGNet	E	0.7338 @ epoch 60	67.02	**80.70**	69.71	14.45	300 time points, 150 kernel length, F1 6, D 2 | LSTM kernel size 16, AVG pooling 4
DeepConvNet	E	0.8293 @ epoch 1172	**69.01**	73.65	78.21	**14.03**	300 time points, dropout 0.5, 12 first layer filters, strides 4, pool size 4

Figure 7.3: Intersubject results for all classification approaches using the longer window. Results are sorted based on the test subject and the obtained classification result on the test set. Colour codes denote either two-step CSP approaches, literature proposed CNN approaches or master thesis proposed CNN-LSTM approaches.

References

Abdeltawab, A., & Ahmad, A. (2020). Classification of Motor Imagery EEG Signals Using Machine Learning. *2020 IEEE 10th International Conference on System Engineering and Technology (ICSET)*, 196–201. https://doi.org/10.1109/ICSET51301.2020.9265364

Abdulkader, S. N., Atia, A., & Mostafa, M.-S. M. (2015). Brain computer interfacing: Applications and challenges. *Egyptian Informatics Journal*, *16*(2), 213–230. https://doi.org/10.1016/j.eij.2015.06.002

Acharya, J. N., Hani, A., Cheek, J., Thirumala, P., & Tsuchida, T. N. (2016). American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature. *Journal of Clinical Neurophysiology*, *33*(4), 308–311. https://doi.org/10.1097/WNP.0000000000000316

Acharya, J. N., Hani, A. J., Thirumala, P. D., & Tsuchida, T. N. (2016). American Clinical Neurophysiology Society Guideline 3: A Proposal for Standard Montages to Be Used in Clinical EEG. *Journal of Clinical Neurophysiology*, *33*(4), 312–316. https://doi.org/10.1097/WNP.0000000000000317

Adama, S. V., & Bogdan, M. (2021). Stroke Rehabilitation and Parkinson's Disease Tremor Reduction Using BCIs Combined With FES: In I. R. Management Association (Ed.), *Research Anthology on Rehabilitation Practices and Therapy* (pp. 679–697). IGI Global. https://doi.org/10.4018/978-1-7998-3432-8.ch033

Aflalo, T., Kellis, S., Klaes, C., Lee, B., Shi, Y., Pejsa, K., Shanfield, K., Hayes-Jackson, S., Aisen, M., Heck, C., Liu, C., & Andersen, R. A. (2015). Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. *Science*, *348*(6237), 906–910. https://doi.org/10.1126/science.aaa5417

Afrakhteh, S., & Mosavi, M. R. (2020). Applying an efficient evolutionary algorithm for EEG signal feature selection and classification in decision-based systems. In *Energy Efficiency of Medical Devices and Healthcare Applications* (pp. 25–52). Elsevier. https://doi.org/10.1016/B978-0-12-819045-6.00002-9

Agarap, A. F. (2018). Deep Learning using Rectified Linear Units (ReLU) [Publisher: arXiv Version Number: 2]. https://doi.org/10.48550/ARXIV.1803.08375

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*, 1–6. https://doi.org/10.1109/ICEngTechnol.2017.8308186

Alcaide, R., Agarwal, N., Candassamy, J., Cavanagh, S., Lim, M., Meschede-Krasa, B., McIntyre, J., Ruiz-Blondet, M. V., Siebert, B., Stanley, D., Valeriani, D., & Yousefi, A. (2021). *EEG-Based Focus Estimation Using Neurable's Enten Headphones and Analytics Platform* (preprint). Neuroscience. https://doi.org/10.1101/2021.06.21.448991

Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains. *ISRN Neuroscience*, *2014*, 1–7. https://doi.org/10.1155/2014/730218

Alimardani, M., Nishio, S., & Ishiguro, H. (2018). Brain-Computer Interface and Motor Imagery Training: The Role of Visual Feedback and Embodiment. In D. Larrivee (Ed.), *Evolving BCI Therapy - Engaging Brain State Dynamics*. InTech. https://doi.org/10.5772/intechopen.78695

Alotaiby, T. N., Alshebeili, S. A., Alotaibi, F. M., & Alrshoud, S. R. (2017). Epileptic Seizure Prediction Using CSP and LDA for Scalp EEG Signals. *Computational Intelligence and Neuroscience*, *2017*, 1–11. https://doi.org/10.1155/2017/1240323

AL-Quraishi, M., Elamvazuthi, I., Daud, S., Parasuraman, S., & Borboni, A. (2018). EEG-Based Control for Upper and Lower Limb Exoskeletons and Prostheses: A Systematic Review. *Sensors*, *18*(10), 3342. https://doi.org/10.3390/s18103342

Alrumiah, S. S. (2020). A Review on Brain-Computer Interface (BCI) Spellers: P300 Speller. *Bioscience Biotechnology Research Communications*, *13*(3), 1191–1199. https://doi.org/10.21786/bbrc/13.3/31

Amunts, K., & Zilles, K. (2015). Architectonic Mapping of the Human Brain beyond Brodmann. *Neuron*, *88*(6), 1086–1107. https://doi.org/10.1016/j.neuron.2015.12.001

Ancona, M., Harkous, H., & Zhang, Y. (2020). *Deepexplain: Attribution methods for deep learning*. Retrieved May 28, 2022, from https://github.com/marcoancona/DeepExplain

Antoine, J.-P. (2003). Wavelet Transforms and Their Applications. *Physics Today*, *56*(4), 68–68. https://doi.org/10.1063/1.1580056

Asif, U., Tang, J., & Harrer, S. (2018). GraspNet: An Efficient Convolutional Neural Network for Real-time Grasp Detection for Low-powered Devices. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 4875–4882. https://doi.org/10.24963/ijcai.2018/677

Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., Carter, R. E., Yao, X., Rabinstein, A. A., Erickson, B. J., Kapa, S., & Friedman, P. A. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction. *The Lancet*, *394*(10201), 861–867. https://doi.org/10.1016/S0140-6736(19)31721-0

Badrakalimuthu, V. R., Swamiraju, R., & de Waal, H. (2011). EEG in psychiatric practice: To do or not to do? *Advances in Psychiatric Treatment*, *17*(2), 114–121. https://doi.org/10.1192/apt.bp.109.006916

Bakshi, K., Pramanik, R., Manjunatha, M., & Kumar, C. S. (2018). Upper Limb Prosthesis Control: A Hybrid EEG-EMG Scheme for Motion Estimation in Transhumeral Subjects. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024–2027. https://doi.org/10.1109/EMBC.2018.8512678

Baldi, P. (2018). Deep Learning in Biomedical Data Science. *Annual Review of Biomedical Data Science*, *1*(1), 181–205. https://doi.org/10.1146/annurev-biodatasci-080917-013343

Barachant, A., Bon, S., Congedo, M., & Jutten, C. (2010). Common Spatial Pattern revisited by Riemannian geometry. *2010 IEEE International Workshop on Multimedia Signal Processing*, 472–476. https://doi.org/10.1109/MMSP.2010.5662067

Berger, H. (1929). Uber das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*(1), 527–570. https://doi.org/10.1007/BF01797193

Bernal, G. (2021). Developing Galea: An open source tool at the intersection of VR and neuroscience. *MIT Media Lab*. https://medium.com/mit-media-lab/developing-galea-an-open-source-tool-at-the-intersection-of-vr-and-neuroscience-61ef60359b96

Bernal, G., Montgomery, S. M., & Maes, P. (2021). Brain-Computer Interfaces, Open-Source, and Democratizing the Future of Augmented Consciousness. *Frontiers in Computer Science*, *3*, 661300. https://doi.org/10.3389/fcomp.2021.661300

Betts, J. G., Young, K. A., Wise, J. A., Johnson, E., Poe, B., Kruse, D. H., Korol, O., Johnson, J. E., Womble, M., & DeSaix, P. (2022). *Anatomy and Physiology 2e* [OCLC: 1322814001]. Rice University. https://openstax.org/details/books/anatomy-and-physiology-2e

Biasiucci, A., Franceschiello, B., & Murray, M. M. (2019). Electroencephalography. *Current Biology*, *29*(3), R80–R85. https://doi.org/10.1016/j.cub.2018.11.052

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*. https://doi.org/10.3389/fninf.2015.00016

Binder, M. D., Hirokawa, N., & Windhorst, U. (Eds.). (2009). Discrete Fourier Transform (DFT). In *Encyclopedia of Neuroscience* (pp. 970–971). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-29678-2_1532

Binnendijk, A., Marler, T., & Bartels, E. (2020). *Brain-Computer Interfaces: U.S. Military Applications and Implications, An Initial Assessment*. RAND Corporation. https://doi.org/10.7249/RR2996

Blankertz, B., Muller, K.-R., Curio, G., Vaughan, T., Schalk, G., Wolpaw, J., Schlogl, A., Neuper, C., Pfurtscheller, G., Hinterberger, T., Schroder, M., & Birbaumer, N. (2004). The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials. *IEEE Transactions on Biomedical Engineering*, *51*(6), 1044–1051. https://doi.org/10.1109/TBME.2004.826692

Blankertz, B., Muller, K.-R., Krusienski, D., Schalk, G., Wolpaw, J., Schlogl, A., Pfurtscheller, G., Millan, J., Schroder, M., & Birbaumer, N. (2006). The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *14*(2), 153–159. https://doi.org/10.1109/TNSRE.2006.875642

Blaus, B. (2014). Medical gallery of Blausen Medical 2014. *WikiJournal of Medicine*, *1*(2). https://doi.org/10.15347/wjm/2014.010

Bontinck, L. (2021). *Bci master thesis at vub 2021 - 2022*. Retrieved September 28, 2021, from https://github.com/pikawika/bci-master-thesis

Boonyakitanont, P., Lek-uthai, A., Chomtho, K., & Songsiri, J. (2020). A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. *Biomedical Signal Processing and Control*, *57*, 101702. https://doi.org/10.1016/j.bspc.2019.101702

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92*, 144–152. https://doi.org/10.1145/130385.130401

Bracewell, R., & Kahn, P. B. (1966). The Fourier Transform and Its Applications. *American Journal of Physics*, *34*(8), 712–712. https://doi.org/10.1119/1.1973431

Bridle, J. (1989). Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. In D. Touretzky (Ed.), *Advances in neural information processing systems* (Vol. 2). Morgan-Kaufmann. https://proceedings.neurips.cc/paper/1989/file/0336dcbab05b9d5ad24f4333c7658a0e-Paper.pdf

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). Openai gym.

