KAUNAS UNIVERSITY OF TECHNOLOGY

TOMAS UKTVERIS

MULTI-CLASS EEG SIGNAL CLASSIFICATION AND ACQUISITION SYSTEM FOR BRAIN-COMPUTER INTERFACE

Doctoral Dissertation
Nature Sciences, Informatics (N 009)

2019, Kaunas

This doctoral dissertation was prepared at Kaunas University of Technology, Faculty of Informatics, Department of Software Engineering, during the period of 2014–2019.

Scientific Supervisor:

Prof. Dr. Vacius JUSAS (Kaunas University of Technology, Nature Sciences, Informatics, N 009).

Doctoral dissertation has been published in:
http://ktu.edu

Editor: Brigita Brasienė

ISBN 978-609-02-1624-8

The bibliographic information about the publication is available at the National Bibliographic Data Bank (NBDB) in Martynas Mažvydas National Library of Lithuania.

KAUNO TECHNOLOGIJOS UNIVERSITETAS

TOMAS UKTVERIS

DAUGIAKLASIO EEG SIGNALO KLASIFIKAVIMO IR ĮRAŠYMO SISTEMA SMEGENŲ-KOMPIUTERIO SĄSAJAI

Daktaro disertacija
Gamtos mokslai, Informatika (N 009)

2019, Kaunas

Disertacija rengta 2014-2019 metais Kauno technologijos universiteto Informatikos fakultete Programų inžinerijos katedroje.

Mokslinis vadovas:

Prof. Dr. Vacius JUSAS (Kauno technologijos universitetas, Gamtos mokslai, Informatika, N 009).

Interneto svetainės, kurioje skelbiama disertacija, adresas:
http://ktu.edu

Redagavo: Brigita Brasienė

ISBN 978-609-02-1624-8

Leidinio bibliografinė informacija pateikiama Lietuvos nacionalinės Martyno Mažvydo bibliotekos Nacionalinės bibliografijos duomenų banke (NBDB).

CONTENTS

DEDICATION

To author's beloved parents, who always were near when he needed them and kept him going when he wanted to give up.

ACKNOWLEDGMENTS

The author of this dissertation is very grateful to his supervisor Prof. Dr. Vacius Jusas for guiding through this journey and always pointing in the right direction.

The author says a big thank you:

- *to the personnel of the Faculty of Informatics and the Department of Software Engineering for providing help,*
- *to all the reviewers for their helpful comments and suggestions to improve this work.*

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ABS - acrylonitrile butadiene styrene
ACSP - augmented common spatial patterns
ADC - analog to digital converter
AFE - analog frontend
BCI - brain-computer interface
BLE - Bluetooth low energy
BOM - bill of materials
BP - band power
CFFT - channel fast Fourier transform
CMRR - common mode rejection ratio
CNN - convolutional neural network
CPU - central processing unit
CSP - common spatial patterns
CV - channel variance
CWT - complex Morlet wavelet transform
DC - direct current
DCT - discrete cosine transform
DL - deep learning
DSP - digital signal processing
ECG - electrocardiogram
EEG - electroencephalogram
ELM - extreme learning machine
EMG - electromyogram
EP - evoked potential
ERD - event-related desynchronization
ERS - event-related synchronization
FBCSP - filter-bank common spatial patterns
FCMS - frequency complementary feature map selection
FFT - fast Fourier transform
FIR - finite impulse response
GPIO - general purpose I/O
GUI - graphical user interface
HCI - human-computer interaction
HR - high resolution
ICA - independent component analysis
ICN - intrinsic channel noise
kNN - k-nearest neighbours
LDA - linear discriminant analysis

LSB - least significant bit
MCE - mean channel energy
MCU - microcontroller
MI - motor imagery
ML - machine learning
MLP - multi-layer perceptron
MPR - myoelectric pattern recognition
MRIC - movement related independent components
MWE - mean window error
PC - personal computer
PCA - principal component analysis
PCB - printed circuit board
PGA - programmable gain amplifier
QDA - quadratic discriminant analysis
RAW - raw EEG signal
RBF - radial basis function
ReLU - rectified linear unit
REM - rapid eye movement
RMS - root mean square (or random map selection)
RSVP - rapid serial visual presentation
SEM - signal energy map
SFM - select all feature maps
SGD - stochastic gradient descent
SNR - signal to noise ratio
SPI - serial peripheral interface
SSVEP - steady-state visually evoked potential
STFT - short-term Fourier transform
SVM - support vector machine
TDP - time domain parameters
TFD - time-frequency distribution
TKEO - Teager-Kaiser energy operator
UART - universal asynchronous receive transmit
VCC - circuit voltage supply
VEP - visually evoked potentials

1. INTRODUCTION[1]

Brain-computer interface (BCI) systems try to narrow the gap between human and computer interaction. The direct control of computer applications by using only human mind and mental abilities can help to solve many rehabilitation, multimedia and gaming challenges. One of the key parts of a BCI system is accurate and fast algorithms that are capable of analysing electroencephalogram (EEG) signal potentials that are recorded along the human scalp. Such signals contain noise and other unwanted artifacts, which prevent from correctly determining-classifying imagined motoric actions (imagery). Though many algorithms were developed to overcome such issues, the problem still requires extensive work.

The motor imagery classification is one of many widespread machine-learning problems of BCI systems. With the need for human mind controlled applications, the recording of EEG has emerged as an optimal solution for non-interventional brain activity analysis. The ability to understand this brain induced electrical signal fully would greatly simplify the life of people with disabilities or break the barrier of natural interaction in the entertainment industry.

This work focuses on four-class motor imagery problem where the recorded EEG signal is classified into four different classes that correspond to four different human subject imagined motoric actions (left hand, right hand, feet and tongue movement). Even if a simpler two-class (binary) problem achieves good classification performance, the four-class still struggles to reach the same results and requires more scientific investigation.

1.1. Aim of the research

The aim of this dissertation is to create a method for solving four-class motor imagery (MI) classification problem.

Objectives of the work are the following:

1. Perform review and analysis of existing MI classification techniques and EEG recording systems in literature,
2. Evaluate common classification algorithms by using MI EEG signal database,
3. Develop an effective method for solving MI classification task,
4. Design the EEG system capable of recording MI signals,
5. Manufacture and validate correctness of the EEG recording system.

[1] This chapter uses parts of the article [1] UKTVERIS, T., and V. JUSAS. Application of Convolutional Neural Networks to Four-Class Motor Imagery Classification Problem. Information Technology And Control, vol. 46 (2), 2017, 260–273 and article [61] UKTVERIS, T., V. JUSAS. Comparison of Feature Extraction Methods for EEG BCI Classification, Information and Software Technologies: 21st International Conference, 2015, 81–92.

1.2. Object of the research

Deep learning Convolutional Neural Networks (CNN) have not been widely analysed and used for motor-imagery (MI) task. The main interest is feature extraction and Convolutional Neural Networks (CNN) classification methods for the motor imagery problem. The development of the design of compact EEG acquisition system for MI task.

1.3. Scientific novelty

A new feature extraction channel difference method has been proposed for the EEG data processing based on Bandpower and Laplace filtering approaches. The proposed algorithm gives a similar filtering performance to a well-known CSP (common spatial patterns) algorithm. A new method for a single dimension (1D) feature vector adaptation to two-dimensional (2D) feature maps has been proposed. The algorithm has been successfully validated during the experiments. The CNN based classification method has been adapted to solve four-class MI problem, and the experimentally acquired results were close to the other state-of-the-art methods. Moreover, a stackable and modular EEG acquisition system for MI has been developed to help record second four-class validation EEG dataset and spread BCI among the wider audience.

1.4. Practical applications

The classification, proposed feature extraction and feature adaptation methods have been validated on publicly available real EEG signal database that was gathered from healthy subjects. The created EEG acquisition system has been constructed and validated in real-life usage scenarios of healthy subjects.

1.5. Approbation of results

Five articles on the topic of the dissertation have been published. Two of the papers were printed in journals indexed in Web of Science. The experimental results of the dissertation have been published in three scientific conferences in Lithuania and abroad.

1.6. Structure of the dissertation

This dissertation is divided into five main chapters. The first chapter introduces reader with the main topic of brain-computer interfaces.

The second chapter analyses existing literature and tackles the problem of motor imagery (MI) task classification by using traditional machine learning approaches.

The third chapter introduces deep learning methods for the same EEG data classification task. A methodology is given to successfully apply Convolutional Neural Networks (CNN) to the four-class MI problem.

The fourth chapter introduces a practical approach of designing an EEG signal acquisition system for MI. A developed printed circuit board (PCB) design is presented and detailed along with the effective system validation techniques.

The last section summarizes the whole work and gives the main conclusions of this dissertation and recommendations for further research.

2. BRAIN COMPUTER INTERFACE

This dissertation chapter is an introduction to the scientific field of brain-computer interfaces (BCI). The basic functioning mechanisms of the human brain are discussed, and multiple BCI systems are presented along with the actual real-world usage scenarios. The basic information is provided for the reader to understand further chapters of this work better.

2.1. Introduction

Brain as an organ of the human body has a unique ability to self-adapt, create new neural bonds while processing and learning new information. Brain is never the same. It progresses and evolves over the lifetime of a human being. Even now, while reading this paragraph, new information is being processed at tremendous speed, and some of it is stored in short or long-term memory. The large memory capacity of the brain serves as an experience and information storage. The environment sensing, body motoric actions are as well the responsibility and coordination of different brain regions. Each of the regions is responsible for the specific body part control, sensing or mental function. Unfortunately, the precise working mechanism of the brain is still unknown. More advanced mental functions such as the birth of an idea is a complete mystery. Without knowing the inner working, it is difficult to map or create a model of the brain. This very complex organ has been analysed by various neuroscientists and neurophysiologists who conduct experiments in order to find out the principal blocks and scientific theories behind the structure and behaviour of the brain. By checking and proving (or disproving) one hypothesis at a time, the whole puzzle is starting to be clearer. There are many problems that prevent deep brain analysis. Much of the problems lie in the variability of the brain. Even though every person has such a mental organ, it is still uniquely evolved in many functional and structural aspects. However, the emerging problems and difficulties do not prevent interest in brain function analysis. Many science fiction films give a huge spark to scientists to bring fiction to the real life. Only by developing and using new created methods to analyse brain structure and behaviour, it will be possible to find out the secrets that lie within the brain.

2.2. Structure of the brain

The anatomical structure of the brain helps to understand the processes happening inside. The structure needs to be reviewed in order to analyse brain functions that relate to motoric actions. This mental organ mainly consists of many connected and electrically active cells called neurons (about 100 billion) and glial cells that are nerve cells, which do not carry any nerve impulses. The neurons only make up 10% of the cells of the brain, while the larger part is glial cells. Unlike most other cells, neurons cannot regrow after damage. The glial cells provide physical, nutritional and digestive support for neurons while at the same time manufacturing myelin. The neurons carry electrical impulses and are connected in groups or

batches that fire electrical signals at the same time. Such coordinated burst of electrical impulses creates the electrical activity of the brain that can be measured by using scientific devices. Brain consists of three main parts: the cerebrum, cerebellum, and brain stem. The cerebellum is responsible for balance, posture and cardiac functions. The brain stem creates motoric and sensory pathway to the whole body. The largest part is the cerebrum consisting of white and grey matter tissue. The outer layer of the cerebrum, or where is folded the grey matter tissue, is called the cortex. The cortex is divided into frontal, parietal, temporal and occipital lobes (Figure 1). Usually, each of these lobes has different functions. The frontal lobe is known to be responsible for behaviour, intellect, smell, attention, motoric actions, etc. Occipital lobe processes vision and is involved in reading process. The primary motor cortex is in the top part of the brain, in the frontal lobe. This part is most active when motoric actions are done or when movement is just imagined.

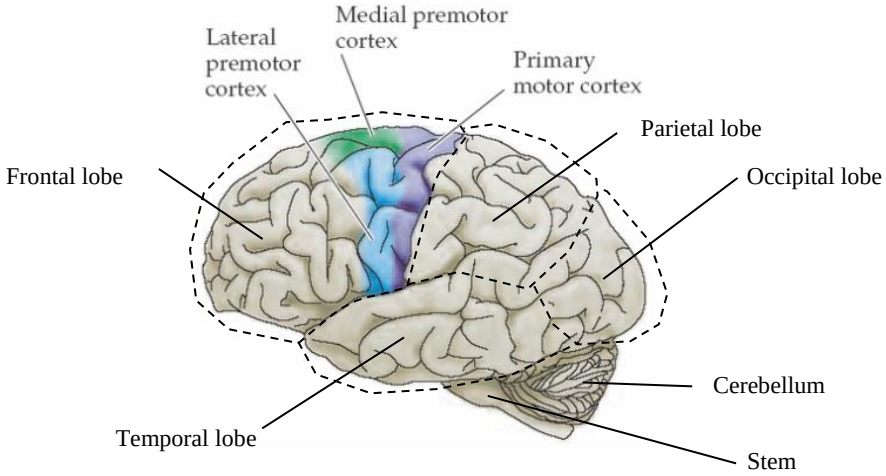


Figure 1. Brain cortex structure[2]

This work is based on the mentioned motoric imagination effect. In the further chapters, the scientific algorithms will be proposed and analysed that will help to discern what kind of motoric actions were imagined based on the effects and state of the brain motoric regions during that time. In order to be able to know the state of the brain, first, some measuring techniques need to be reviewed and selected. This is done in the next section.