Brodwin, E., & Robbins, R. (2020). As Elon Musk's Neuralink prepares to draw back the curtain, ex-employees describe rushed timelines clashing with science's slow pace. https://www.statnews.com/2020/08/25/elon-musk-neuralink-update-brain-machine-implants/

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020).

Language Models are Few-Shot Learners [Publisher: arXiv Version Number: 4]. https://doi.org/10.48550/ARXIV.2005.14165

Brunner, P., Joshi, S., Briskin, S., Wolpaw, J. R., Bischof, H., & Schalk, G. (2010). Does the 'P300' speller depend on eye gaze? *Journal of Neural Engineering*, *7*(5), 056013. https://doi.org/10.1088/1741-2560/7/5/056013

Buhrmester, V., Münch, D., & Arens, M. (2019). Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey [Publisher: arXiv Version Number: 1]. https://doi.org/10.48550/ARXIV.1911.12116

Burkov, A. (2019). *The hundred-page machine learning book*. Andriy Burkov.

Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., & Vidal, F. (2015). Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *International Journal of Psychophysiology*, *97*(3), 210–220. https://doi.org/10.1016/j.ijpsycho.2015.05.004

Capati., F. A., Bechelli., R. P., & Castro., M. C. F. Hybrid ssvep/p300 bci keyboard - controlled by visual evoked potential. In: In *Proceedings of the 9th international joint conference on biomedical engineering systems and technologies - biosignals, (biostec 2016)*. INSTICC. SciTePress, 2016, 214–218. ISBN: 978-989-758-170-0. https://doi.org/10.5220/0005705202140218

Cardoso, V. F., Delisle-Rodriguez, D., Romero-Laiseca, M. A., Loterio, F. A., Gurve, D., Floriano, A., Krishnan, S., Frizera-Neto, A., & Filho, T. F. B. (2022). BCI based on pedal end-effector triggered through pedaling imagery to promote excitability over the feet motor area. *Research on Biomedical Engineering*. https://doi.org/10.1007/s42600-021-00196-7

Cardoso, V. F., Delisle-Rodriguez, D., Romero-Laiseca, M. A., Loterio, F. A., Gurve, D., Floriano, A., Valadão, C., Silva, L., Krishnan, S., Frizera-Neto, A., & Freire Bastos-Filho, T. (2021). Effect of a Brain–Computer Interface Based on Pedaling Motor Imagery on Cortical Excitability and Connectivity. *Sensors*, *21*(6), 2020. https://doi.org/10.3390/s21062020

Castelvecchi, D. (2022). The race to save the Internet from quantum hackers. *Nature*, *602*(7896), 198–201. https://doi.org/10.1038/d41586-022-00339-5

Center for Security and Emerging Technology, Konaev, M., Chahal, H., Fedasiuk, R., Huang, T., & Rahkovksy, I. (2020). *U.S. Military Investments in Autonomy and AI: A Budgetary Assessment* (tech. rep.). Center for Security and Emerging Technology. https://doi.org/10.51593/20200069

Chen, S. (2022). Mind-reading device to detect porn could speed China's policing of illicit content, say researchers. Retrieved July 13, 2022, from https://www.scmp.com/news/china/science/article/3182087/mind-reading-device-detect-porn-could-speed-chinas-policing

Chen, W.-L., Wagner, J., Heugel, N., Sugar, J., Lee, Y.-W., Conant, L., Malloy, M., Heffernan, J., Quirk, B., Zinos, A., Beardsley, S. A., Prost, R., & Whelan, H. T. (2020). Functional Near-Infrared Spectroscopy and Its Clinical Application in the Field of Neuroscience: Advances and Future Directions. *Frontiers in Neuroscience*, *14*, 724. https://doi.org/10.3389/fnins.2020.00724

Chen, Z., Wang, Y., & Song, Z. (2021). Classification of Motor Imagery Electroencephalography Signals Based on Image Processing Method. *Sensors*, *21*(14), 4646. https://doi.org/10.3390/s21144646

Chervonenkis, A. Y. (2013). Early History of Support Vector Machines. In B. Schölkopf, Z. Luo, & V. Vovk (Eds.), *Empirical Inference* (pp. 13–20). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41136-6_3

Chin-Teng Lin, Shu-Fang Tsai, & Li-Wei Ko. (2013). EEG-Based Learning System for Online Motion Sickness Level Estimation in a Dynamic Vehicle Environment. *IEEE Transac-*

tions on Neural Networks and Learning Systems, *24*(10), 1689–1700. https://doi.org/10.1109/TNNLS.2013.2275003

Chollet, F., et al. (2015). Keras.

Chouhan, T. (2020). *Fbcsp toolbox*. Retrieved August 24, 2022, from https://github.com/fbcsptoolbox/fbcsp_code

Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2015). Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) [Publisher: arXiv Version Number: 5]. https://doi.org/10.48550/ARXIV.1511.07289

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Medicine*, *13*(1), 1. https://doi.org/10.1186/s12916-014-0241-z

Cong Wang, Bin Xia, Jie Li, Wenlu Yang, Dianyun, Alejandro Cardona Velez, & Yang, H. (2011). Motor imagery BCI-based robot arm system. *2011 Seventh International Conference on Natural Computation*, 181–184. https://doi.org/10.1109/ICNC.2011.6021923

Cristianini, N., & Ricci, E. (2008). Support Vector Machines. In M.-Y. Kao (Ed.), *Encyclopedia of Algorithms* (pp. 928–932). Springer US. https://doi.org/10.1007/978-0-387-30162-4_415

Cruz-Garza, J. G., Brantley, J. A., Nakagome, S., Kontson, K., Megjhani, M., Robleto, D., & Contreras-Vidal, J. L. (2017). Deployment of Mobile EEG Technology in an Art Museum Setting: Evaluation of Signal Quality and Usability. *Frontiers in Human Neuroscience*, *11*, 527. https://doi.org/10.3389/fnhum.2017.00527

Cuthbertson, A. (2021). Valve is working on brain-computer interface gaming, president reveals. *Independent*. https://www.independent.co.uk/life-style/gadgets-and-tech/valve-brain-computer-interface-video-game-b1792225.html

Dadia, T., & Greenbaum, D. (2019). Neuralink: The Ethical 'Rithmatic of Reading and Writing to the Brain. *AJOB Neuroscience*, *10*(4), 187–189. https://doi.org/10.1080/21507740.2019.1665129

Daly, J. J., & Wolpaw, J. R. (2008). Brain–computer interfaces in neurological rehabilitation. *The Lancet Neurology*, *7*(11), 1032–1043. https://doi.org/10.1016/S1474-4422(08)70223-0

D'Amour, A., Heller, K., Moldovan, D., Adlam, B., Alipanahi, B., Beutel, A., Chen, C., Deaton, J., Eisenstein, J., Hoffman, M. D., Hormozdiari, F., Houlsby, N., Hou, S., Jerfel, G., Karthikesalingam, A., Lucic, M., Ma, Y., McLean, C., Mincu, D., ... Sculley, D. (2020). Underspecification Presents Challenges for Credibility in Modern Machine Learning [Publisher: arXiv Version Number: 2]. https://doi.org/10.48550/ARXIV.2011.03395

da Silva Souto, C., Lüddemann, H., Lipski, S., Dietz, M., & Kollmeier, B. (2016). Influence of attention on speech-rhythm evoked potentials: First steps towards an auditory brain-computer interface driven by speech. *Biomedical Physics & Engineering Express*, *2*(6), 065009. https://doi.org/10.1088/2057-1976/2/6/065009

David Hairston, W., Whitaker, K. W., Ries, A. J., Vettel, J. M., Cortney Bradford, J., Kerick, S. E., & McDowell, K. (2014). Usability of four commercially-oriented EEG systems. *Journal of Neural Engineering*, *11*(4), 046018. https://doi.org/10.1088/1741-2560/11/4/046018

Davis, K. A., Devries, S. P., Krieger, A., Mihaylova, T., Minecan, D., Litt, B., Wagenaar, J. B., & Stacey, W. C. (2018). The effect of increased intracranial EEG sampling rates in clinical practice. *Clinical Neurophysiology*, *129*(2), 360–367. https://doi.org/10.1016/j.clinph.2017.10.039

De Smet, R. (2021). *Vub latex huisstijl* [GitHub commit: 2de903e5...]. Retrieved August 29, 2022, from https://gitlab.com/rubdos/texlive-vub

De Wulf, W. (2022). *Transfer learning in brain-computer interfaces: Language-pretrained transformers for classifying electroencephalography* [Doctoral dissertation, VUB]. Brussels. Retrieved July 16, 2022, from https://github.com/wulfdewolf/lpt-for-eeg

Deiss, O., Biswal, S., Jin, J., Sun, H., Westover, M. B., & Sun, J. (2018). HAMLET: Interpretable Human And Machine co-LEarning Technique [arXiv: 1803.09702]. *arXiv:1803.09702 [cs, stat]*. Retrieved December 6, 2021, from http://arxiv.org/abs/1803.09702

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Dillen, A. (2018). *EEGPLUS: Making a Hybrid Brain Computer Interface Using EEG and EMG Signals* [Doctoral dissertation, VUB]. Brussels. Retrieved September 28, 2021, from https://researchportal.vub.be/en/studentTheses/eegplus-making-a-hybrid-brain-computer-interface-using-eeg-and-em

Dillen, A., Steckelmacher, D., Efthymiadis, K., Langlois, K., De Beir, A., Marusic, U., Vanderborght, B., Nowé, A., Meeusen, R., Ghaffari, F., Romain, O., & De Pauw, K. (2022). Deep learning for biosignal control: Insights from basic to real-time methods with recommendations. *Journal of Neural Engineering*, *19*(1), 011003. https://doi.org/10.1088/1741-2552/ac4f9a

Drew, L. (2019). The ethics of brain–computer interfaces. *Nature*, *571*(7766), S19–S21. https://doi.org/10.1038/d41586-019-02214-2

European Commission. Joint Research Centre. (2020). *Robustness and explainability of Artificial Intelligence: From technical to policy solutions.* Publications Office. Retrieved March 1, 2022, from https://data.europa.eu/doi/10.2760/57493

Facebook. (2017). F8 2017: AI, Building 8 and More Technology Updates From Day Two. Retrieved April 11, 2022, from https://about.fb.com/news/2017/04/f8-2017-day-2/

Facebook. (2021). BCI milestone: New research from UCSF with support from Facebook shows the potential of brain-computer interfaces for restoring speech communication. https://tech.fb.com/bci-milestone-new-research-from-ucsf-with-support-from-facebook-shows-the-potential-of-brain-computer-interfaces-for-restoring-speech-communication/

Fadzal, C., Mansor, W., & Khuan, L. Y. (2011). Review of brain computer interface application in diagnosing dyslexia. *2011 IEEE Control and System Graduate Research Colloquium*, 124–128. https://doi.org/10.1109/ICSGRC.2011.5991843

Fahimi, F., Dosen, S., Ang, K. K., Mrachacz-Kersting, N., & Guan, C. (2021). Generative Adversarial Networks-Based Data Augmentation for Brain–Computer Interface. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(9), 4039–4051. https://doi.org/10.1109/TNNLS.2020.3016666

Fan, X., Bi, L., & Wang, Z. (2012). Detecting emergency situations by monitoring drivers' states from EEG. *2012 ICME International Conference on Complex Medical Engineering (CME)*, 245–248. https://doi.org/10.1109/ICCME.2012.6275717

Farwell, L., & Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, *70*(6), 510–523. https://doi.org/10.1016/0013-4694(88)90149-6

Fazel-Rezai, R., Allison, B. Z., Guger, C., Sellers, E. W., Kleih, S. C., & Kübler, A. (2012). P300 brain computer interface: Current challenges and emerging trends. *Frontiers in Neuroengineering*, *5*. https://doi.org/10.3389/fneng.2012.00014

Ferree, T., Clay, M., & Tucker, D. (2001). The spatial resolution of scalp EEG. *Neurocomputing*, *38-40*, 1209–1216. https://doi.org/10.1016/S0925-2312(01)00568-9

Fisher, R. A. (1936). THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. *Annals of Eugenics*, *7*(2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Fleischer, C. (2007). Controlling Exoskeletons with EMG signals and a Biomechanical Body Model [Publisher: Technische Universität Berlin]. https://doi.org/10.14279/DEPOSITONCE-1674

Floridi, L. (2020). AI and Its New Winter: From Myths to Realities. *Philosophy & Technology*, *33*(1), 1–3. https://doi.org/10.1007/s13347-020-00396-6

Frey, J. (2016). Comparison of an open-hardware electroencephalography amplifier with medical grade device in brain-computer interface applications [Publisher: arXiv Version Number: 1]. https://doi.org/10.48550/ARXIV.1606.02438

Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, *20*(3-4), 121–136. https://doi.org/10.1007/BF00342633

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. https://doi.org/10.1007/BF00344251

Fulham, M. (2004). Neuroimaging. In *Encyclopedia of Neuroscience* (pp. 459–469). Elsevier. https://doi.org/10.1016/B978-008045046-9.00309-0

Fuller, M. (2019). Big data and the Facebook scandal: Issues and responses. *Theology*, *122*(1), 14–21. https://doi.org/10.1177/0040571X18805908

Garcia-Moreno, F. M., Bermudez-Edo, M., Rodriguez-Fortiz, M. J., & Garrido, J. L. (2020). A CNN-LSTM Deep Learning Classifier for Motor Imagery EEG Detection Using a Low-invasive and Low-Cost BCI Headband. *2020 16th International Conference on Intelligent Environments (IE)*, 84–91. https://doi.org/10.1109/IE49459.2020.9155016

Gaur, P., Gupta, H., Chowdhury, A., McCreadie, K., Pachori, R. B., & Wang, H. (2021). A Sliding Window Common Spatial Pattern for Enhancing Motor Imagery Classification in EEG-BCI. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–9. https://doi.org/10.1109/TIM.2021.3051996

Geethanjali, P., Mohan, Y. K., & Sen, J. (2012). Time domain Feature extraction and classification of EEG data for Brain Computer Interface. *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, 1136–1139. https://doi.org/10.1109/FSKD.2012.6234336

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (Second edition). O'Reilly Media, Inc.

Gilbert, F., Cook, M., O'Brien, T., & Illes, J. (2019). Embodiment and Estrangement: Results from a First-in-Human "Intelligent BCI" Trial. *Science and Engineering Ethics*, *25*(1), 83–96. https://doi.org/10.1007/s11948-017-0001-5

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Gregg, M., Hall, C., & Butler, A. (2010). The MIQ-RS: A Suitable Option for Examining Movement Imagery Ability. *Evidence-Based Complementary and Alternative Medicine*, *7*(2), 249–257. https://doi.org/10.1093/ecam/nem170

Grosse-Wentrup, M., & Buss, M. (2008). Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction. *IEEE Transactions on Biomedical Engineering*, *55*(8), 1991–2000. https://doi.org/10.1109/TBME.2008.921154

Guerrero, M. C., Parada, J. S., & Espitia, H. E. (2021). EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks. *Heliyon*, *7*(6), e07258. https://doi.org/10.1016/j.heliyon.2021.e07258

Gupta, K., Rahimi, A., Ajanthan, T., Mensink, T., Sminchisescu, C., & Hartley, R. (2020). Calibration of Neural Networks using Splines [Publisher: arXiv Version Number: 2]. https://doi.org/10.48550/ARXIV.2006.12800

Guy, V., Soriani, M.-H., Bruno, M., Papadopoulo, T., Desnuelle, C., & Clerc, M. (2018). Brain computer interface with the P300 speller: Usability for disabled people with amyotrophic lateral sclerosis. *Annals of Physical and Rehabilitation Medicine*, *61*(1), 5–11. https://doi.org/10.1016/j.rehab.2017.09.004

Haas, L. F. (2003). Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*(1), 9–9. https://doi.org/10.1136/jnnp.74.1.9

Halder, S., Takano, K., & Kansaku, K. (2018). Comparison of Four Control Methods for a Five-Choice Assistive Technology. *Frontiers in Human Neuroscience*, *12*, 228. https://doi.org/10.3389/fnhum.2018.00228

Halford, J. J., Sabau, D., Drislane, F. W., Tsuchida, T. N., & Sinha, S. R. (2016). American Clinical Neurophysiology Society Guideline 4: Recording Clinical EEG on Digital Media. *Journal of Clinical Neurophysiology*, *33*(4), 317–319. https://doi.org/10.1097/WNP.0000000000000318

He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems*, *212*, 106622. https://doi.org/10.1016/j.knosys.2020.106622

Herath, H. M. K. K. M. B., & de Mel, W. (2021). Controlling an Anatomical Robot Hand Using the Brain-Computer Interface Based on Motor Imagery (A. Piccinno, Ed.). *Advances in Human-Computer Interaction*, *2021*, 1–15. https://doi.org/10.1155/2021/5515759

Hinrichs, H., Scholz, M., Baum, A. K., Kam, J. W. Y., Knight, R. T., & Heinze, H.-J. (2020). Comparison between a wireless dry electrode EEG system with a conventional wired wet electrode EEG system for clinical applications. *Scientific Reports*, *10*(1), 5218. https://doi.org/10/gm3k2p

Hodson, R. (2019). The brain. *Nature*, *571*(7766), S1–S1. https://doi.org/10.1038/d41586-019-02206-2

Holmes, K. (2014). Hans Berger and the E.E.G. *Thresholds*, *42*, 88–97. https://doi.org/10.1162/thld_a_00080

Hou, Y., Jia, S., Lun, X., Zhang, S., Chen, T., Wang, F., & Lv, J. (2020). GCNs-Net: A Graph Convolutional Neural Network Approach for Decoding Time-resolved EEG Motor Imagery Signals [Publisher: arXiv Version Number: 3]. https://doi.org/10.48550/ARXIV.2006.08924

Hou, Y., Jia, S., Lun, X., Zhang, S., Chen, T., Wang, F., & Lv, J. (2022). Deep Feature Mining via the Attention-Based Bidirectional Long Short Term Memory Graph Convolutional Neural Network for Human Motor Imagery Recognition. *Frontiers in Bioengineering and Biotechnology*, *9*, 706229. https://doi.org/10.3389/fbioe.2021.706229

Hou, Y., Zhou, L., Jia, S., & Lun, X. (2020). A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN. *Journal of Neural Engineering*, *17*(1), 016048. https://doi.org/10.1088/1741-2552/ab4af6

Hu, M. (2020). Cambridge Analytica's black box. *Big Data & Society*, *7*(2), 205395172093809. https://doi.org/10.1177/2053951720938091

Hubner, D., Verhoeven, T., Muller, K.-R., Kindermans, P.-J., & Tangermann, M. (2018). Unsupervised Learning for Brain-Computer Interfaces Based on Event-Related Potentials:

Review and Online Comparison [Research Frontier]. *IEEE Computational Intelligence Magazine*, *13*(2), 66–77. https://doi.org/10.1109/MCI.2018.2807039

Humayun, M. S., Dorn, J. D., da Cruz, L., Dagnelie, G., Sahel, J.-A., Stanga, P. E., Cideciyan, A. V., Duncan, J. L., Eliott, D., Filley, E., Ho, A. C., Santos, A., Safran, A. B., Arditi, A., Del Priore, L. V., & Greenberg, R. J. (2012). Interim Results from the International Trial of Second Sight's Visual Prosthesis. *Ophthalmology*, *119*(4), 779–788. https://doi.org/10.1016/j.ophtha.2011.09.028

Hunter Christie, E. (2022). Defence cooperation in artificial intelligence: Bridging the transatlantic gap for a stronger Europe. *European View*, *21*(1), 13–21. https://doi.org/10.1177/17816858221089372

Hussein, D., Ibrahim, D., Alrumiah, S., Alhajjaj, L., & Alshobaili, J. (2020). A review on brain-computer interface (bci) spellers: P300 speller. *Bioscience Biotechnology Research Communications*, *13*, 1191–1199. https://doi.org/10.21786/bbrc/13.3/31

Ifft, P. J., Shokur, S., Li, Z., Lebedev, M. A., & Nicolelis, M. A. L. (2013). A Brain-Machine Interface Enables Bimanual Arm Movements in Monkeys. *Science Translational Medicine*, *5*(210). https://doi.org/10.1126/scitranslmed.3006159

Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [Publisher: arXiv Version Number: 3]. https://doi.org/10.48550/ARXIV.1502.03167

Isaksen, J. L., Baumert, M., Hermans, A. N. L., Maleckar, M., & Linz, D. (2022). Artificial intelligence for the detection, prediction, and management of atrial fibrillation. *Herzschrittmachertherapie + Elektrophysiologie*, *33*(1), 34–41. https://doi.org/10.1007/s00399-022-00839-x

Javed Mehedi Shamrat, F. M., Ranjan, R., Hasib, K. M., Yadav, A., & Siddique, A. H. (2022). Performance Evaluation Among ID3, C4.5, and CART Decision Tree Algorithm [Series Title: Lecture Notes in Networks and Systems]. In G. Ranganathan, R. Bestak, R. Palanisamy, & Á. Rocha (Eds.), *Pervasive Computing and Social Networking* (pp. 127–142, Vol. 317). Springer Singapore. https://doi.org/10.1007/978-981-16-5640-8_11

Jawad, A. J. (2020). Engineering Ethics of Neuralink Brain Computer Interfaces Devices. *Annals of Bioethics & Clinical Applications*, *4*(1). https://doi.org/10.23880/abca-16000160

Jeong, J.-H., Shim, K.-H., Kim, D.-J., & Lee, S.-W. (2020). Brain-Controlled Robotic Arm System Based on Multi-Directional CNN-BiLSTM Network Using EEG Signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(5), 1226–1238. https://doi.org/10.1109/TNSRE.2020.2981659

Jia, S., Hou, Y., Shi, Y., & Li, Y. (2020). Attention-based Graph ResNet for Motor Intent Detection from Raw EEG signals [Publisher: arXiv Version Number: 1]. https://doi.org/10.48550/ARXIV.2007.13484

Jing, H., & Takigawa, M. (2000). Low sampling rate induces high correlation dimension on electroencephalograms from healthy subjects. *Psychiatry and Clinical Neurosciences*, *54*(4), 407–412. https://doi.org/10.1046/j.1440-1819.2000.00729.x

Kai Keng Ang, Zhang Yang Chin, Haihong Zhang, & Cuntai Guan. (2008). Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2390–2397. https://doi.org/10.1109/IJCNN.2008.4634130

Kalika, D., Collins, L., Caves, K., & Throckmorton, C. (2017). Fusion of P300 and eye-tracker data for spelling using BCI2000. *Journal of Neural Engineering*, *14*(5), 056010. https://doi.org/10.1088/1741-2552/aa776b

Kapgate, D. (2022). Efficient Quadcopter Flight Control Using Hybrid SSVEP + P300 Visual Brain Computer Interface. *International Journal of Human–Computer Interaction*, *38*(1), 42–52. https://doi.org/10.1080/10447318.2021.1921482

Kasess, C. H., Windischberger, C., Cunnington, R., Lanzenberger, R., Pezawas, L., & Moser, E. (2008). The suppressive influence of SMA on M1 in motor imagery revealed by fMRI and dynamic causal modeling. *NeuroImage*, *40*(2), 828–837. https://doi.org/10.1016/j.neuroimage.2007.11.040

Käthner, I., Kübler, A., & Halder, S. (2015). Comparison of eye tracking, electrooculography and an auditory brain-computer interface for binary communication: A case study with a participant in the locked-in state. *Journal of NeuroEngineering and Rehabilitation*, *12*(1), 76. https://doi.org/10.1186/s12984-015-0071-z

Kawala-Sterniuk, A., Browarska, N., Al-Bakri, A., Pelc, M., Zygarlicki, J., Sidikova, M., Martinek, R., & Gorzelanczyk, E. J. (2021). Summary of over Fifty Years with Brain-Computer Interfaces—A Review. *Brain Sciences*, *11*(1), 43. https://doi.org/10.3390/brainsci11010043

Kaya, M., Binli, M. K., Ozbay, E., Yanar, H., & Mishchenko, Y. (2018). A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. *Scientific Data*, *5*(1), 180211. https://doi.org/10.1038/sdata.2018.211

Khan, G., Hashmi, M., Awais, M., Khan, N., & Basir, R. (2020). High Performance Multi-class Motor Imagery EEG Classification [Backup Publisher: INSTICC]. *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 4: BIOSIGNALS,* 149–155. https://doi.org/10.5220/0008864501490155

Kindermans, P.-J., Verschore, H., Verstraeten, D., & Schrauwen, B. (2012). A P300 BCI for the masses: Prior information enables instant unsupervised spelling. http://lib.ugent.be/catalog/pug01:3008919

King, W. C., Hinerman, A. S., & White, G. E. (2022). Changes in Marital Status Following Roux-en-Y Gastric Bypass and Sleeve Gastrectomy: A US Multicenter Prospective Cohort Study. *Annals of Surgery Open*, *3*(3), e182. https://doi.org/10.1097/AS9.0000000000000182

Kirar, J. S., & Agrawal, R. K. (2018). Relevant Frequency Band Selection using Sequential Forward Feature Selection for Motor Imagery Brain Computer Interfaces. *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, 52–59. https://doi.org/10.1109/SSCI.2018.8628719

Kirschstein, T., & Köhling, R. (2009). What is the Source of the EEG? *Clinical EEG and Neuroscience*, *40*(3), 146–149. https://doi.org/10.1177/155005940904000305

Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. [Place: Netherlands]. *Electroencephalography and clinical neurophysiology. Supplement*, *52*, 3–6.

Ko, W., Jeon, E., Yoon, J. S., & Suk, H.-I. (2022). Semi-supervised generative and discriminative adversarial learning for motor imagery-based brain–computer interface. *Scientific Reports*, *12*(1), 4587. https://doi.org/10.1038/s41598-022-08490-9

Koles, Z. J., Lazar, M. S., & Zhou, S. Z. (1990). Spatial patterns underlying population differences in the background EEG. *Brain Topography*, *2*(4), 275–284. https://doi.org/10.1007/BF01129656

Koshev, N., Butorina, A., Skidchenko, E., Kuzmichev, A., Ossadtchi, A., Ostras, M., Fedorov, M., & Vetoshko, P. (2021). Evolution of ¡span style="font-variant:small-caps;"¿MEG¡/span¿ : A first ¡span style="font-variant:small-caps;"¿MEG¡/span¿ -feasible fluxgate magne-

tometer. *Human Brain Mapping*, *42*(15), 4844–4856. https://doi.org/10.1002/hbm.25582

Kosmyna, N., Tarpin-Bernard, F., Bonnefond, N., & Rivet, B. (2016). Feasibility of BCI Control in a Realistic Smart Home Environment. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00416

Kotegawa, K., Yasumura, A., & Teramoto, W. (2020). Activity in the prefrontal cortex during motor imagery of precision gait: An fNIRS study. *Experimental Brain Research*, *238*(1), 221–228. https://doi.org/10.1007/s00221-019-05706-9

Kowalik, Z. J., Wrobel, A., & Rydz, A. (1996). Why does the human brain need to be a nonlinear system? *Behavioral and Brain Sciences*, *19*(2), 302–303. https://doi.org/10.1017/S0140525X0004276X

Krigolson, O. E., Williams, C. C., Norton, A., Hassall, C. D., & Colino, F. L. (2017). Choosing MUSE: Validation of a Low-Cost, Portable EEG System for ERP Research. *Frontiers in Neuroscience*, *11*. https://doi.org/10.3389/fnins.2017.00109

Kübler, A. (2020). The history of BCI: From a vision for the future to real support for personhood in people with locked-in syndrome. *Neuroethics*, *13*(2), 163–180. https://doi.org/10.1007/s12152-019-09409-4

Kuhner, D., Fiederer, L., Aldinger, J., Burget, F., Völker, M., Schirrmeister, R., Do, C., Boedecker, J., Nebel, B., Ball, T., & Burgard, W. (2019). A service assistant combining autonomous robotics, flexible goal formulation, and deep-learning-based brain–computer interfacing. *Robotics and Autonomous Systems*, *116*, 98–113. https://doi.org/10.1016/j.robot.2019.02.015