2.3. Measuring activity

In order to understand and analyse how the brain works, the brain state data extraction methods are required. One of the methods used for brain scanning is Magnetic Resonance Imaging (fMRI). This method detects blood flow changes in active brain regions. It is known that blood flow increases in regions that are more

[2] https://www.researchgate.net/publication/281198195_Motor_impairments_following_stroke

active. Since fMRI machines became available in the early 90s, this brain mapping method became the most used one. It as well reached its popularity because it does not require any other medical intervention (e.g., surgery or ingesting radioactive substances). Long duration of the scans and non-realtime performance are the main drawbacks of this method. However, newer methods that improve spatial and temporal resolution are researched as well. Another approach for measuring brain activity is electroencephalography (EEG). EEG allows to measure electrical activity of the brain by using metal electrodes that are connected to specialized EEG signal recording hardware. Two EEG types are the most common, i.e., invasive and non-invasive. The invasive type requires surgery and embedding an electrode deep inside the brain tissue. The non-invasive type places metal electrodes along the head scalp. Even though the invasive method provides a less noisy and more accurate signal, the non-invasive approach is the most preferred one due to the less dangerous experimental conditions and possible negative consequences. EEG allows to capture and monitor spontaneous brain electrical activity so it could be used for real-time applications as compared to the much slower fMRI method. The electrical signal on the scalp surface is much weaker than recorded directly from the surface of the brain cortex. Since the skull and scalp skin suppresses the electrical signal, a more sensitive recording hardware is required. Compared to fMRI, EEG approach has a high temporal resolution, but a much lower spatial resolution. The nature of low spatial resolution is the fact that the electrical activity on the surface of the scalp is a combination of activities from multiple groups of neurons, just like a "chorus echo" of the neuron electrical signal. Since the EEG hardware is usually much cheaper than fMRI, it is as well a more preferred solution for real-time applications and tasks.

Due to the high availability and simplicity of the usage of EEG, the further sections and chapters of this work will analyse the EEG approach for brain data/signal acquisition. After the data is recorded by using specialized hardware, some digital signal processing needs to be done to filter and extract the state information for further analysis. This preprocessing step will be further discussed in the next section of this work.

2.4. Signal processing

Signal analysis and filtering methods are used to process EEG data. EEG is recorded via multiple electrical leads on the scalp; thus, the resulting recording is a multi-channel time series data. An example of such EEG is given in Figure 2.

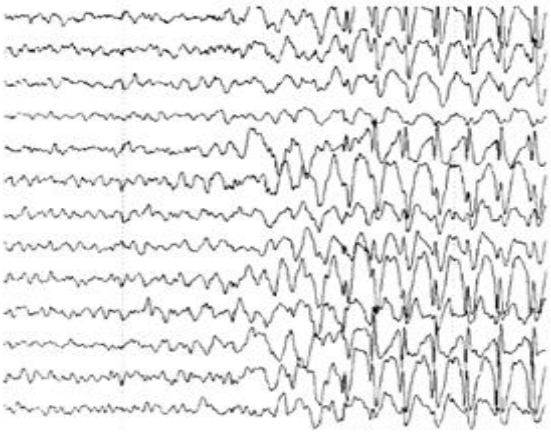


Figure 2. Normal EEG (left) and EEG during a seizure (right)[3]

The electrical leads can be placed anywhere on the skin surface. However, the standard placement positions are defined by the international 10–20 System of Electrode placement. This system maps the locations (Figure 3) of an electrode and the underlying area of the cortex. Each placement position [2] has a letter (to identify the lobe) and a number or another letter to identify the hemisphere location. The letters denote cortex regions: (F)rontal, (T)emporal, (C)entral, (P)arietal and (O)ccipital. Even numbers refer to the right hemisphere, and odd numbers refer to the left hemisphere. The "z" denotes the midline.

Depending on the hardware supported sampling frequency/resolution (in hertz), a different temporal resolution signal is obtained. The recorded EEG signal is very weak (typically tens of micro-volts) and susceptible to noise; thus, the amplification and good filtering is required. Usually, the recording hardware samples the analog signal N times faster than the sampling speed and averages the results to cancel out the noise effectively. In order to eliminate the common mode noise coming from the wall 50/60 Hz power line, a dedicated notch filter is used.

The brain waves (or EEG signal) oscillate in the range of 0–40 Hz; thus, a bandpass filter is commonly used to remove higher frequency noise above 40 Hz. The frequency domain of EEG is very important (Figure 4). Five types of EEG frequencies can be distinguished: delta (1–4 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (12–25 Hz) and gamma (30–100 Hz). Delta waves are noticeable in deep sleep or dreamless state. Theta waves usually appear in light sleep or REM (rapid eye movement) dream state, alpha waves appear in deep relaxation with eyes closed, beta waves in conscious concentrated, focused state. Gamma waves are associated

[3] https://www.health.harvard.edu/mind-and-mood/electroencephalogram-eeg

with bursts of insight or sudden insight and usually are ignored when analysing EEG. The brain wave ranges do not strictly reside in the mentioned regions and are subject to variations and overlaps.

Since signal analysis and preprocessing is only one of the key blocks to build a BCI system, the further discussion in the next section will be an introduction to the full BCI paradigm.

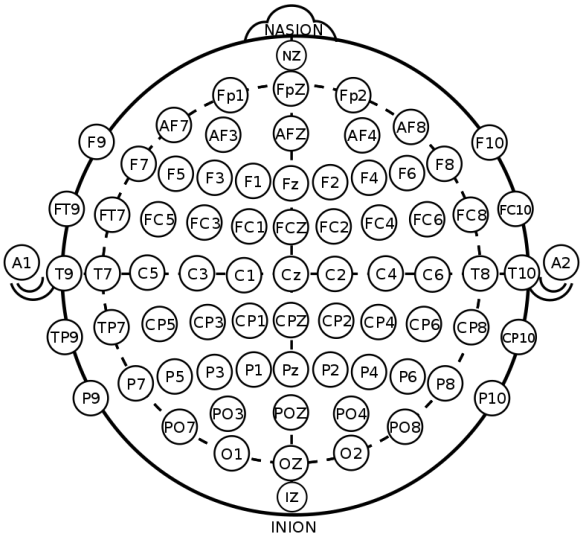


Figure 3. Electrode placement positions based on the 10–20 system[4]

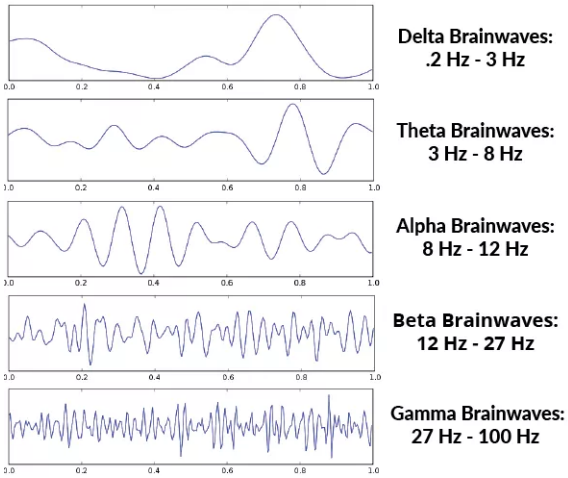


Figure 4. EEG signal frequencies[5]

[4] https://en.wikipedia.org/wiki/10–20_system_(EEG)
[5] https://rewiringtinnitus.com/science-brainwave-entrainment/

2.5. Brain-computer interface

A brain-computer interface (BCI) is a system that transforms signals originating from the human brain into commands that can control other devices. The main purpose of the BCI is to help ill or disabled people to communicate without the need of muscular movement. However, BCI has as well received a lot of interest from the entertainment sector due to the new human-computer interaction capabilities. In each BCI system, there is a method to convert brain signals into computer control commands. The method can use different hardware, i.e., fMRI, EEG or any other kind, and must be able to extract relevant control information from the recorded brain signals. This as well includes preprocessing, filtering, removing artifacts from signals that are recorded. The control algorithm must be able to adapt to each subject during the learning step, since each brain is unique. There must be some form of a task that the subject could complete to express his intention, and the system would interpret and execute. The feedback or notifications to the human subject should be presented on a computer screen or some other form (e.g., blinking a light). There are multiple functional BCI systems that already work by this principle.

It is known that sensory stimulation creates an electrical response in the brain. Such response is called evoked potential (EP). Stimulation can be as simple as playing a fixed frequency sound tone to the subject. It is known that certain frequency visual stimulation (for a human visual system, it is 3.5–75 Hz) causes same frequency oscillations in the EEG signals. There are BCI systems that use such stimulation approach for their mental task, and such EP is called visually evoked potentials (VEP). BCI system can flash different computer screen regions or lights at different fixed rates. The subject wanting to select one of the two options would need to focus and concentrate on the one with blinking regions/lights. An increase in energy in the EEG signal at the frequency related to the focused blinking frequency would be noticeable. Such VEP of the visual system is called Steady-State Visual Evoked Potential (SSVEP).

Another type of the BCI system is motor-imagery (MI), and this work is based on it. The next section of the dissertation is dedicated to give a brief overview of the MI concept and the use-case. More in-depth BCI structure, analysis and implementation are given in the next chapters of the work.

2.6. Motor imagery BCI[6]

Motor imagery (MI) is a type of BCI, where the human subject has to imagine a motoric action (e.g., wave right hand) without executing this movement physically (Figure 5). It has been shown that the imagination of a movement stimulates/activates the motoric cortex regions of the brain. Depending on the side of the moved limb, appropriate brain hemisphere is activated, and the opposite is

[6] This section uses parts of the article [60] UKTVERIS, T., V. JUSAS. Comparison of Feature Extraction Methods for EEG BCI Classification, Information and Software Technologies: 21st International Conference, 2015, 81–92.

deactivated. By imagining/selecting from a list of possible movements (left hand, right hand, etc.), it is possible to control devices or communicate.

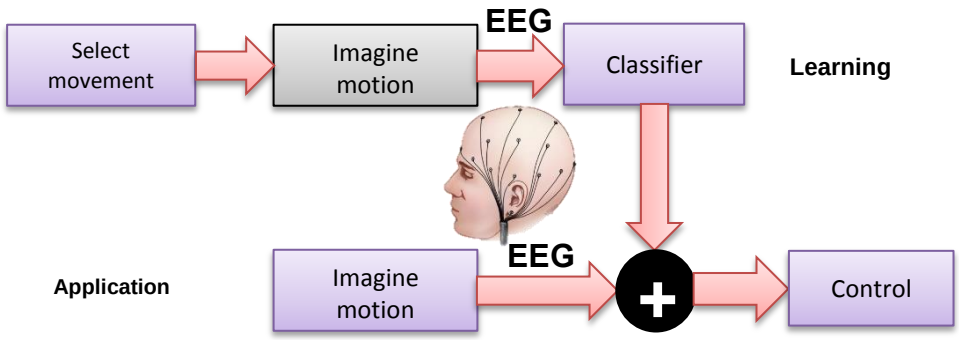


Figure 5. Motor imagery BCI workflow

The classification of a motoric action is to determine the action that the subject was thinking about when the MI EEG signal was recorded, given the EEG signal itself. The classification methods vary; however, machine learning (ML) approaches are used mostly due to complex brain oscillation patterns. In order to be able to classify motoric actions with great accuracy, correct and significant features must be extracted from the EEG signal for classifier training. In the process of motor imagery, various regions of the brain are induced differently: signal energy decreases or increases based on the motoric actions that are imagined. The effect of energy decrease is called Event-Related Desynchronization (ERD), and the opposite effect is named as Event-Related Synchronization (ERS). Since ERD/ERS describes transient changes in the brain signal oscillatory activity, a correct pattern of such information allows the classification of motor imagery tasks. However, the pattern extraction is error-prone due to the nature and highly non-deterministic brain activity even for the same test subject. After the classifier is trained for a specific subject, its output can be used for other device control.

2.7. Conclusions

This introduction chapter presented the motor imagery BCIs. After the short review, such conclusions can be made:

1. The BCI field creates new possibilities to human-computer interaction (HCI).
2. Motor imagery is one of many use cases for making the HCI possible. However, it can only work if the correct methods are chosen to identify and interpret brain electrical signals. Since these signals contain a lot of complex brain oscillations, an automated pattern extraction approach is needed.
3. New methods such as machine learning and deep learning have opened a new perspective frontier for complex pattern analysis, which is suitable for

brain signal analysis. By intelligently using these new tools more effective solutions can be implemented for the motor imagery BCI.

3. FEATURE EXTRACTION AND CLASSIFICATION[7]

This chapter discusses the use of various signal processing algorithms for the EEG feature extraction and classification. In the first section, a short literature review and feature extraction techniques for the EEG signal analysis are presented along with the new Channel difference feature extraction method. Further sections give an overview of the classification methods that are applied in practice along with more details about the experiment procedure and data that was selected to assess the algorithms. The results and findings, concluding remarks are given in the last section.

3.1. Introduction

Over the years, many multi-class BCI solutions were proposed that use different feature extraction methods. A feature extraction algorithm is one of the most critical parts of the EEG classification task or any BCI system processing pipeline. Since the BCI system accuracy directly depends on the quality of extracted feature vectors; thus, care must be taken to ensure quality. The pipeline can be viewed as an EEG signal processing filter chain as shown in Figure 6.

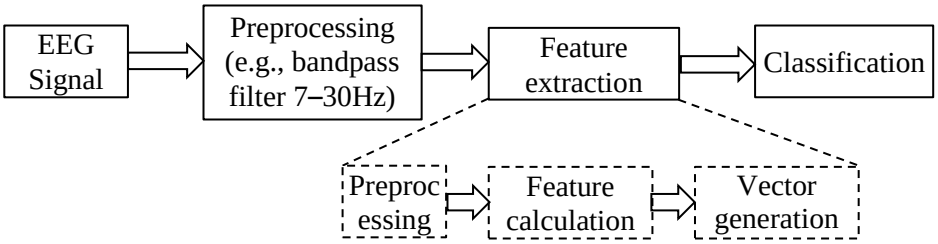


Figure 6. EEG signal processing filter pipeline

After the initial required signal preprocessing (that must be done for all signal channels), the feature extraction stage comes into view. For each extraction algorithm, such stage can be decomposed into three distinct processing blocks that influence the total performance of the algorithm:

1. *Preprocessing* [optional] is additional signal filtering (e.g., into different frequency bands).
2. *Feature calculation* process the EEG channel data (e.g., channel energy calculation).

[7] This chapter uses parts of the article [1] UKTVERIS, T., and V. JUSAS. Application of Convolutional Neural Networks to Four-Class Motor Imagery Classification Problem. Information Technology And Control, vol. 46 (2), 2017, 260–273 (Section 3.4) and article [60] UKTVERIS, T., V. JUSAS. Comparison of Feature Extraction Methods for EEG BCI Classification, Information and Software Technologies: 21st International Conference, 2015, 81–92 (Section 3.1, 3.3–3.8).

3. *Vector generation* compute elements of the final vector (e.g., mean of channel energy).

By controlling and changing the implementation in blocks, many algorithm variations can be acquired for the evaluation. This schema is handful in order to analyse subtle algorithm discrepancies as well.

3.2. Literature review

Various solutions to the motor-imagery classification problem have been proposed in literature. A brief review of each paper will be presented, while a summary is given in Table 1.

W. Yi et al. in [3] investigated the patterns of simple limb and compound limb MI and their separability for the use in BCI. By using event-related spectral perturbation (ERSP) and power spectral entropy (PSE) methods to observe the spectral power changes in the EEG, the authors of the paper found differences between simple and compound limb motor-imagery. Moreover, the multi-class common spatial patterns (CSP) filter was used and compared to multi-class stationary Tikhonov regularized CSP (TRCSP) filter for the EEG signal feature extraction. It was found that TRCSP outperforms other CSP methods. Support vector machine (SVM) classifier was used to classify the extracted feature vectors in the mentioned work. Another approach involving ERSP, power spectral density (PSD) and phase locking value (PLV) for phase synchronization analysis was done again by W. Yi et al. in [4] in order to verify the feasibility of motor sequences involving multiple limb MI. The results confirmed that the PSD outperformed multiclass CSP and showed that the motor sequences of multiple limbs can be utilized to build a multimodal BCI system. Quite different technique by A. Úbeda et al. was implemented in [5] where the assessment of feasibility to decode (classify) upper limb kinematics from the EEG signals was done. A multidimensional linear regression was used for the EEG kinematics decoding. The final obtained results suggested that the decoding accuracy might not be high enough (above chance levels) to perform a real-time control, and more analysis is needed. An interesting experiment to stimulate the motor cortex for classification of MI was done by I. N. Angulo-Sherman et al. in [6]. The stimulation was executed by using transcranial direct current stimulation (tDCS) method. A bandpass filter and independent component analysis (ICA) with spectral power estimation was used for feature preprocessing and extraction. The proposed montage improved the classification by 10%, while a high variable effect of stimulation was found in different motoric areas. The linear discriminant analysis (LDA) method was used for the classification of two motoric actions. The ability to discern self and third person MI analysis has been conducted by J. Andrade et al. in [7]. The time-frequency and source analysis (TFSA), including independent component analysis (ICA), CSP and principal component analysis (PCA), were the main methods used for the EEG signal processing. The classification has been implemented by using the SVM classifier. It was found that self and third person MI use distinct electrophysiological mechanisms that can be detectable at the scalp and thus used for MI BCI. K. Wang

et al. [8] proposed a MI-BCI paradigm in which commands are made by imagining clenching the right hand with different force loads. The work used three different loads for the right hand (light load, high load and relaxed), and the electromyogram (EMG) was used for the additional muscle artifact validation. The extracted features included time-frequency spectral power, ERSP, and they were classified by using SVM. It was showcased that multi-force load on the same limb can be used for the MI tasks successfully. A subject's intention analysis and decoding from a streaming EEG data was done by K. Georgiadis et al. in [9]. Six healthy subjects and six subjects with neuromuscular disease (NMD) were drawn into the experiment. Three MI tasks: left hand clenching, right hand clenching and rest were used. Phase synchrony using PLV method was employed to describe the functional coupling between all the recordings along with the ICA for the feature preparation. The features were classified by using the SVM classifier. The results showed that the increased phase synchrony in NMD patients could turn to a valuable tool for the MI decoding. Healthy patients and patients with complete spinal cord injury for 3-class MI were analysed by G. Pfurtscheller et al. in [10]. The CSP was used to filter the EEG signal, and time-frequency energy maps were extracted as features for the classification with LDA classifier. It was found that MI patterns are not very pronounced for the injured patients; thus, the classification accuracy was rather low. However, the EEG patterns for healthy subjects were clearly discriminable. Given the provided evidence, it was clear that the extensive training sessions are necessary to achieve a good BCI performance at least in some subjects. The improvement based on the multiclass posterior probability for twin SVM classifier was proposed by Q. She et al. in [11] for classifying MI EEG. Two different datasets were used for the performance validation. CSP method was used to extract relevant features from the bandpass filtered EEG signal. A twin SVM (TSVM) was used for the classification. The experiment results demonstrated that the proposed method yields slightly higher averaged mean kappa value than TSVM; it can achieve comparatively close performance to the SVM competitors with lower time complexity on the used datasets. Mutual information feature selection (MIFS) method and max-relevance min-redundancy (mRMR) criterion have been used by B. Xu et al. in [12] to select the most relevant features from the MI EEG signal of a 3-task right hand clench speed problem. Additional methods as Hilbert transform for phase extraction and CSP for feature filtering have been applied. The obtained feature vectors were classified by using SVM and extreme learning machine (ELM) classifiers. No significant difference in the classification rate between SVMs and ELM was found. However, it was found that time-frequency-phase feature can improve the classification rate by about 20% more than the time-frequency feature alone. A vast number of methods for feature extraction were used by H. Higashi et al. in [13] to solve a spatial weight determination problem for MI. The authors of the paper developed a data-driven criterion method (an extension to CSP) named common spatio-time-frequency patterns (CSTFP). The enhanced CSP feature extraction method was compared against filter-bank CSP (FBCSP), common sparse spectral spatial patterns (CSSSP) and discriminative FBCSP (DFBCSP). A linear LDA classifier was used for the classification. The proposed CSTFP method

achieved high classification accuracy and showed that it can effectively extract discriminative features for motor imagery. A new noise sensitivity reduction mechanism for CSP was proposed by R. Zhang et al. in [14] with an introduction of local temporal correlation CSP (LTCCSP). The new feature extraction method was compared to the local temporal CSP (LTCSP) and CSP. The data were classified by using the SVM classifier. LTCCSP achieved the highest average classification accuracies and showed that it can be a promising method for practical motor imagery BCI applications. A. Astigarraga et al. [15] analysed the way to select only relevant EEG channels for the EEG BCIs. A novel two-step method was introduced, in which, firstly, a computationally inexpensive greedy algorithm finds an adequate search range, and, then, the Estimation of Distribution Algorithm (EDA) is applied in the reduced range to obtain the optimal channel subset. Feature extraction and filtering was done with the CSP. SVM classifier was used as a fitness function for channel selection. It was found that the number of channels could be reduced drastically without losing accuracy. Thus, the making of specific selection of channels for each subject is necessary. The first attempt to control virtual and real robot by using four-class motor-imagery via optical BCI was implemented by A. M. Batula et al. in [16]. The correlation-based signal improvement (CBSI), common average referencing (CAR) and task-related component analysis (TRCA) methods were used for feature extraction, while the classification was carried out by using LDA classifier after the feature reduction and normalization. The results showed that motor-imagery can be improved with feedback, and a four-class motor-imagery-based fNIRS-BCI could be feasible with sufficient subject training. Separate Mu and Beta rhythm extraction and classification was proposed by Y. Kim et al. [17]. Using fully data-driven multivariate empirical mode decomposition (MEMD), the authors obtained mu and beta rhythms from the nonlinear EEG signals and filtered them via strong uncorrelating transform complex common spatial patterns (SUTCCSP) method. The extracted features were classified by using random forest (RF), logistic model tree (LMT), model tree (MT), k-nearest neighbours (kNN), logitboost (LB) classifiers. The proposed SUTCCSP outperformed both CSP and CCSP. MEMD proved to be a preferred preprocessing method for the nonlinear and nonstationary EEG. Rensong Liu et al. in [18] proposed regularized common spatial pattern (R-CSP) algorithm for the EEG feature extraction by incorporating the principle of generic learning. A combined kNN-SVM classifier by using kNN and SVM approaches was used to classify four anisomerous states. Feature extraction has been done by using Wavelet canonical correlation analysis (wCCA) and wavelet threshold denoising (WTD). The results indicate that the KNN-SVM classifier is more suitable for the recognition of the four MI states than the five mainstream classifiers. The work in [19] by V. Mondini et al. described a cue-paced, EEG-based BCI system using motor-imagery that has flexible training sessions, unbalancing in the training conditions, adaptive thresholds when giving feedback. Two-class BCI tasks were classified by using SVM. The EEG signal was preprocessed by using bandpass and CSP filters, and the final feature vectors were made out of log-transformed normalized variances of the time-series. The results showed that the participants should see only positive feedback (as negative induces EEG non-