Kuratani, J., Pearl, P. L., Sullivan, L., Riel-Romero, R. M. S., Cheek, J., Stecker, M., San-Juan, D., Selioutski, O., Sinha, S. R., Drislane, F. W., & Tsuchida, T. N. (2016). American Clinical Neurophysiology Society Guideline 5: Minimum Technical Standards for Pediatric Electroencephalography. *Journal of Clinical Neurophysiology*, *33*(4), 320–323. https://doi.org/10.1097/WNP.0000000000000321

Kwon, M., Han, S., Kim, K., & Jun, S. C. (2019). Super-Resolution for Improving EEG Spatial Resolution using Deep Convolutional Neural Network—Feasibility Study. *Sensors*, *19*(23), 5317. https://doi.org/10.3390/s19235317

LaRocco, J., Le, M. D., & Paeng, D.-G. (2020). A Systemic Review of Available Low-Cost EEG Headsets Used for Drowsiness Detection. *Frontiers in Neuroinformatics*, *14*, 553352. https://doi.org/10.3389/fninf.2020.553352

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2016). EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces [Publisher: arXiv Version Number: 4]. https://doi.org/10.48550/ARXIV.1611.08024

Lebedev, M. A. (2014). How to read neuron-dropping curves? *Frontiers in Systems Neuroscience*, *8*. https://doi.org/10.3389/fnsys.2014.00102

Lee, D. (2020). Comparison of Reinforcement Learning Activation Functions to Improve the Performance of the Racing Game Learning Agent. *Journal of Information Processing Systems*, *16*(5), 1074–1082. https://doi.org/10.3745/JIPS.02.0141

Lee, S. (2020). A Study on Consent of the GDPR in Advertising Technology Focusing on Programmatic Buying. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3616651

Lee, W. T., Nisar, H., Malik, A. S., & Yeap, K. H. (2013). A brain computer interface for smart home control. *2013 IEEE International Symposium on Consumer Electronics (ISCE)*, 35–36. https://doi.org/10.1109/ISCE.2013.6570240

Lemm, S., Blankertz, B., Curio, G., & Muller, K.-R. (2005). Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Transactions on Biomedical Engineering*, *52*(9), 1541–1548. https://doi.org/10.1109/TBME.2005.851521

Lévesque, L. (2014). Nyquist sampling theorem: Understanding the illusion of a spinning wheel captured with a video camera. *Physics Education*, *49*(6), 697–705. https://doi.org/10.1088/0031-9120/49/6/697

Li, T., Zhu, S., & Ogihara, M. (2006). Using discriminant analysis for multi-class classification: An experimental investigation. *Knowledge and Information Systems*, *10*(4), 453–472. https://doi.org/10.1007/s10115-006-0013-y

Li Deng. (2012). The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Processing Magazine*, *29*(6), 141–142. https://doi.org/10.1109/MSP.2012.2211477

Liang, E., Liaw, R., Moritz, P., Nishihara, R., Fox, R., Goldberg, K., Gonzalez, J. E., Jordan, M. I., & Stoica, I. (2017). Rllib: Abstractions for distributed reinforcement learning. https://doi.org/10.48550/ARXIV.1712.09381

Lin, C.-T., Chen, Y.-C., Huang, T.-Y., Chiu, T.-T., Ko, L.-W., Liang, S.-F., Hsieh, H.-Y., Hsu, S.-H., & Duann, J.-R. (2008). Development of Wireless Brain Computer Interface With Embedded Multitask Scheduling and its Application on Real-Time Driver's Drowsiness Detection and Warning. *IEEE Transactions on Biomedical Engineering*, *55*(5), 1582–1591. https://doi.org/10.1109/TBME.2008.918566

Linden, D. E. J. (2005). The P300: Where in the Brain Is It Produced and What Does It Tell Us? *The Neuroscientist*, *11*(6), 563–576. https://doi.org/10.1177/1073858405280524

Liu, X., Makeyev, O., & Besio, W. (2020). Improved Spatial Resolution of Electroencephalogram Using Tripolar Concentric Ring Electrode Sensors. *Journal of Sensors*, *2020*, 1–9. https://doi.org/10.1155/2020/6269394

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update. *Journal of Neural Engineering*, *15*(3), 031005. https://doi.org/10.1088/1741-2552/aab2f2

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *4*(2), R1–R13. https://doi.org/10.1088/1741-2560/4/2/R01

Lu, G., Zhang, W., & Wang, Z. (2022). Optimizing Depthwise Separable Convolution Operations on GPUs. *IEEE Transactions on Parallel and Distributed Systems*, *33*(1), 70–87. https://doi.org/10.1109/TPDS.2021.3084813

Malin, B. A., Emam, K. E., & O'Keefe, C. M. (2013). Biomedical data privacy: Problems, perspectives, and recent advances. *Journal of the American Medical Informatics Association*, *20*(1), 2–6. https://doi.org/10.1136/amiajnl-2012-001509

Mallat, S. G. (2009). *A wavelet tour of signal processing: The sparse way* (3rd ed). Elsevier/Academic Press.

Malouin, F., Richards, C. L., Jackson, P. L., Lafleur, M. F., Durand, A., & Doyon, J. (2007). The Kinesthetic and Visual Imagery Questionnaire (KVIQ) for Assessing Motor Imagery in Persons with Physical Disabilities: A Reliability and Construct Validity Study. *Journal of Neurologic Physical Therapy*, *31*(1), 20–29. https://doi.org/10.1097/01.NPT.0000260567.24122.64

Mane, R., Chew, E., Chua, K., Ang, K. K., Robinson, N., Vinod, A. P., Lee, S.-W., & Guan, C. (2021). FBCNet: A Multi-view Convolutional Neural Network for Brain-Computer Interface [arXiv: 2104.01233]. *arXiv:2104.01233 [cs, eess]*. Retrieved September 28, 2021, from http://arxiv.org/abs/2104.01233

Markand, O. N. (1976). Electroencephalogram in "locked-in" syndrome. *Electroencephalography and Clinical Neurophysiology*, *40*(5), 529–534. https://doi.org/10.1016/0013-4694(76)90083-3

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Jia, Y., Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, ... Xiaoqiang Zheng. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems [Software available from tensorflow.org]. https://www.tensorflow.org/

Martínez Beltrán, E. T., Quiles Pérez, M., López Bernal, S., Huertas Celdrán, A., & Martínez Pérez, G. (2022). Noise-based cyberattacks generating fake P300 waves in brain–computer interfaces. *Cluster Computing*, *25*(1), 33–48. https://doi.org/10.1007/s10586-021-03326-z

Martini, M. L., Oermann, E. K., Opie, N. L., Panov, F., Oxley, T., & Yaeger, K. (2020). Sensor Modalities for Brain-Computer Interface Technology: A Comprehensive Literature Review. *Neurosurgery*, *86*(2), E108–E117. https://doi.org/10.1093/neuros/nyz286

Mathewson, K. E., Harrison, T. J. L., & Kizuk, S. A. D. (2017). High and dry? Comparing active dry EEG electrodes to active and passive wet electrodes: Active dry vs. active & passive wet EEG electrodes. *Psychophysiology*, *54*(1), 74–82. https://doi.org/10.1111/psyp.12536

McAvinue, L. P., & Robertson, I. H. (2008). Measuring motor imagery ability: A review. *European Journal of Cognitive Psychology*, *20*(2), 232–251. https://doi.org/10.1080/09541440701394624

McCrimmon, C. M., Fu, J. L., Wang, M., Lopes, L. S., Wang, P. T., Karimi-Bidhendi, A., Liu, C. Y., Heydari, P., Nenadic, Z., & Do, A. H. (2017). Performance Assessment of a Custom, Portable, and Low-Cost Brain–Computer Interface Platform. *IEEE Transactions on Biomedical Engineering*, *64*(10), 2313–2320. https://doi.org/10.1109/TBME.2017.2667579

McDermott, E. J., Raggam, P., Kirsch, S., Belardinelli, P., Ziemann, U., & Zrenner, C. (2021). Artifacts in EEG-Based BCI Therapies: Friend or Foe? *Sensors*, *22*(1), 96. https://doi.org/10.3390/s22010096

McFarland, D. J., Daly, J., Boulay, C., & Parvaz, M. A. (2017). Therapeutic applications of BCI technologies. *Brain-Computer Interfaces*, *4*(1-2), 37–52. https://doi.org/10.1080/2326263X.2017.1307625

Michel, C. M., & Brunet, D. (2019). EEG Source Imaging: A Practical Review of the Analysis Steps. *Frontiers in Neurology*, *10*, 325. https://doi.org/10.3389/fneur.2019.00325

Miller, K. J., Schalk, G., Fetz, E. E., den Nijs, M., Ojemann, J. G., & Rao, R. P. N. (2010). Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proceedings of the National Academy of Sciences*, *107*(9), 4430–4435. https://doi.org/10.1073/pnas.0913697107

Mishra, P. K., Jagadish, B., Kiran, M. P. R. S., Rajalakshmi, P., & Reddy, D. S. (2018). A Novel Classification for EEG Based Four Class Motor Imagery Using Kullback-Leibler Regularized Riemannian Manifold. *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 1–5. https://doi.org/10.1109/HealthCom.2018.8531086

Montoya-Martínez, J., Vanthornhout, J., Bertrand, A., & Francart, T. (2021). Effect of number and placement of EEG electrodes on measurement of neural tracking of speech (W. Mumtaz, Ed.). *PLOS ONE*, *16*(2), e0246769. https://doi.org/10.1371/journal.pone.0246769

Morris, H. H., Hem, G., & Gilmore-Pollak, W. (1992). Hair loss after prolonged EEG/video monitoring. *Neurology*, *42*(7), 1401–1401. https://doi.org/10.1212/WNL.42.7.1401

Moses, D. A., Metzger, S. L., Liu, J. R., Anumanchipalli, G. K., Makin, J. G., Sun, P. F., Chartier, J., Dougherty, M. E., Liu, P. M., Abrams, G. M., Tu-Chan, A., Ganguly, K., & Chang, E. F. (2021). Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria. *New England Journal of Medicine*, *385*(3), 217–227. https://doi.org/10.1056/NEJMoa2027540

Mulder, T. (2007). Motor imagery and action observation: Cognitive tools for rehabilitation. *Journal of Neural Transmission*, *114*(10), 1265–1278. https://doi.org/10.1007/s00702-007-0763-z

Musk, E., & Neuralink. (2019). *An integrated brain-machine interface platform with thousands of channels* (preprint). Neuroscience. https://doi.org/10.1101/703801

Mussi, M., Spindola, M., & Marques, P. (2019). Classification of EEG Motor Imagery signal processed in time-frequency with Convolutional Neural Network.