stationarities). S. Guan et al. in [20] propose a novel classification framework and a novel data reduction method based on the manifold of covariance matrices in a Riemannian perspective. The features were extracted via semi supervised joint mutual information (semi-JMI) general discriminate analysis (GDA) or SJGDA. A subject-specific decision tree (SSDT-KNN) was designed to identify/classify MI tasks. The results showed that the proposed method is effective for the MI tasks in a BCI system. An improved tensor (time, frequency, channel) time based analysis scheme for the hybrid BCIs was proposed by H. Ji et al. in [21]. A weighted fisher criterion was designed to select multimodal discriminative EEG patterns, and nonredundant rank-one tensor decomposition model was introduced to select only the most effective tensors. The projection coefficients of the selected rank-one tensors for each hybrid task were calculated and concatenated as the feature vectors. The classification was done by using the SVM classifier. It was found that the proposed scheme is efficient in extracting multimodal discriminative patterns; however, the tensor generation and decomposition is very time-consuming. A single-channel hybrid BCI system combining motor imagery (MI) and steady-state visually evoked potential (SSVEP) approaches were proposed by L. W. Ko et al. in [22]. Short time Fourier transform (STFT) and common frequency-patterns (CFP) were used for the feature extraction, and the classification was done with LDA. The proposed hybrid system demonstrates a comparable level of classification accuracy combining important advantages of utilizing just a single EEG channel and providing more freedom in the channel placement as compared with a single-mode SSVEP-based BCIs. A study by I. Martišius et al. [23] proposed to use wave atom transform (WAT), band-power (BP) and adaptive CSP (ACSP) for feature extraction for the SSVEP BCI gaming system. The typical architecture, paradigms, requirements, and limitations of electroencephalogram-based gaming systems were discussed in the work, and 3-class tasks classified by using LDA and SVM classifiers. The average accuracy of 80.5% using SVM was achieved. The study showed that BCI can be feasible even when using low-resolution low-cost customer-grade EEG acquisition devices. Other multiclass BCI decoding algorithm was developed by I. Xygonakis et al. [24]. The features were extracted in the cortical source space from the selected Regions of Interest (ROIs). CSP filtering was used along with kNN, Naïve Bayes, Decision Tree and LDA classifiers. LDA had superior performance with the highest prediction accuracy among all the subjects; however, the developed algorithm did not reach accuracy levels of the state-of-the-art methods. A novel approach to MI, which combines the discriminative power of extreme learning machine (ELM) with the reconstruction capability of sparse representation, was proposed by Q She et al. in [25]. The methodology was tested on 2-class and 4-class problem datasets. The feature vectors were constructed via bandpass and CSP filtering the EEG signal and applying Fisher Discrimination Dictionary Learning (FDDL). The classification was done by using ELM. The method achieved superior performance results than the other existing algorithms, while comparable performance with other state-of-the-art methods for the UCI dataset was reached. M. Dai et al. [26] proposed a transfer kernel CSP (TKCSP) approach for the MI BCI feature extraction that learns a domain-invariant kernel

directly by matching distributions of source subjects and target subjects. The approach was a combination of Kernel CSP (KCSP) and Transfer kernel learning (TKL). The algorithm was compared to subject-to-subject transfer (SJ-to-SJ) CSP, regularizing CSP (RCSP), stationary subspace CSP (ssCSP), multitask CSP (mtCSP), and the combined mtCSP and ssCSP (ss + mtCSP) method. The SVM classifier was used for the classification. An improved performance was achieved by TKCSP. An attempt to use Power spectral density (PSD) was used by N. G. Ozmen et al. [27] for the feature extraction in the frequency domain by using frequency domain analysis. LDA and SVM classifiers were trained for classification. The results that were obtained were similar to the other state-of-the-art methods. M. Li et al. [28] presented a novel feature extraction method based on the Locally Linear Embedding (LLE) algorithm and discrete wavelet transform (DWT). Neural networks (NN) and genetic algorithm (GA) was used for the classification. The results confirmed that the proposed method outperforms the existing methods in classification accuracy with fewer feature dimensions. The nonlinear information that was contained in the MI-EEG signals can be represented by wavelet coefficients that depict the time-frequency energy distributions of the original signal. P. Batres-Mendoza et al. [29] presented an improvement to the quaternion-based signal analysis (QSA) as iQSA (improved QSA). The improved method extracts features in a more efficient way by reducing the number of samples needed to classify the signal and improving the classification percentage. Boosting and decision trees (DT) were used for the classification. iQSA showed significant ~ 50% improvement compared to the original QSA.

Table 1. Summary of the MI classification methods in literature

Source	Subjects	Classes	Methods	EEG system	Classifier	Data set	Cross-validation	Accuracy or result
[3]	10	3, 4	ERSP, PSE, CSP, TRCSP	64ch, Neuroscan SynAmps2, *Fs = 1 kHz	SVM	n/a	10-fold	Mean 70%
[4]	12	4	ERSP, PSD, PLV	64ch, Neuroscan SynAmps2, Fs = 1 kHz	SVM	n/a	10-fold	Mean 74.14%
[5]	5	2, 4, 8	Multidimensional linear regression	16ch, gUSBamp, Fs = 1200 Hz	Nearest neighbour	n/a	10-fold, 5-fold	Slightly above chance levels
[6]	5	2	tDCS	32ch, Enobio system, Neuroelectrics, Fs = 500 Hz	LDA	n/a	n/a	Improvement by +10%

[7]	20	2	TFSA, ICA, CSP, PCA	64ch, BrainVision acti-CHAMP, Fs = 1 kHz	SVM	n/a	5-fold	Mean 67%
[8]	11	2	ERSP, CSP	64ch, Neuroscan SynAmps2, Fs = 1 kHz	SVM	n/a	n/a	Mean 70%
[9]	12	2	PLV, ICA	63ch, Fs = 256 Hz, BePlusLTM Bioelectric Signal Amplifier	SVM	n/a	Custom (39/40)	Novel scheme
[10]	15	3	Time-Frequency energy maps, CSP	15ch, Easycap, g.Tec, Fs = 256 Hz	LDA	n/a	10-fold	Mean 67%
[11]	9	4	CSP, posterior probability model, Platt's estimating method	22ch, Fs = 250 Hz	SVM	BCI IV 2a, UCI	10-fold	Mean 72%
[12]	6	3	Hilbert transform, PCA, MIFS, mRMR, CSP	21ch, Neuroscan synamps2, Fs = 1000 Hz	SVM, extreme learning machine (ELM)	n/a	5-fold	+20% for time-freq-phase vs time-freq
[13]	4, 5	2	CSP, FBCSP, CSSSP, DFBCSP	BCI IV: 118ch, Fs = 1000 Hz, BCI III: 64ch, Fs = 1000 Hz	LDA	BCI III 4a, BCI IV 1	5-fold	92.5% for BCI III, 92.2% for BCI IV
[14]	5, 9, 13	2	LTCCSP, LTCSP, CSP	BCI III: 118ch, Fs = 1000 Hz, BCI IV: 22ch, Fs = 250 Hz, custom: 15ch, Fs = 1000 Hz	SVM	BCI III 4a, BCI IV 2a, custom	10-fold	LTCCSP 77.6–82.9%
[15]	3	4	Greedy selection, EDA, CSP	BCI III: 60ch, Fs = 250 Hz	SVM	BCI III dataset 3a	10-fold	Consistent
[16]	13	4	CBSI, CAR, TRCA	24 optodes, Hitachi ETG-4000, Fs = 10	LDA	n/a	n/a	Mean 27.12%

				Hz				
[17]	10 5	2	MEMD, SUTCCSP	64ch, Fs = 160 Hz	RF, LMT, MT, kNN, LB	Physiob ank MMI	5-fold	SUTCCS P 80.05%
[18]	5	4	R-CSP, wCCA	14ch+2 *ref, Fs = 256 Hz, g.Tec device	kNN-SVM, LDA, RF, Naïve Bayes	n/a	Cust om	Mean 87%
[19]	10	2	CAR, CSP	11ch, Brainbox EEG-1166, Fs = 128 Hz	SVM	n/a	n/a	Performa nce increase
[20]	9, 3, 7	2, 3	Semi-JMI, GDA or SJGDA	BCI IV: 22ch, Fs = 250 Hz, BCI III: 64ch, Fs = 250 Hz, custom: 14ch, Emotiv Epoc, Fs = 128 Hz	SSDT-KNN	BCI IV 2a, BCI III 3a, custom	10-fold	Kappa +0.037
[21]	9	2	Tensor multimodal analysis	15ch, Fs = 256 Hz	SVM	n/a	5-fold	Scheme is efficient
[22]	17	2	STFT, CFP	32ch, Neuroscan, Fs = 500 Hz	LDA	n/a	5-fold	Mean 85.6%
[23]	2	3	WAT, BP, ACSP	14ch+2 ref, Emotiv EPOC, Fs = 128 Hz	LDA, SVM	n/a	10-fold	Mean 80.5%
[24]	9	4	Ensemble classification model, Cortical ROIs, CSP	BCI IV: 22ch, Fs = 250	kNN, Naïve Bayes, DT, LDA	BCI IV 2a	10-fold	Mean 59.7%
[25]	9, 5, 3	4, 2	CSP, FDDL	BCI IV: 22ch, Fs = 250 Hz, BCI III: 118ch and 60ch, Fs = 1000 Hz	Non-linear ELM	BCI IV 2a, BCI III 3a and 4a, UCI	5-fold	Mean 2a: 64.46% 3a: 87.54% 4a: 80.68%
[26]	5	2	TKCSP, KCSP, TKL	BCI III: 118ch, Fs = 1000 Hz	SVM	BCI III 4a	n/a	TKCSP 81.4%

[27]	6	2, 5	PSD	64ch, Fs = 512 Hz, Biosemi ActiveTwo	LDA, SVM	n/a	n/a	Mean 91.85%
[28]	3	2	LLE, time-frequency features, DWT	2ch bipolar, Fs = 125 Hz	3-layer neural network (NN), genetic algorithm (GA)	BCI III 3b	10-fold	Mean 91.43%
[29]	39, 4	2, 3	Quaternion-based signal analysis (QSA), improved QSA (iQSA)	14ch+2 ref, Emotiv EPOC, Fs = 128 Hz, BCI IV: 22ch, Fs = 250 Hz	Decision trees (DT) using boosting	BCI IV 2a	n/a	BCI IV 2a: mean 82.30%

*Fs = sampling frequency, ref = reference

A short review of existing methods shows the popularity of machine learning (ML) approaches (LDA, SVM, kNN, neural networks, etc.) for the MI classification task. The classifier method summary is given in Table 2.

Table 2. Summary of the classification methods

ML branch	**Method examples**	**Structure or behaviour**	**Calculated metrics**	**Accuracy**	**Training speed**
Instance based	kNN	Lazy learning, non-parametric, data memorization	Euclidean distance	Lower	Fast
Bayesian	Naïve Bayes	Statistical	Membership probability	Lower	Fast
Linear/ Nonlinear	LDA/QDA	Linear functions	Linear or quadratic parameters (weights)	Lower	Fast
	SVM	Binary (non)linear	Hyperplane parameters	Higher	Slow
Neural networks	Perceptron, ELM	Multilayer neural architecture	Weights	Higher	Slow
Decision trees	Decision tree	Rule based	Leafs/nodes	Lower	Fast
Deep learning	Convolutional Neural Networks	Advanced neural architecture	Filter weights	Higher	Slow
Ensemble	Random forest, Boosting	Vote based	Votes	Higher	Slow

The selection of feature extraction methods in literature is more varied; however, the EEG signal power and frequency domain analysis techniques show the largest interest because of its intuitive use and greater comprehension. The CSP method is the leader amongst feature filtering techniques.

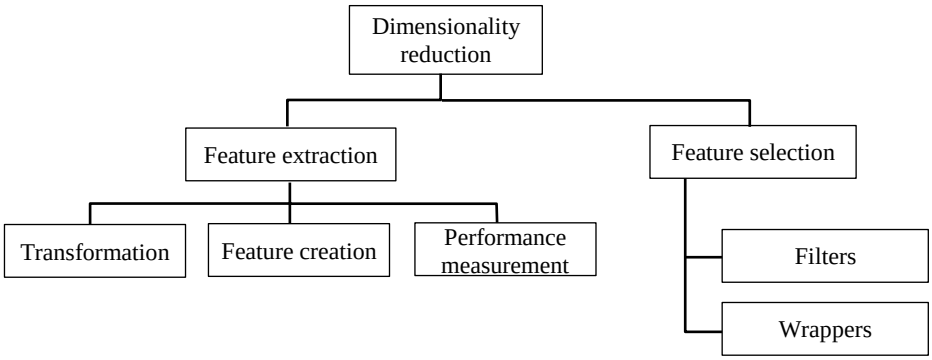


Figure 7. Dimensionality reduction techniques

In general, the feature extraction and feature selection methods are dimensionality reduction techniques (Figure 7). They eliminate redundant data; thus, the classification can be more efficient [30].

Feature selection tries to choose the most significant features from all the available input data. Feature selection methods can be organised into three categories: filters, wrappers and embedded/hybrid approaches. Filter methods tend to be fast, but not very reliable. They are good for large datasets. Wrappers use classification to optimize the feature selection for the specific problem, while introducing large computational cost and slowness. Hybrid methods use a mixture of two previous techniques.

Feature extraction works by creating new features that summarize or describe the input data differently or more efficiently. Three types of feature extraction exist: transformation, feature creation and performance measurement. Transformation type extraction applies some form of a transformation (e.g., linear transformation, etc.) in order to shape the data differently. Feature creation produces new (not seen) smaller group of features from the existing feature information. The performance measurement evaluates the existing features (e.g., statistically) and produces new features that describe some input data characteristics. The next section of this work will concentrate on the feature extraction methods. Some of the approaches will be detailed and analysed further.