Nam, C. S., Nijholt, A., & Lotte, F. (Eds.). (2018). *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances* (1st ed.). CRC Press. https://doi.org/10.1201/9781351231954

Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain Computer Interfaces, a Review. *Sensors*, *12*(2), 1211–1279. https://doi.org/10.3390/s120201211

Nijboer, F., van de Laar, B., Gerritsen, S., Nijholt, A., & Poel, M. (2015). Usability of Three Electroencephalogram Headsets for Brain–Computer Interfaces: A Within Subject Comparison. *Interacting with Computers*, *27*(5), 500–511. https://doi.org/10.1093/iwc/iwv023

Niketeghad, S., & Pouratian, N. (2019). Brain Machine Interfaces for Vision Restoration: The Current State of Cortical Visual Prosthetics. *Neurotherapeutics*, *16*(1), 134–143. https://doi.org/10.1007/s13311-018-0660-1

Nunez, P. L., & Cutillo, B. A. (Eds.). (1995). *Neocortical dynamics and human EEG rhythms*. Oxford University Press.

Olivas-Padilla, B. E., & Chacon-Murguia, M. I. (2019). Classification of multiple motor imagery using deep convolutional neural networks and spatial filters. *Applied Soft Computing*, *75*, 461–472. https://doi.org/10.1016/j.asoc.2018.11.031

Pal, R. (2017). Comparison of the design of FIR and IIR filters for a given specification and removal of phase distortion from IIR filters. *2017 International Conference on Advances in Computing, Communication and Control (ICAC3)*, 1–3. https://doi.org/10.1109/ICAC3.2017.8318772

Palmini, A. (2006). The concept of the epileptogenic zone: A modern look at Penfield and Jasper's views on the role of interictal spikes. [Place: France]. *Epileptic disorders : international epilepsy journal with videotape*, *8 Suppl 2*, S10–15.

Panoulas, K. J., Hadjileontiadis, L. J., & Panas, S. M. (2010). Brain-Computer Interface (BCI): Types, Processing Perspectives and Applications [Series Title: Smart Innovation, Systems and Technologies]. In R. J. Howlett, L. C. Jain, G. A. Tsihrintzis, & L. C. Jain (Eds.), *Multimedia Services in Intelligent Environments* (pp. 299–321, Vol. 3). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13396-1_14

Pasqualotto, E., Matuz, T., Federici, S., Ruf, C. A., Bartl, M., Olivetti Belardinelli, M., Birbaumer, N., & Halder, S. (2015). Usability and Workload of Access Technology for People With Severe Motor Impairment: A Comparison of Brain-Computer Interfacing and Eye Tracking. *Neurorehabilitation and Neural Repair*, *29*(10), 950–957. https://doi.org/10.1177/1545968315575611

Pathirana, S., Asirvatham, D., & Johar, G. (2018). A Critical Evaluation on Low-Cost Consumer-Grade Electroencephalographic Devices. *2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, 160–165. https://doi.org/10.1109/ICBAPS.2018.8527413

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2018). Scikit-learn: Machine Learning in Python [arXiv: 1201.0490]. *arXiv:1201.0490 [cs]*. Retrieved March 14, 2022, from http://arxiv.org/abs/1201.0490

Peterson, V., Galván, C., Hernández, H., & Spies, R. (2020). A feasibility study of a complete low-cost consumer-grade brain-computer interface system. *Heliyon*, *6*(3), e03425. https://doi.org/10.1016/j.heliyon.2020.e03425

Pfurtscheller, G., Neuper, C., Flotzinger, D., & Pregenzer, M. (1997). EEG-based discrimination between imagination of right and left hand movement. *Electroencephalography and Clinical Neurophysiology*, *103*(6), 642–651. https://doi.org/10.1016/S0013-4694(97)00080-1

Picton, T. W. (1988). The Endogenous Evoked Potentials [Series Title: Springer Series in Brain Dynamics]. In E. Başar, W. J. Freeman, W.-D. Heiss, D. Lehmann, F. H. Lopes da Silva, D. Speckmann, & E. Başar (Eds.), *Dynamics of Sensory and Cognitive Processing by the Brain* (pp. 258–265, Vol. 1). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-71531-0_17

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1464*(1), 5–29. https://doi.org/10.1111/nyas.13948

Podder, P., Zaman Khan, T., Haque Khan, M., & Muktadir Rahman, M. (2014). Comparative Performance Analysis of Hamming, Hanning and Blackman Window. *International Journal of Computer Applications*, *96*(18), 1–7. https://doi.org/10.5120/16891-6927

Poorvitha, H. R., Aishwarya, G., Kavya, K. B., & Nayana, R. (2020). Brain Computer Interface for Sleep Apnea Detection. *IJERT*, *8*(15). https://www.ijert.org/brain-computer-interface-for-sleep-apnea-detection

Poulos, M., Felekis, T., & Evangelou, A. (2012). Is it possible to extract a fingerprint for early breast cancer via EEG analysis? *Medical Hypotheses*, *78*(6), 711–716. https://doi.org/10.1016/j.mehy.2012.02.016

Priyanka, A. A., Bharti, W. G., & Suresh, C. M. (2016). *Introduction to EEG- and Speech-Based Emotion Recognition*. Elsevier. https://doi.org/10.1016/C2015-0-01959-1

Ptito, M., Bleau, M., Djerourou, I., Paré, S., Schneider, F. C., & Chebat, D.-R. (2021). Brain-Machine Interfaces to Assist the Blind. *Frontiers in Human Neuroscience*, *15*, 638887. https://doi.org/10.3389/fnhum.2021.638887

Qin, Z., Yu, F., Liu, C., & Chen, X. (2018). How convolutional neural network see the world - A survey of convolutional neural network visualization methods [Publisher: arXiv Version Number: 2]. https://doi.org/10.48550/ARXIV.1804.11191

Rahman, M. A., Khanam, F., Kazem, M., Khurshed, M., & Ahmad, M. (2019). Four-Class Motor Imagery EEG Signal Classification using PCA, Wavelet and Two-Stage Neural Network. *International Journal of Advanced Computer Science and Applications*, *10*(5). https://doi.org/10.14569/IJACSA.2019.0100562

Ramadan, R. A., & Vasilakos, A. V. (2017). Brain computer interface: Control signals review. *Neurocomputing*, *223*, 26–44. https://doi.org/10.1016/j.neucom.2016.10.024

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). Zero-Shot Text-to-Image Generation [Publisher: arXiv Version Number: 2]. https://doi.org/10.48550/ARXIV.2102.12092

Rao, P. (2020). Funding for Brain-Computer Interface Ventures. Retrieved February 23, 2022, from https://www.from-the-interface.com/BCI-venture-funding/

Rasheed, S. (2021). A Review of the Role of Machine Learning Techniques towards Brain–Computer Interface Applications. *Machine Learning and Knowledge Extraction*, *3*(4), 835–862. https://doi.org/10.3390/make3040042

Rashid, U., Niazi, I., Signal, N., & Taylor, D. (2018). An EEG Experimental Study Evaluating the Performance of Texas Instruments ADS1299. *Sensors*, *18*(11), 3721. https://doi.org/10.3390/s18113721

Ratti, E., Waninger, S., Berka, C., Ruffini, G., & Verma, A. (2017). Comparison of Medical and Consumer Wireless EEG Systems for Use in Clinical Trials. *Frontiers in Human Neuroscience*, *11*, 398. https://doi.org/10.3389/fnhum.2017.00398

Renton, A. I., Mattingley, J. B., & Painter, D. R. (2019). Optimising non-invasive brain-computer interface systems for free communication between naïve human participants. *Scientific Reports*, *9*(1), 18705. https://doi.org/10.1038/s41598-019-55166-y

Rezeika, A., Benda, M., Stawicki, P., Gembler, F., Saboor, A., & Volosyak, I. (2018). Brain–Computer Interface Spellers: A Review. *Brain Sciences*, *8*(4), 57. https://doi.org/10.3390/brainsci8040057

Rezende, I. N. (2020). Facial recognition in police hands: Assessing the 'Clearview case' from a European perspective. *New Journal of European Criminal Law*, *11*(3), 375–389. https://doi.org/10.1177/2032284420948161

Riccio, A., Mattia, D., Simione, L., Olivetti, M., & Cincotti, F. (2012). Eye-gaze independent EEG-based brain–computer interfaces for communication. *Journal of Neural Engineering*, *9*(4), 045001. https://doi.org/10.1088/1741-2560/9/4/045001

Robinson, A. J., Jain, A., Sherman, H. G., Hague, R. J. M., Rahman, R., Sanjuan-Alberte, P., & Rawson, F. J. (2021). Toward Hijacking Bioelectricity in Cancer to Develop New Bioelectronic Medicine. *Advanced Therapeutics*, *4*(3), 2000248. https://doi.org/10.1002/adtp.202000248

Rosenow, F., Klein, K. M., & Hamer, H. M. (2015). Non-invasive EEG evaluation in epilepsy diagnosis. *Expert Review of Neurotherapeutics*, *15*(4), 425–444. https://doi.org/10.1586/14737175.2015.1025382