3.3. Techniques for feature extraction

Five common EEG signal feature extraction methods found in the literature were implemented and analysed in this work: Band Power features (BP), Time Domain Parameters (TDP), Teager-Kaiser Energy Operator (TKEO), Signal power

and new Channel difference. The additional EEG preprocessing step was done by using Common Spatial Patterns (CSP) filtering. Each of the algorithms will be briefly detailed further.

3.3.1. Signal power features

One of the simplest methods for signal energy calculation. The power of a signal is the sum of squares of its time-domain samples divided by the signal length (1). The power is computed for every EEG channel, and the result is used further as a feature vector.

$$P = \frac{1}{N}\sum_{k=1}^{N} x^2[k], \tag{1}$$

where x is the discrete EEG channel signal, N is the number of EEG signal samples taken, and k is the index of the EEG channel vector.

3.3.2. Band Power features

The algorithm calculates multiple band power by band-pass filtering [31] the signal. In order to apply the algorithm, the signal frequency range must be divided into multiple regions. First, the initial EEG signal is filtered by using a band-pass filter designed for each frequency band, e.g., 4-th order Butterworth finite impulse response (FIR) filter. The resulting signal is squared to obtain its power. The resulting signal values p_{ij} of each i-th band and j-th EEG channel are then squared to obtain the power, and a w-size smoothing window operation is performed to filter the signal as shown in (2):

$$\bar{p}_{ij}[n] = ln\left(\frac{1}{w}\sum_{k=0}^{w} p_{ij}[n-k]^2\right), i = \overline{1,3}, \tag{2}$$

where n is the vector element index. The computed result is used for the feature vector generation. Such method is already implemented in MATLAB signal processing library.

Three different frequency bands were used in the work: 8–14 Hz, 19–24 Hz and 24–30 Hz, which correspond to *Mu*, *Alfa* and *Beta* brain waves. A similar approach was used in a closed loop system for navigating in a virtual environment via ERD-BCI [32]. The complex band power features were selected from three major frequency bands of cortical oscillations: μ (8–12 Hz), sensorimotor rhythm (12–15 Hz) and β (15–30 Hz).

3.3.3. Time Domain Parameters

Similar to the BP algorithm, time domain parameters compute time-varying power of the first k derivatives of the signal. The obtained derivative values (3) are

smoothed by using the exponential moving average, and a logarithm is taken as given by (5). The resulting signal is used in the feature vector generation.

$$p_i(t) = \frac{d^j x_i(t)}{dt^j}, j = 0,1, \dots, k, \tag{3}$$

$$p_i[n] = (x_i[n+1] - x_i[n-1])/2, \tag{4}$$

$$\bar{p_i}[n] = \ln(u \cdot p_i[n] - (1-u) \cdot p_i[n-1]), \tag{5}$$

where x_i is the initial *i-th* channel EEG signal at time *t*, $p_i(t)$ is the continuous signal derivative value at time *t*, $p_i[n]$ is the discrete signal derivative approximation (e.g., (4)) value at index *n*, *u* is the moving average parameter ($u \in [0; 1]$), and $\bar{p_i}[n]$ is the *n*-th component of smoothed signal derivative vector.

3.3.4. Teager-Kaiser Energy Operator (TKEO)

A non-linear algorithm for a more accurate signal energy calculation was presented by Teager and further analysed by Kaiser [33]. The advantage of TKEO is the ability to discover high-frequency low-amplitude components and take into account the frequency component and signal amplitude of the signal [34]. The algorithm for a continuous signal at time *t* can be written as shown in the Equation (6), while an approximation (7) exists for the discrete signals.

$$\Psi[x_i(t)] = \left(\frac{\partial x_i}{\partial t}\right)^2 - x_i(t) \cdot \frac{\partial^2 x_i}{\partial t^2}, \tag{6}$$

$$\Psi\big[x_i[n]\big] = x_i^2[n] - x_i[n-1]x_i[n+1], \tag{7}$$

where x_i is the *i-th* channel EEG signal vector, $\Psi\big[x_i[n]\big]$ are energy values computed for the discrete signal at EEG signal vector *n*-th index. The algorithm was applied for each of the EEG channels; then, the final feature vectors were generated.

3.3.5. Channel Difference method

Since the feature extraction method is one of the critical parts of a BCI, a more effective solution to this step inevitably brings higher classification performance. Channel difference algorithm is a new method presented in this dissertation for extracting EEG signal features. The method is an extension to the Band power feature algorithm with an extra signal filtering step. The algorithm (Figure 8) works by computing filtered features only for those EEG channels that have at least four symmetrical neighbouring electrodes around them. The logic behind this approach is to let Laplace filter reduce the number of EEG input channels (and data redundancy) that are used for the feature extraction. Less noise helps the classifier to learn new

features more effectively. Moreover, in order to extract only the non-common (unique) signal part of the spatial EEG channel neighbourhood for classification, the neighbourhood channels, usually, oscillate similarly or have the same common signal basis (noise). This baseline noise signal is not desired and should be eliminated. It was already shown that Laplace signal filtering is effective at enhancing EEG spatial resolution [36] and discerning EEG signals from the background [37] noise. Thus, the Laplace approach (Figure 9) has been selected for the filtering step.

For motor-imagery problems, the electrodes have to be chosen to be symmetric and cover both hemispheres in order to be able to capture all energy changes induced by the motor-imagery ERD/ERS processes of different sides of the body [35]. Thus, a good selection of electrodes having at least four neighbours (from international 10–20 system) could be: C3, C4, F3, F4, P3, P4, Cz, Pz, Fz, if they are available in the EEG system.

The Band power method implies the decomposition of EEG signal into multiple frequency components. It is common to use the frequencies related to *Mu*, *Alfa* and *Beta* brain waves. The measurement of signal power for motor-imagery seems intuitive due to the ERD/ERS processes happening in the brain during the MI BCI experiment. The number of bands is a matter of choice and discussion, since the brain waves do not strictly fall into the frequency ranges, are subject specific and tend to overlap due to the brain dynamics/non-stationarity.

Algorithm: Channel difference

Input: E ← matrix of the EEG signals of all electrodes (size: channels × samples)
u ← matrix of indices of *n* selected electrodes (having at least 4 symmetric neighbours around) and their neighbour electrodes (size: $n \times 9$)

Output: $v[1..m]$ ← constructed feature vector (length: $m = 3n$)

```
1.  k = 1
2.  for i=1 to n do
3.      Q = laplace(E, u[i])          /* Filter the electrode EEG signal by using
                                         Laplace filter.*/
4.      Z = bandpower(Q)              /* Apply Band Power method to filtered
                                         signal to get signals in Mu, Alfa and Beta
                                         frequencies/bands. */
5.      for j=1 to 3 do
6.          v[k] = mean(Z_j)          /* Compute the feature vector element (e.g.,
7.          k = k+1                      mean, variance, etc.) from a signal Z band
8.      end for                       */
9.  end for
```

Figure 8. Channel difference pseudocode

Algorithm: Laplace filter

Input: E ← matrix of the EEG signals of all electrodes (channels × samples)
u[1..9] ← indices of group of electrodes (flattened 3x3 configuration)

Output: e ← filtered EEG signal samples (array of size: 1× samples)

```
1.  function laplace(E, u)                          /* List of Laplace
2.    K = [-0.5, -1, -0.5; -1, 6, -1; -0.5, -1, -0.5]  coefficients.*/
3.    e = {0}                                       /* Array of zeroes */
4.    k = 0
      for i=1 to 3 do
5.      for j=1 to 3 do
6.        e = e + K[i, j] * E[u[k]]                 /* Convolution operation with
7.        k = k + 1                                 coefficient matrix.*/
8.      end for
9.    end for

10.   return e
11. end
```

Figure 9. Laplace algorithm pseudocode

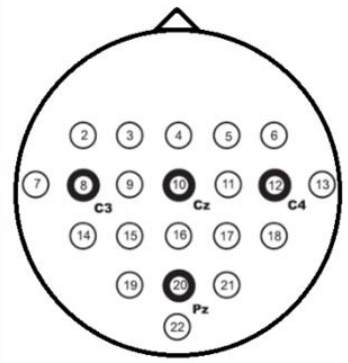


Figure 10. EEG channels from the 10–20 system used in the calculation

In this work, four EEG channels that match the 1st criteria of the method, i.e., electrodes C3, Cz, C4 and Pz (in international system 10–20), were selected as shown in Figure 10. Each of the selected channels has 4 or more neighbours and was filtered by using a Laplace filter (8) in a single channel radius neighbourhood with a standard 3x3 Laplace kernel as given in (9).

$$Q = \sum_{i=1}^{3} \sum_{j=1}^{3} K_{ij} E_{ij}, \tag{8}$$

$$K = \begin{bmatrix} -0.5 & -1 & -0.5 \\ -1 & 6 & -1 \\ -0.5 & -1 & -0.5 \end{bmatrix}, \tag{9}$$

where E_{ij} is the neighbouring EEG channel signal, K_{ij} is the corresponding weight from the matrix at *i*-th row and *j*-th column. For example, by taking Cz as the selected channel, the K_{11} element would denote a weight of -0.5 for the 3rd EEG channel, and K_{22} would denote weight 6 for the Cz channel etc. For non-existing channels, zero weight was used. Larger matrix sizes were not analysed as it would require more EEG channels in the dataset.

After filtering, the band power of frequency ranges: 8–14 Hz, 14–19 Hz, 19–24 Hz, 24–30 Hz, was computed for each of the four signals. The 16 resulting energy bands were used for the feature vector generation.

3.4. CSP preprocessing

Common spatial patterns (CSP) is a preprocessing technique (filter) for separating a multivariate signal into subcomponents that have maximum differences in variance [38]. The separation allows for easier signal classification. In general, the filter can be described (10) as a spatial coefficient matrix *W*:

$$S = W^T E, \tag{10}$$

where *S* is the filtered signal matrix, *E* is the original EEG signal vector. The columns of *W* denote spatial filters, while the inverse of *W* (i.e., W^{-1}) are spatial patterns of the EEG signal. The criterion of CSP for a two C_1, C_2 class problem is given by:

$$maximize: tr(W^T \sum_1 W), \tag{11}$$

$$subject\ to: W^T(\sum_1 + \sum_2)W = I, \tag{12}$$

where

$$\sum_1 = exp_{E \in C_1}\left(\frac{EE^T}{trace(EE^T)}\right), \tag{13}$$

$$\sum_2 = exp_{E \in C_2}\left(\frac{EE^T}{trace(EE^T)}\right) \tag{14}$$

are the class covariance matrices. The solution can be acquired by solving generalized eigenvalue problem by decomposing the problem into multiple standard eigenvalue sub-problems. Multiclass solutions are combined of multiple spatial filters. For more information, see [39]. Due to the broad and positive acknowledgement of the CSP, the method was used in this work to filter the EEG data before commencing feature extraction.

3.5. Methods for classification

From an extensive list of known EEG signal classifiers, the most commonly used in other literature were selected for the initial analysis, i.e., Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and k-Nearest Neighbours (kNN). A brief overview of the classification algorithms will be given further.

3.5.1. Linear and quadratic discriminant analysis

A classifier employs the Bayes' theorem for the classification. The discriminant analysis estimates the parameters of the Gaussian distribution for each class (15), and the trained classifier finds the class with the smallest misclassification cost. The posterior probability that a point *x* belongs to the class *C* is the product of the prior probability and multivariate normal density. The density function of the multivariate normal with mean μ_c and covariance Σ_c at a point *x* is given by:

$$P(x|C) = \frac{1}{(2\pi|\Sigma_c|)^{1/2}} exp\left(-\frac{1}{2}(x-\mu_c)^T \Sigma_c^{-1}(x-\mu_c)\right), \quad (15)$$

where $|\Sigma_c|$ is the determinant of Σ_c, and Σ_c^{-1} is the inverse matrix[8].

If *P(C)* represents the prior probability of class *C*, then, the posterior probability that an observation *x* is of class *C* is given by (16):

$$P(C|x) = \frac{P(x|C)P(C)}{P(x)}, \quad (16)$$

where *P(x)* is a constant equal to the sum over *C* of *P(x|C)P(C)*.

The Linear discriminant (or Fisher discriminant) analysis model is named for its inventor, R. A. Fisher [40]. Linear method (LDA) has the same covariance matrix for each class, and only means vary. For quadratic discriminant analysis (QDA), both means and covariances of each class vary.

[8] https://www.mathworks.com/help/stats/prediction-using-discriminant-analysis-models.html

3.5.2. Support Vector Machine

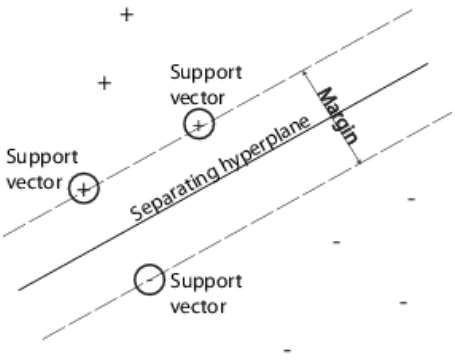


Figure 11. Finding separating SVM hyperplane between features

The Support Vector Machines (SVM) were introduced by Boser, Guyon and Vapnik [41] in 1992. SVM is a two-class algorithm that classifies data by finding the best hyperplane (17) separating all one class points from all of the other classes (with the largest margin) as shown in Figure 11.

$$\boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{b} = 0, \tag{17}$$

where $\boldsymbol{w}$ is the hyperplane coefficients vector, $\boldsymbol{x}$ is the point, and $\boldsymbol{b}$ is the hyperplane constant vector.

The problem is of dual quadratic programming nature that can be reduced to the Lagrangian optimization problem. A scheme of ONE *vs* ALL or ONE *vs* ONE is used if more than two classes are needed.

3.5.3. k-Nearest Neighbours

kNN is one of the simplest algorithms for classification. A feature vector is classified by a majority vote of its neighbours. The object class is assigned to the most common one found among *k* nearest neighbours (e.g., to class "square" as given in Figure 12).

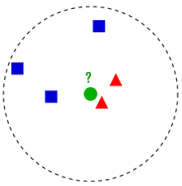


Figure 12. Classifying object with a kNN classifier (*k = 5)*

3.6. Data selection and experiment

3.6.1. Evaluation procedure

The feature extraction methods that were mentioned in the previous sections were implemented and compared in the experiment. All experiments were completed by using MATLAB numerical computation environment, BioSig library for biomedical signal processing and libSVM for multiclass SVM classification tasks. The ideas of using different EEG signal energy processing methods and CSP filtering (with initial code implementation) were acquired from the earlier work of Piotr Szachewicz [42]. The classifiers were trained and validated by using tenfold cross-validation. The default parameters were used for LDA and QDA classifiers as provided by MATLAB package. The grid search method was used for SVM RBF (radial basis function) gamma and cost parameter optimization [43]. The used values were: $C = 10$, $\gamma = 0.25$.

3.6.2. BCI IV 2a dataset

A BCI signal database "2a" [44] from the BCI IV competition held in 2008 was used for classifier training and testing. At the time of writing this dissertation, 4-class motor-imagery test dataset was the only one known data that was freely available online.

The analysed experiment dataset consisted of 22 channels that EEG signal recorded at 250 Hz sample rate for 9 healthy test subjects (total 288 motor imagery trials per subject). The signal stored in the dataset was already additionally preprocessed by using a bandpass-filter between 0.5 Hz and 100 Hz, and 50 Hz notch filter was enabled to suppress power line noise.

The EEG signals in the dataset were recorded by using a cue-paced (synchronous) mode of operation in two sessions on different days for each subject. During each session, the test subjects were asked to imagine movement of one out of four different motoric motions (left hand, right hand, feet, tongue) for 3 seconds. Each of the trials (Figure 13) in the dataset started with an audible signal (beep), followed by visual information (cue) to perform one of the mental tasks and a short break after the mental task.

Before using the data in all experiments described in this work, additional artifact correction of EEG data was done to discard invalid trials as mentioned in [44] by Brunner et al. The corrected EEG data were bandpass-filtered between 7 Hz and 30 Hz in order to cover μ and β brain rhythm frequencies and further used for the feature extraction in this work.

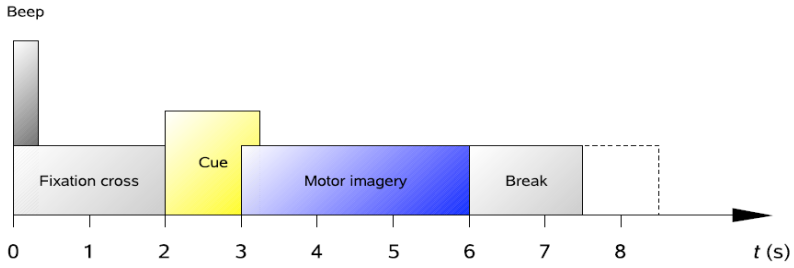


Figure 13. Single trial timing scheme

3.6.3. Accuracy calculation

The accuracy of the BCI data classification results was computed by calculating the Cohen's kappa [45] coefficient κ as given by Equation (18):

$$\kappa = \frac{p_0 - p_e}{1 - p_e}, \tag{18}$$

where p_0 is the classification accuracy, p_e is the hypothetical accuracy of a random classifier for the data (p_e = 1/4 for four-class problems).

3.7. Results

All experiment results are given in Table 3 and Table 4. As it can be seen from the tables, by using simple EEG features, the mean kappa for the LDA method scores the highest accuracy (and average) almost everywhere. This could be an indication of good linearization or good linear separation of the EEG features. Still, a combination of CSP filtering and SVM classification beats the LDA for single feature type, but the difference is negligible. A maximum average kappa of 0.495 was reached in tests (i.e., accuracy of 62%), which is still far from 90% accuracy achieved in other [46] work that uses more advanced techniques. Since Channel difference method does not support CSP filtering, N/A values are presented.

Table 3. Classification results using CSP filtering (kappa values)

Classifier	Ch.diff	Teager	TDP	Mean power	BP	C-Avrg
kNN	N/A	0.3890	0.4327	0.4249	0.3393	0.3965
LDA	**N/A**	**0.4437**	**0.4950**	**0.4638**	**0.4154**	**0.4545**
QDA	N/A	0.3868	0.4344	0.4259	0.3515	0.3997
SVM	N/A	0.3645	0.3835	0.4605	0.4075	0.4040
F-avrg	N/A	0.3960	0.4364	**0.4438**	0.3784	

Table 4. Classification results without CSP filtering (kappa values)

Classifier	Ch.diff	Teager	TDP	Mean power	BP	C-Avrg
kNN	0.3678	0.1538	0.2045	0.1867	0.1225	0.2071
LDA	0.4346	**0.4147**	**0.4134**	**0.4400**	**0.2487**	**0.3903**
QDA	0.3179	0.2864	0.2790	0.3185	0.1480	0.2700
SVM	**0.4372**	0.1553	0.1791	0.2950	0.1877	0.2509
F-avrg	**0.3894**	0.2525	0.2690	0.3101	0.1767	

The average feature performance (F-avrg) indicates that Mean power and TDP features are the best feature methods when using CSP filtering. However, Channel difference method achieves the best result when CSP filtering is not being done. Band power algorithm was the worst performer in tests. A graphical view of the same data is given further in Figure 14 where CSP filtering normalization influence can be seen better (left with CSP, right no CSP).

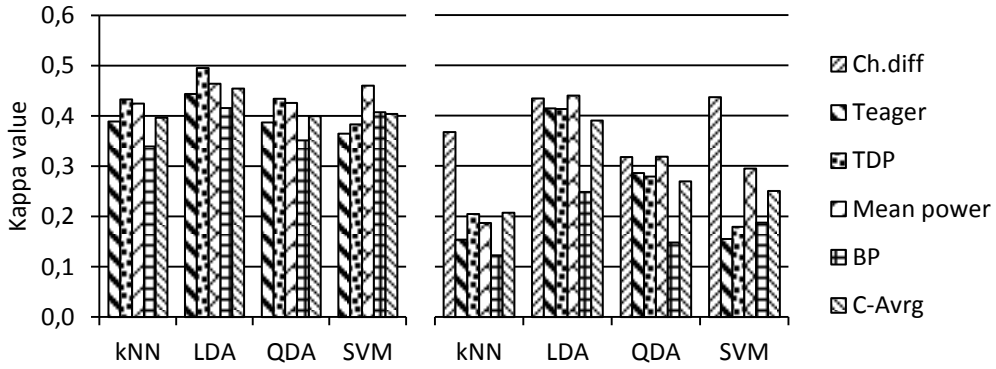


Figure 14. Average feature results for classifiers

It should be noted that the Channel difference method performance follows along the Band Power performance when CSP is used due to the obvious reason: the method needs to extract different signal frequency bands by using the BP algorithm. This can be seen in Figure 15 further. Since Channel difference method cannot use CSP filtering, one result bar is not shown.

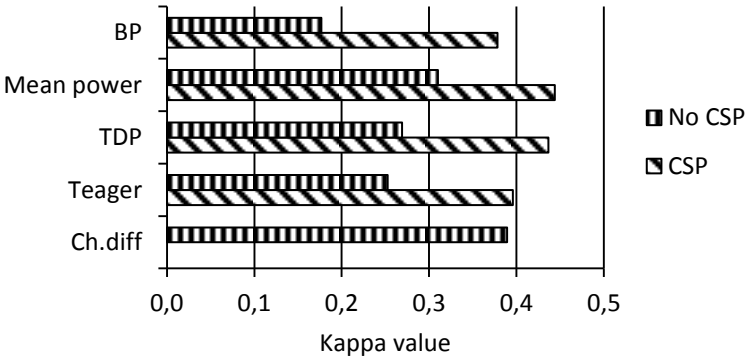


Figure 15. Average accuracy per feature type

The view showing method accuracy per subject as in Figure 16 (left no CSP, right with CSP) is quite important. Some subjects are resistant to the existing EEG methods; thus, the subject-specific (adaptive) techniques are required in order to achieve higher classification accuracy. The positive normalizing effect can be seen in the CSP case by giving greater average accuracy.

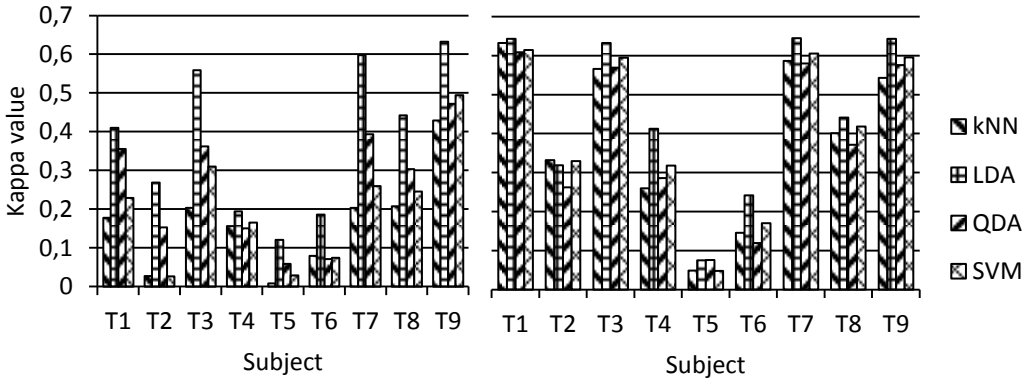


Figure 16. Average accuracy per subject

All experiment source code, test data and detailed results can be acquired from repository: https://github.com/tomazas/icist2015.

3.8. Conclusions

This chapter analysed multiple signal energy feature extraction methods and their usage for 4-class motor imagery BCI classification problem. After conducting experimental analysis, such conclusions can be made:

1. Mean power and/or TDP among the tested feature extraction methods show the best accuracy (0.4638 kappa ~ 59% and 0.4950 kappa ~ 62%, respectively), when doing EEG signal classification with CSP filtering.

2. A positive influence on the accuracy and test results were visible (10–15% increase) when the CSP filter was applied. This clearly states that CSP is effective and should be used for the EEG feature filtering.
3. The best classification performance was demonstrated by LDA among the tested classifiers (0.4545 kappa ~ 59%). Such results provide insight that LDA algorithm can be successfully used for the EEG signal classification if the EEG features can be linearly separable and/or EEG data is properly preprocessed.
4. A proposed Channel difference algorithm for signal feature extraction was able to achieve the best average classification result (0.3894 kappa ~ 54%) among the other tested feature extraction algorithms, when CSP filtering was not used (outperforming others by 6–16%). The ability to reach classification results close to CSP (0.4438 kappa ~ 58%) shows that the approach is effective.
5. Some subjects (T5, T6) show high resistance to current feature extraction methods; thus, the subject-specific (adaptive) techniques are needed in order to achieve higher classification accuracy.
6. Many improvement possibilities exist for future work, including the development of better feature extraction algorithms that are able to adapt to the subject specific EEG information. Moreover, other new classification techniques can be used for solving the motor-imagery problem.