Saha, S., Mamun, K. A., Ahmed, K., Mostafa, R., Naik, G. R., Darvishi, S., Khandoker, A. H., & Baumert, M. (2021). Progress in Brain Computer Interface: Challenges and Opportunities. *Frontiers in Systems Neuroscience*, *15*, 578875. https://doi.org/10.3389/fnsys.2021.578875

Sahaya, J., Haquea, A., Ananda, A., Gautama, A., & Venkatesana, N. (2020). A Deep Learning Approach for Robotic Arm Control using Brain-Computer Interface. *International Journal of Biology and Biomedical Engineering*, *14*. https://doi.org/10.46300/91011.2020.14.18

Sanchez-Panchuelo, R. M., Besle, J., Beckett, A., Bowtell, R., Schluppeck, D., & Francis, S. (2012). Within-Digit Functional Parcellation of Brodmann Areas of the Human Primary Somatosensory Cortex Using Functional Magnetic Resonance Imaging at 7 Tesla. *Journal of Neuroscience*, *32*(45), 15815–15822. https://doi.org/10.1523/JNEUROSCI.2501-12.2012

Sazgar, M., & Young, M. G. (2019). EEG Artifacts. In *Absolute Epilepsy and EEG Rotation Review* (pp. 149–162). Springer International Publishing. https://doi.org/10.1007/978-3-030-03511-2_8

Schaller, R. (1997). Moore's law: Past, present and future. *IEEE Spectrum*, *34*(6), 52–59. https://doi.org/10.1109/6.591665

Schirrmeister, R. T., Gramfort, A., & Lawhern, V. J. (2022). *Army research laboratory (arl) eegmodels*. Retrieved May 10, 2022, from https://github.com/vlawhern/arl-eegmodels

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization: Convolutional Neural Networks in EEG Analysis. *Human Brain Mapping*, *38*(11), 5391–5420. https://doi.org/10.1002/hbm.23730

Schnitzler, A., Salenius, S., Salmelin, R., Jousmäki, V., & Hari, R. (1997). Involvement of Primary Motor Cortex in Motor Imagery: A Neuromagnetic Study. *NeuroImage*, *6*(3), 201–208. https://doi.org/10.1006/nimg.1997.0286

Schuster, C., Hilfiker, R., Amft, O., Scheidhauer, A., Andrews, B., Butler, J., Kischka, U., & Ettlin, T. (2011). Best practice for motor imagery: A systematic literature review on motor imagery training elements in five different disciplines. *BMC Medicine*, *9*(1), 75. https://doi.org/10.1186/1741-7015-9-75

Schwarz, A., Brandstetter, J., Pereira, J., & Müller-Putz, G. R. (2019). Direct comparison of supervised and semi-supervised retraining approaches for co-adaptive BCIs. *Medical & Biological Engineering & Computing*, *57*(11), 2347–2357. https://doi.org/10.1007/s11517-019-02047-1

Selvam, V. S., & Shenbagadevi, S. (2011). Brain tumor detection using scalp eeg with modified Wavelet-ICA and multi layer feed forward neural network. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 6104–6109. https://doi.org/10.1109/IEMBS.2011.6091508

Semmlow, J. (2018). The Big Picture. In *Circuits, Signals and Systems for Bioengineers* (pp. 3–50). Elsevier. https://doi.org/10.1016/B978-0-12-809395-5.00001-1

Seol, H. Y., Park, S., Ji, Y. S., Hong, S. H., & Moon, I. J. (2020). Impact of hearing aid noise reduction algorithms on the speech-evoked auditory brainstem response. *Scientific Reports*, *10*(1), 10773. https://doi.org/10.1038/s41598-020-66970-2

Shang-Lin Wu, Chun-Wei Wu, Pal, N. R., Chih-Yu Chen, Shi-An Chen, & Chin-Teng Lin. (2013). Common spatial pattern and linear discriminant analysis for motor imagery classification. *2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, 146–151. https://doi.org/10.1109/CCMB.2013.6609178

Sheu, Y.-h. (2020). Illuminating the Black Box: Interpreting Deep Neural Network Models for Psychiatric Research. *Frontiers in Psychiatry*, *11*, 551299. https://doi.org/10.3389/fpsyt.2020.551299

Shih, J. J., Krusienski, D. J., & Wolpaw, J. R. (2012). Brain-Computer Interfaces in Medicine. *Mayo Clinic Proceedings*, *87*(3), 268–279. https://doi.org/10.1016/j.mayocp.2011.12.008

Shim, K.-H., Jeong, J.-H., Kwon, B.-H., Lee, B.-H., & Lee, S.-W. (2019). Assistive Robotic Arm Control based on Brain-Machine Interface with Vision Guidance using Convolution Neural Network. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2785–2790. https://doi.org/10.1109/SMC.2019.8914058

Shyu, K.-K., Lee, P.-L., Lee, M.-H., Lin, M.-H., Lai, R.-J., & Chiu, Y.-J. (2010). Development of a Low-Cost FPGA-Based SSVEP BCI Multimedia Control System. *IEEE Transactions on Biomedical Circuits and Systems*, *4*(2), 125–132. https://doi.org/10.1109/TBCAS.2010.2042595

Silipo, R., Deco, G., & Bartsch, H. (2013). Detection of Primary Brain Tumor Present in EEG signal using Wavelet Transform and Neural Network.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961

Sinha, S. R., Sullivan, L., Sabau, D., San-Juan, D., Dombrowski, K. E., Halford, J. J., Hani, A. J., Drislane, F. W., & Stecker, M. M. (2016). American Clinical Neurophysiology Society Guideline 1: Minimum Technical Requirements for Performing Clinical Electroencephalography. *Journal of Clinical Neurophysiology*, *33*(4), 303–307. https://doi.org/10.1097/WNP.0000000000000308

Sonam, Y. S. (2018). A Review Paper on Brain Computer Interface. *IJERT*, *3*(10). https://www.ijert.org/a-review-paper-on-brain-computer-interface

Sonoda, S., & Murata, N. (2017). Neural network with unbounded activation functions is universal approximator. *Applied and Computational Harmonic Analysis*, *43*(2), 233–268. https://doi.org/10.1016/j.acha.2015.12.005

Srinivasan, R. (1999). Methods to improve the spatial resolution of EEG. *International journal of bioelectromagnetism*, *1*(1), 102–111.

Srinivasan, R., Nunez, P. L., Tucker, D. M., Silberstein, R. B., & Cadusch, P. J. (1996). Spatial sampling and filtering of EEG with spline Laplacians to estimate cortical potentials. *Brain Topography*, *8*(4), 355–366. https://doi.org/10.1007/BF01186911

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, *15*(1), 1929–1958.

Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM – a tutorial into Long Short-Term Memory Recurrent Neural Networks [Publisher: arXiv Version Number: 1]. https://doi.org/10.48550/ARXIV.1909.09586

Stecker, M. M., Sabau, D., Sullivan, L., Das, R. R., Selioutski, O., Drislane, F. W., Tsuchida, T. N., & Tatum, W. O. (2016). American Clinical Neurophysiology Society Guideline 6: Minimum Technical Standards for EEG Recording in Suspected Cerebral Death. *Journal of Clinical Neurophysiology*, *33*(4), 324–327. https://doi.org/10.1097/WNP.0000000000000322

Steyrl, D., Kobler, R. J., & Müller-Putz, G. R. (2016). On Similarities and Differences of Invasive and Non-Invasive Electrical Brain Signals in Brain-Computer Interfacing. *Journal of Biomedical Science and Engineering*, *09*(08), 393–398. https://doi.org/10.4236/jbise.2016.98034

Stockman, C. (2020). Can a Technology Teach Meditation? Experiencing the EEG Headband InteraXon Muse as a Meditation Guide. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(08), 83. https://doi.org/10.3991/ijet.v15i08.12415

Strickland, E., & Harris, M. (2022). Their bionic eyes are now obsolete and unsupported. https://spectrum.ieee.org/bionic-eye-obsolete

Sudarsan, S., & Sekaran, E. C. (2012). Design and Development of EMG Controlled Prosthetics Limb. *Procedia Engineering*, *38*, 3547–3551. https://doi.org/10.1016/j.proeng.2012.06.409

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., Leeb, R., Mehring, C., Miller, K. J., Müller-Putz, G. R., Nolte, G., Pfurtscheller, G., Preissl, H., Schalk, G., Schlögl, A., Vidaurre, C., Waldert, S., & Blankertz, B. (2012). Review of the BCI Competition IV. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00055

Taran, S., & Bajaj, V. (2020). Sleep Apnea Detection Using Artificial Bee Colony Optimize Hermite Basis Functions for EEG Signals. *IEEE Transactions on Instrumentation and Measurement*, *69*(2), 608–616. https://doi.org/10.1109/TIM.2019.2902809

Tatum, W. O., Olga, S., Ochoa, J. G., Munger Clary, H., Cheek, J., Drislane, F., & Tsuchida, T. N. (2016). American Clinical Neurophysiology Society Guideline 7: Guidelines for

EEG Reporting. *Journal of Clinical Neurophysiology*, *33*(4), 328–332. https://doi.org/10.1097/WNP.0000000000000319

Tayeb, Z., Fedjaev, J., Ghaboosi, N., Richter, C., Everding, L., Qu, X., Wu, Y., Cheng, G., & Conradt, J. (2019). Validating Deep Neural Networks for Online Decoding of Motor Imagery Movements from EEG Signals. *Sensors*, *19*(1), 210. https://doi.org/10.3390/s19010210

The Royal Society. (2019). *Explainable AI: The basics*. Publications Office. https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf

Tosi, J., Taffoni, F., Santacatterina, M., Sannino, R., & Formica, D. (2017). Performance Evaluation of Bluetooth Low Energy: A Systematic Review. *Sensors*, *17*(12), 2898. https://doi.org/10.3390/s17122898