4. DEEP LEARNING AND NEURAL NETWORKS[9]

4.1. Overview

It was seen in the previous chapter that simple, linear and other commonly used feature extraction and classification methods do not provide the necessary accuracy for 4-class MI BCI, and other still unexplored techniques or solutions are needed to solve the MI task.

A relatively new and perspective approach to the EEG data classification was found in the deep learning branch of machine learning. The convolutional neural network (CNN) is a novel animal visual cortex inspired method for image based classification that has not been widely used with EEG, let alone motor-imagery task. With the abilities to generalize/pool and self-learn the needed features in non-linear ways, it can provide benefits for the EEG classification. Since the EEG motor imagery task lacks accurate solutions, the CNN could be the new perspective way to look deeper into the same problem. Regarding its novelty and success in other fields, it was chosen as the main tool for four-class EEG motor imagery problem analysis in this work.

By using CNN for classification, subtle fine tuning is required to receive the best results. This involves selecting a proper neural network architecture, feature method and feature map size. These nuances and their effect on classification performance are further analysed and discussed in this work.

Furthermore, the feature extraction and feature map (image) generation methods for classification are of great significance. In the simplest cases, the EEG signal and feature vector can be treated as a one-dimensional signal. In order to move to two-dimensional image classification, two dimensional features or feature transformation methods are required. Possible techniques for such a task are presented and discussed in the next sections.

4.2. Deep learning

In recent years, an increasing number of papers that use CNN for EEG classification task have been published. Multiple approaches have been proposed for solving motor imagery and other related problems. A short review of the common techniques is presented in the remainder of this section.

CNN was successfully used by Mirowski et al. [47] to predict epileptic seizures from the EEG. The authors have proposed to use four types of bivariate statistical properties of the EEG signal as features for classification. They argue that the commonly used univariate features (computed on each EEG channel separately) lack the required channel relationship information. Cross-correlation, non-linear interdependence, Lyapunov exponent and wavelet synchrony feature information

[9] This chapter uses parts of the article [1] UKTVERIS, T., and V. JUSAS. Application of Convolutional Neural Networks to Four-Class Motor Imagery Classification Problem. Information Technology And Control, vol. 46 (2), 2017, 260–273 (Sections 4.1–4.11).

was packed into 2D images for the classification. The prediction accuracy of 70% was achieved. Another work in the field of EEG analysis was dedicated to solving the SSVEP (Steady State Visually Evoked Potential) signal classification problem by Cecotti and Gräser [48], where a subject is introduced to the visual stimulation at a specific frequency. A four layer CNN network topology with a Fourier transform filter in second layer was tested. The selected architecture proved to achieve up to 97% classification accuracy. It was noted that the switch from time domain to frequency domain gave a positive effect on the classification performance. However, the introduced reliability rejection criteria for each class made the final solution less robust, produced a lot of sample rejections and gave average generalization. The different application of CNN to the SSVEP is described in a paper by Bevilacqua et al. [49]. The authors used a four layer network architecture with a hidden L2 Fast Fourier Transform (FFT) layer for frequency extraction. Due to the nature of the problem, the signal analysis was done in the frequency domain. Channels Pz, PO3, PO4, Oz (of 10-20 electrode system) were used to record EEG samples at 256 Hz within 2 second windows. The images of 4x512 elements were composed of filtered EEG data and used as an input for the CNN classifier. The network was trained for 1000 epochs. The mean accuracy of 88% was obtained by this method.

CNN capability of detecting P300 events from the EEG was showcased by Cecotti and Gräser [50] with the accuracy of 95%. The signal analysis was conducted separately in time and space domains. The images of 64x64 in size created from 64 channels of down-sampled EEG data were used for the classification. Seven different CNN models were verified. Additionally, the work employed a strategy to use vector based CNN kernels instead of matrix kernels in order to prevent mixing features related to space and time domains. A technique based on trained network first layer weight analysis was used to extract 8 most relevant electrodes for each subject.

Recently, CNN has been used by Manor et al. [51] to solve RSVP (Rapid Serial Visual Presentation) task (where a subject has to detect a target image within five possible categories). The authors introduced a spatio-temporal regularization penalty for the EEG classification to reduce network overfitting. The accuracy of 75% was reached with CNN architecture of three layers, having 64x1 convolutional, two pooling and two fully connected filters. The images of 64x64 (64 channels by 64 time samples) were used as an input for the network. The advantages of using neural network models against manually designed feature extraction algorithms were presented along with the criticism of the manual method for unclear and endless possibilities of combining different methods in an efficient way.

Various techniques directly related to the current motor imagery problem have been proposed over the years in literature. Qin and He in [52] describe the analysis of a two-class motor imagery problem. The authors proposed a technique to analyse the EEG in the frequency domain. A time-frequency distribution (TFD) images were constructed based on the complex Morlet wavelet decomposition for electrode pairs. The TFDs were subtracted from symmetrical channels to form weight matrices that were used to compute in weighted energy for the classification. A Laplacian filter was used for signal preprocessing. The average classification rate of 78% was

achieved for this method. Another approach based on the energy entropy preprocessing and Fisher class separability criteria was proposed in [53] by Xiao et al. The authors analysed a two-class motor imagery problem in time-frequency domain. Similar TFD distributions (spectrograms) were constructed from the EEG short-term Fourier transform (STFT) data. Three different classification methods were compared. The classification accuracy for the two-class problem was 85%. A more sophisticated approach for 3-class motor imagery analysis was done by Zhou et al. in [54]. The study proposed a new method to extract the MRICs (movement related independent components) and utilized ICA (Independent Component Analysis) spatial distribution patterns for such a task. Different ICA filter designs were tested. The ICA filter design was confirmed to be subject invariant. The classification accuracy of 62% was received.

A more recent study by Bai et al. [55] on 4-class motor imagery proposed a novel Wavelet-CSP (Common Spatial Patterns) with the ICA-filter method. The EEG artifacts were removed by using negative entropy-based ICA. The mean accuracy of 76% was achieved by using SVM (Support Vector Machine) classifier.

One of the latest works in the field of CNN and 4-class motor imagery is the paper by Yang et al. [56]. The authors proposed a frequency complementary feature map selection (FCMS) method. ACSP (Augmented CSP) feature filtering was used in their work. Two other feature selection methods, i.e., the random map selection (RMS) and selection of all feature maps (SFM), were analysed. FCMS was the best performing method due to its ability to limit the ACSP feature redundancy in different frequency bands. The CNN used 5 layer architecture with 5x5 filters (kernels). The work as well demonstrated that CNNs are capable of learning discriminant, deep structure features for the EEG classification without relying on the handcrafted features. The average classification accuracy that was achieved was 69%.

4.3. Convolutional Neural Networks (CNN)

Convolutional neural networks are biologically-inspired variants of MLPs (multi-layer perceptrons). They have been successfully used for character recognition in the past by LeCun et al. [57] and currently have gained interest from the researchers due to the performance capabilities. CNNs consist of one or more convolutional layers with the weights of the layer shared across the input. Multiple layers form a non-linear “filter” chain. The convolution is designed to handle 2D data, as opposed to the other neural networks that operate on 1D vectors. This ability makes the extracted features easier to view and interpret.

4.3.1. Feed-Forward Neural Network

A typical neural network function as presented by Vedaldi and Lenc [58] is defined as:

$$f(x) = f_n(\dots f_2(f_1(x; w_1); w_2) \dots, w_n), \tag{19}$$

$$f: \mathbb{R}^{M\times N\times K} \rightarrow \mathbb{R}^{M'\times N'\times K'}, \tag{20}$$

where $\boldsymbol{x} = (x_1, \ldots, x_k)$ is the network layer input (a $M\times N$ size image with K channels), $\boldsymbol{w} = (w_1, \ldots, w_n)$ is the vector of learned parameters (weights). The function f transforms a set of K images into a new set of K' images of size $M'\times N'$.

4.3.2. Convolution

A 3-dimensional convolution operation for *k'-th* filter (from a set of filters) can be expressed as (21):

$$y_{i'j'k'} = \sum_{ijk} w_{ijkk'} x_{i+i',j+j',k}, \tag{21}$$

where y is the output (image) sample of the convolution (for k'-th filter) at coordinates (i', j'), x_{ij} is *k-th* image input sample at coordinates $(i + i', j + j')$, w_{ij} are the *k-th* filter (set of weights), and i, j are the indices (coordinates) of the filter.

4.3.3. Pooling

The CNN concept of pooling is a form of non-linear down-sampling. Pooling partitions the input image into a set of non-overlapping rectangles and for each such sub-region, outputs the maximum or average value. In this way, it is possible to reduce the feature size (and computation) as required and provide translation invariance. Pooling function is given by (22):

$$y_{ijk} = max\{y_{i'j'k}: i \leq i' < i + p, j \leq j' < j + p\}, \tag{22}$$

where y is the output, p is padding, and i, j are the indices of the k filters.

4.3.4. Non-Linear Gating

Typical CNN non-linear filters use linear functions with a non-linear gating function, applied identically to each component of the feature map. The simplest function is the Rectified Linear Unit (ReLU). Such filter can be written as:

$$y_{ijk} = max\{0, x_{ijk}\}, \tag{23}$$

where y is the output, x are the input filters, and i, j are the indices of the k filters.

4.3.5. Normalization

Another important CNN building block is channel-wise normalization. This operator normalizes the vector over the feature channels at each spatial location in the input map x. The form of the normalization operator is the following:

$$y_{ijk'} = \frac{x_{ijk}}{\left(\kappa + \alpha \sum_{k \in G(k')} x_{ijk}^2\right)^{\beta}}, \tag{24}$$

where y is the output; κ, α, β are normalization parameters, $G(k) = [k-\lfloor \rho/2 \rfloor, k+\lceil \rho/2 \rceil] \cap \{1,2,\ldots,K\}$ is a group of ρ consecutive feature channels in the input map.

4.3.6. Softmax

The operation computes the softmax operator across the feature channels x_{ijk}, and, in a convolutional manner, at all spatial locations $t = \overline{1, D}$. It is a combination of an activation function (exponential) and a normalization operator:

$$y_{ijk} = \frac{e^{x_{ijk}}}{\sum_{t=1}^{D} e^{x_{ijt}}} \tag{25}$$

where y is the output of the softmax operation, and i, j are the indices of the k filters.

4.3.7. Need for Common Spatial Patterns (CSP)

The CSP method (more thoroughly described in section 3.4) has been selected for the EEG signal filtering, since it is a widely adopted signal preprocessing method and has other interesting features as shown by Naeem et al. [59]: it decomposes the raw EEG into subcomponents (spatial patterns) having maximum differences in variance. Furthermore, Wang et al. in [60] concluded that this technique allows better feature separation in feature space and more accurate signal classification. Moreover, the property of CSP to decrease the feature dimensionality is very suitable for the EEG data complexity reduction. It has been shown by Uktveris and Jusas in [61] and other works that this method gives a substantial EEG signal classification performance increase; thus, it is a highly recommended filtering method.

4.4. Feature extraction methods

A multitude of EEG feature extraction methods have been studied by Uktveris and Jusas in [61] and other literature. Their output usually is a one dimensional feature vector that can be used for classification. The ability to adapt the algorithms for two-dimensional CNN has not been thoroughly analysed. It is also important to know if the adapted methods can give similar or better results when applied in 2D for CNN. Thus, a review of the most common feature extraction techniques and their implementations for the CNN is presented in this work. A short description of the EEG feature methods that were tested and analysed in this work is given next.

4.4.1. Mean channel energy (MCE)

The energy of each *i*-th EEG channel x_i is computed as the mean of *N* squared time domain samples (26). The result is then transformed by using a Box-Cox [62] transformation (i.e., logarithm) in order to make the features more normally distributed, and, finally, the resulting values are combined into a feature vector:

$$y_i = log\left(\frac{1}{N}\sum_{k=1}^{N} x_i[k]^2\right), i = \overline{1,n}\,. \tag{26}$$

4.4.2. Channel variance (CV)

The variance for each *i*-th EEG channel is the second moment of the signal x_i computed about its mean $\overline{x_i}$. The result is normalized by using Box-Cox for the final feature vector:

$$y_i = log\left(\frac{1}{N}\sum_{k=1}^{N} (x_i[k] - \overline{x_i})^2\right), i = \overline{1,n}\,. \tag{27}$$

An example of a feature map generated by using this technique is given in Figure 17. The EEG trial data from the dataset described in section 4.9.1 was used to make the image.