Tseghai, G. B., Malengier, B., Fante, K. A., & Langenhove, L. V. (2021). The Status of Textile-Based Dry EEG Electrodes. *Autex Research Journal*, *21*(1), 63–70. https://doi.org/10.2478/aut-2019-0071

Tsow, F., Kumar, A., Hosseini, S. H., & Bowden, A. (2021). A low-cost, wearable, do-it-yourself functional near-infrared spectroscopy (DIY-fNIRS) headband. *HardwareX*, *10*, e00204. https://doi.org/10.1016/j.ohx.2021.e00204

Tullis, P. (2019). The US military is trying to read minds. https://www.technologyreview.com/2019/10/16/132269/us-military-super-soldiers-control-drones-brain-computer-interfaces/

Utsumi, K., Takano, K., Okahara, Y., Komori, T., Onodera, O., & Kansaku, K. (2018). Operation of a P300-based brain-computer interface in patients with Duchenne muscular dystrophy. *Scientific Reports*, *8*(1), 1753. https://doi.org/10.1038/s41598-018-20125-6

V. Utkin, L. (2019). An imprecise extension of SVM-based machine learning models. *Neurocomputing*, *331*, 18–32. https://doi.org/10.1016/j.neucom.2018.11.053

van Ginneken, B. (2017). Fifty years of computer analysis in chest imaging: Rule-based, machine learning, deep learning. *Radiological Physics and Technology*, *10*(1), 23–32. https://doi.org/10.1007/s12194-017-0394-5

Vidal, J. J. (1973). Toward Direct Brain-Computer Communication. *Annual Review of Biophysics and Bioengineering*, *2*(1), 157–180. https://doi.org/10.1146/annurev.bb.02.060173.001105

Vijayarangan, S., Murugesan, B., R, V., SP, P., Joseph, J., & Sivaprakasam, M. (2020). Interpreting Deep Neural Networks for Single-Lead ECG Arrhythmia Classification [Publisher: arXiv Version Number: 1]. https://doi.org/10.48550/ARXIV.2004.05399

Vlahou, A., Hallinan, D., Apweiler, R., Argiles, A., Beige, J., Benigni, A., Bischoff, R., Black, P. C., Boehm, F., Céraline, J., Chrousos, G. P., Delles, C., Evenepoel, P., Fridolin, I., Glorieux, G., van Gool, A. J., Heidegger, I., Ioannidis, J. P., Jankowski, J., ... Vanholder, R. (2021). Data Sharing Under the General Data Protection Regulation: Time to Harmonize Law and Research Ethics? *Hypertension*, *77*(4), 1029–1035. https://doi.org/10.1161/HYPERTENSIONAHA.120.16340

Vuckovic, A. (2010). Motor imagery questionnaire as a method to detect BCI illiteracy. *2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2010)*, 1–5. https://doi.org/10.1109/ISABEL.2010.5702803

Vuckovic, A., & Osuagwu, B. A. (2013). Using a motor imagery questionnaire to estimate the performance of a Brain–Computer Interface based on object oriented motor imagery. *Clinical Neurophysiology*, *124*(8), 1586–1595. https://doi.org/10.1016/j.clinph.2013.02.016

Waldert, S. (2016). Invasive vs. Non-Invasive Neuronal Signals for Brain-Machine Interfaces: Will One Prevail? *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00295

Wang, B., Wong, C. M., Kang, Z., Liu, F., Shui, C., Wan, F., & Chen, C. L. P. (2021a). Common Spatial Pattern Reformulated for Regularizations in Brain–Computer Interfaces. *IEEE Transactions on Cybernetics*, *51*(10), 5008–5020. https://doi.org/10.1109/TCYB.2020.2982901

Wang, B., Wong, C. M., Kang, Z., Liu, F., Shui, C., Wan, F., & Chen, C. L. P. (2021b). Common Spatial Pattern Reformulated for Regularizations in Brain–Computer Interfaces. *IEEE Transactions on Cybernetics*, *51*(10), 5008–5020. https://doi.org/10.1109/TCYB.2020.2982901

Wang, P., Jiang, A., Liu, X., Shang, J., & Zhang, L. (2018). LSTM-Based EEG Classification in Motor Imagery Tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(11), 2086–2095. https://doi.org/10.1109/TNSRE.2018.2876129

Wang, Y.-T., Wang, Y., & Jung, T.-P. (2011). A cell-phone-based brain–computer interface for communication in daily life. *Journal of Neural Engineering*, *8*(2), 025018. https://doi.org/10.1088/1741-2560/8/2/025018

Warren, B., & Randolph, A. B. (2019). Facebrain: A p300 BCI to facebook [Series Title: Lecture Notes in Information Systems and Organisation]. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Information systems and neuroscience* (pp. 119–124, Vol. 29). Springer International Publishing. https://doi.org/10.1007/978-3-030-01087-4_14

Waytowich, N. R., Lawhern, V., Garcia, J. O., Cummings, J., Faller, J., Sajda, P., & Vettel, J. M. (2018). Compact Convolutional Neural Networks for Classification of Asynchronous Steady-state Visual Evoked Potentials [Publisher: arXiv Version Number: 2]. https://doi.org/10.48550/ARXIV.1803.04566

Wei, X., Dong, E., & Zhu, L. (2021). Multi-class MI-EEG Classification: Using FBCSP and Ensemble Learning Based on Majority Voting. *2021 China Automation Congress (CAC)*, 872–876. https://doi.org/10.1109/CAC53003.2021.9728576

Wei, Y., Li, J., Ji, H., Jin, L., Liu, L., Bai, Z., & Ye, C. (2022). A Semi-Supervised Progressive Learning Algorithm for Brain–Computer Interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 2067–2076. https://doi.org/10.1109/TNSRE.2022.3192448

Weng, J., Chen, H., Yan, D., You, K., Duburcq, A., Zhang, M., Su, H., & Zhu, J. (2021). Tianshou: A highly modularized deep reinforcement learning library. *arXiv preprint arXiv:2107.14171*.

Willingham, E. J. (2021). *The tailored brain: From ketamine, to keto, to companionship, a user's guide to feeling better and thinking smarter* (First U.S. edition). Basic Books.

Wolff, R. F., Moons, K. G., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., Mallett, S., & for the PROBAST Group†. (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*, *170*(1), 51. https://doi.org/10.7326/M18-1376

Wolpaw, J., & Wolpaw, E. W. (2012). *Brain–Computer InterfacesPrinciples and Practice*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195388855.001.0001

Wolpaw, J., Birbaumer, N., Heetderks, W., McFarland, D., Peckham, P., Schalk, G., Donchin, E., Quatrano, L., Robinson, C., & Vaughan, T. (2000). Brain-computer interface technology: A review of the first international meeting [Conference Name: IEEE Transactions on Rehabilitation Engineering]. *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 164–173. https://doi.org/10/dz4238

Won, K., Kwon, M., Jang, S., Ahn, M., & Jun, S. C. (2019). P300 Speller Performance Predictor Based on RSVP Multi-feature. *Frontiers in Human Neuroscience*, *13*, 261. https://doi.org/10.3389/fnhum.2019.00261

Wong, L. L. N., Chen, Y., Wang, Q., & Kuehnel, V. (2018). Efficacy of a Hearing Aid Noise Reduction Function. *Trends in Hearing*, *22*, 233121651878283. https://doi.org/10.1177/2331216518782839

World Health Organization. (2018). *Global status report on road safety 2018*. Retrieved March 15, 2022, from https://apps.who.int/iris/handle/10665/276462

Xu, B., Li, W., Liu, D., Zhang, K., Miao, M., Xu, G., & Song, A. (2022). Continuous Hybrid BCI Control for Robotic Arm Using Noninvasive Electroencephalogram, Computer Vision, and Eye Tracking. *Mathematics*, *10*(4), 618. https://doi.org/10.3390/math10040618

Xu, G., Shen, X., Chen, S., Zong, Y., Zhang, C., Yue, H., Liu, M., Chen, F., & Che, W. (2019). A Deep Transfer Convolutional Neural Network Framework for EEG Signal Classification [Conference Name: IEEE Access]. *IEEE Access*, *7*, 112767–112776. https://doi.org/10.1109/ACCESS.2019.2930958

Xu, J., Mitra, S., Van Hoof, C., Yazicioglu, R. F., & Makinwa, K. A. A. (2017). Active Electrodes for Wearable EEG Acquisition: Review and Electronics Design Methodology [Conference Name: IEEE Reviews in Biomedical Engineering]. *IEEE Reviews in Biomedical Engineering*, *10*, 187–198. https://doi.org/10.1109/RBME.2017.2656388

Yao, D., Qin, Y., Hu, S., Dong, L., Bringas Vega, M. L., & Valdés Sosa, P. A. (2019). Which Reference Should We Use for EEG and ERP practice? *Brain Topography*, *32*(4), 530–549. https://doi.org/10.1007/s10548-019-00707-x

Zancanaro, A. (2021). *Fbcsp python*. Retrieved August 24, 2022, from https://github.com/jesus-333/FBCSP-Python

Zerafa, R., Camilleri, T., Falzon, O., & Camilleri, K. P. (2018). A comparison of a broad range of EEG acquisition devices – is there any difference for SSVEP BCIs? *Brain-Computer Interfaces*, *5*(4), 121–131. https://doi.org/10.1080/2326263X.2018.1550710

Zhang, R., Zong, Q., Dou, L., & Zhao, X. (2019). A novel hybrid deep learning scheme for four-class motor imagery classification. *Journal of Neural Engineering*, *16*(6), 066004. https://doi.org/10.1088/1741-2552/ab3471

Zhuang, J., & Yin, G. (2017). Motion control of a four-wheel-independent-drive electric vehicle by motor imagery EEG based BCI system. *2017 36th Chinese Control Conference (CCC)*, 5449–5454. https://doi.org/10.23919/ChiCC.2017.8028220