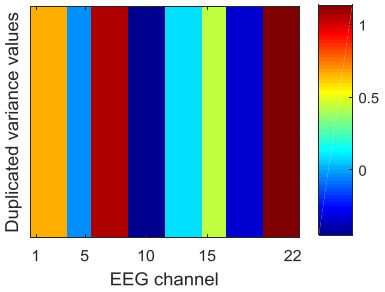


Figure 17. Feature map generated with CV method

4.4.3. Mean window energy (MWE)

This technique computes (28) the mean signal energy of *N* windows of size *W* = *s*/*N* for each *i*-th EEG channel (where *s* is the EEG channel sample count). The resulting coefficients are Box-Cox transformed (28) to form the final map:

$$H_{ij} = log\left[\frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{W}\sum_{k=0}^{W} x_i[k]^2\right)\right], N = \overline{1,p}\,. \tag{28}$$

The maximum window count in the experiments was selected as $p = n$ (where n is the EEG channel count) in order to form rectangle feature maps.

4.4.4. Principal Component Analysis (PCA)

PCA is a filtering technique that decomposes an input signal into the main components by using orthogonal transformations. Wang et al. showed in [63] that it could be used to suppress artifacts and noise in the EEG signal as well. The decomposition (29) is carried out multiple times: initially, to determine the principal components, secondly, to suppress noisy components at the decomposition levels 1–3:

$$\hat{x}_i = |PCA(x_i)|, \qquad i = \overline{1,n}\,, \tag{29}$$

where n is the number of EEG channels. The final feature vector consists of the filtered $\hat{x}_i$ EEG mean energy elements (30) that were normalized via Box-Cox:

$$y_i = log\left(\frac{1}{N}\sum_{k=1}^{N}\hat{x}_i[k]^2\right), i = \overline{1,n}. \tag{30}$$

The multi-resolution of 5 levels with Daubechies least-asymmetric wavelet (4 vanishing moments) was used for the decomposition in this work.

4.4.5. Mean band power (BP)

The algorithm is already described in section 3.3.2; thus, it will not be repeated. By using the obtained result $\bar{p}_{ij}$ from the initial BP step, the required mean power values y_{ij} are computed via (31):

$$y_{ij} = \frac{1}{N}\sum_{k=1}^{N}\bar{p}_{ij}[k], i = \overline{1,3}, \tag{31}$$

where i is the band number and j is the EEG channel of length N. The computed result is used as the feature vector components.

4.4.6. Channel FFT energy (CFFT)

As analysed by Cecotti and Gräser in [48], this method employs the Fast Fourier Transform (FFT) for computing i-th EEG channel signal x_i energy estimation in the frequency domain. The FFT result is squared, and the sum of all elements is computed:

$$y_i = log\left(\sum_{k=1}^{N}FFT(x_i)^2\right), i = \overline{1,n}, \tag{32}$$

where *n* is the number of the EEG channels. The final feature vector components y_i are formed after the Box-Cox normalization transformation is applied.

4.4.7. Channel Discrete Cosine Transform (DCT)

The signal energy concentration can be estimated via DCT as shown by Birvinskas et al. in [64]. The sum of *N* squared DCT coefficients of each *i*-th EEG channel signal x_i forms the feature vector components of this method (33). Features y_i are normalized by using Box-Cox transform:

$$y_i = log\left(\sum_{k=1}^{N} DCT(x_i)^2\right), i = \overline{1,n}\,. \tag{33}$$

4.4.8. Time Domain Parameters (TDP)

TDP has already been described in section 3.3.3; thus, it will not be repeated. The components y_i of the final feature vector have been computed by using (34):

$$y_i = \frac{1}{N}\sum_{k=1}^{N} \bar{p}_i[k]\,, i = \overline{1,n}\,, \tag{34}$$

where *n* is the number of EEG channels, and $\bar{p}_i$ are the smoothed derivative values.

4.4.9. Teager-Kaiser Energy Operator (TKEO)

TKEO has already been described in section 3.3.4; thus, it will not be repeated. The components of the final feature vector were computed by using (35).

$$y_i = log\left(\frac{1}{N}\sum_{k=1}^{N} \Psi_i[k]\right), i = \overline{1,n,} \tag{35}$$

where *n* is the number of EEG channels, and Ψ_i are TKEO energy vector for *i-th* EEG channel.

4.4.10. FFT energy map (FFTEM)

This method generates a 2D feature map from EEG by using FFT. Each *i-th* EEG channel signal *x* is transformed into the frequency domain and forms a single row in the feature map as shown in (36). Full signal window was used to gain a global energy view(state) as opposed to the work by Hu et al. [65], which used short-term FFT windows to capture the energy dynamics. The full view has the advantage of incorporating the entire frequency domain without losing data at the cost of additional noise that does not necessarily belong to the valuable MI EEG.

$$H_i = |FFT(x_i)|, \qquad i = \overline{1,n}. \tag{36}$$

The computed map H was scaled to required feature map size for the CNN classification. Figure 18 shows an example result map that was generated with this method. EEG trial data from the dataset described in section 4.9.1 was used to make the image.

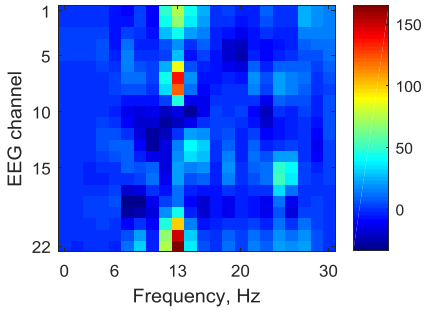


Figure 18. FFT energy map example

4.4.11. Complex Morlet Wavelet Transform (CWT)

CWT is a time-frequency analysis method used by Le Van Quyen et al. in [67] for obtaining wavelet coefficient maps W_x (37) at specific frequencies f and time τ, which were analysed more by Qin and He in [52]:

$$W_x(\tau, f) = \int_{-\infty}^{+\infty} x(u)\, \Psi_{\tau,f}(u) du\,. \tag{37}$$

All EEG channel signals combined as one *x(t)* signal were convolved with a number of different frequency Morlet Wavelets (38), where $\sigma = n/2\pi f$ and n is the number of wavelet cycles (38):

$$\Psi_{\tau,f}(u) = \sqrt{f} e^{i2\pi f(u-\tau)} e^{-\frac{(u-\tau)^2}{2\sigma^2}}. \tag{38}$$

Finally, the $W_x(\tau, f)$ was decomposed back to the initial EEG dimensions, and the mean energy coefficients of each channel formed a single row (39) in the feature map:

$$H_i = \frac{1}{N} \sum_{k=1}^{N} |W_{xk}(\tau, f_i)|\,, \qquad i = \overline{1,n}\,, \tag{39}$$

where n is the frequency/row count and N is the W_x row count. In this work, 22 different frequencies were used of [0; 30] Hz range band along with the wavelet cycles of range [0.5; 5]. An example output of this method is given in Figure 19. The EEG trial data from the dataset described in section 4.9.1 was used to make the image.

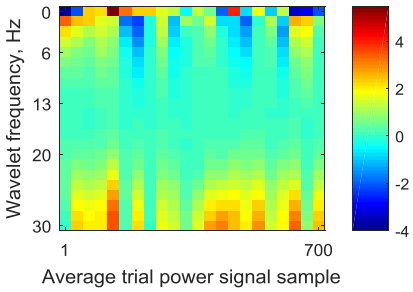


Figure 19. Feature map generated with CWT

4.4.12. Raw signal features (RAW)

RAW is a baseline method that uses the initially preprocessed EEG signal as values for the feature map. Each *i-th* EEG signal *x* channel directly maps to the feature map H_i *i-th* row as shown in (40):

$$H_i = x_i, \qquad i = \overline{1,n}\,. \tag{40}$$

If necessary, the resulting feature map is scaled (upsampled) to the required image size for the CNN training.

4.4.13. Signal energy map (SEM)

This method is using raw EEG signal energy values for the feature map generation. The Box-Cox normalized energy of each *i-th* EEG signal *x* channel is computed, and the resulting vector is directly mapped to the feature map H_i *i-th* row as shown in (41):

$$H_i = log\, x_i^2\,, \qquad i = \overline{1,n}\,. \tag{41}$$

If necessary, the resulting feature map is scaled (upsampled) to the required image size for the CNN training. An example map generated with this method is given in Figure 20. The EEG trial data from the dataset described in section 4.9.1 was used to make the image.

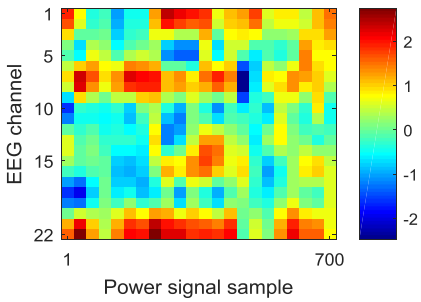


Figure 20. Feature map generated with SEM

4.5. CNN architecture selection

Choosing the correct network architecture for the problem gives a greater probability of getting better classification results. CNN supports serially connected layers. Due to the large number of different layer types, it is not trivial to find an optimal chain that closely matches the given problem.

The tests for 11 different CNN architectures were completed. Starting from the simplest and ending with the more complex ones, the tested architecture configurations in a simplified notation are given in Table 5. The used notation is explained in Table 6. A visual example of CNN architecture ICRPFSO can be seen in Figure 21.

The dataset described in section 4.9.1 was used for the architecture evaluation. The fixed test parameters for CNN were the following: image size 44x44px, initial learning rate 0.01, momentum 0.01, epochs 500, batchsize 128, features algorithm MCE.

Table 5. List of evaluated CNN architectures

#	CNN configuration	Notes
1	IC(4)RPFSO	4 filters
2	IC(4)RP(4)FSO	stride 4
3	IC(8)RPFSO	8 filters
4	ICRPFSO	
5	IC(32)RPFSO	32 filters
6	IC(64)RPFSO	64 filters
7	ICRPCRPFSO	
8	ICRFSO	
9	ICFSO	
10	IC(7x1)RC(1x7)RPFSO	Non-rect filters
11	IC(1x7)RPC(7x1)RPFSO	Non-rect filters

Table 6. CNN layer symbolic notation

Notation	Description (default parameters)
I	input layer of size (44x44x1)
C	convolutional layer (7x7, 16 filters)
R	ReLU layer
P	max pooling layer (2x2, stride 2)
F	fully connected layer (4 classes)
S	softmax layer
O	classification (output) layer

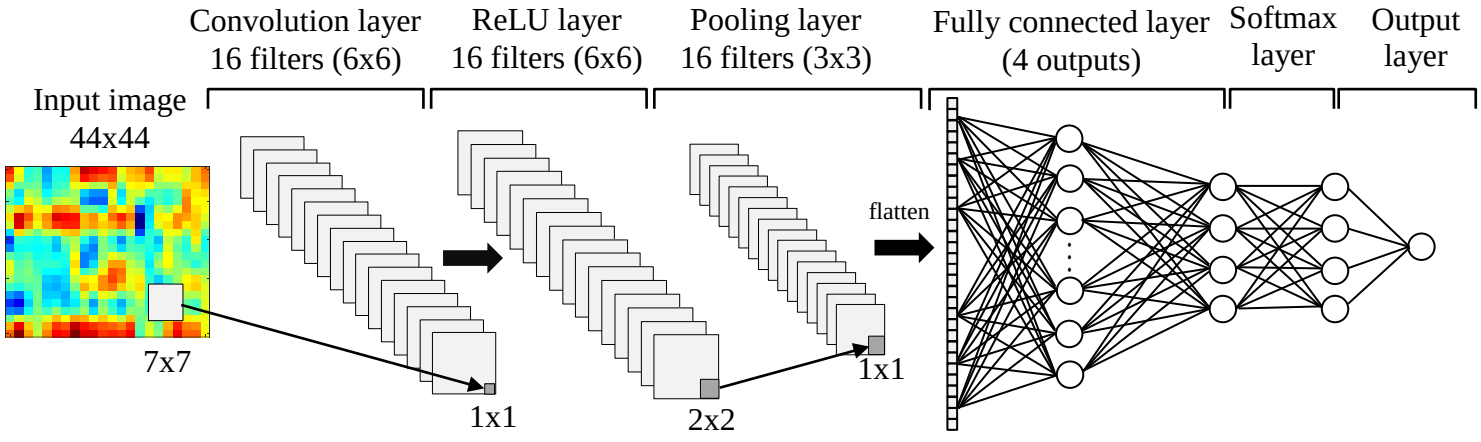


Figure 21. Example view of architecture ICRPFSO

The evaluation results are shown in Figure 22. It can be noted that the testing accuracy is around ~ 65% between the most of the configurations. However, the training accuracy displays a more dynamic profile from 50% to 80%. In this case, the CNN configuration with the least amount of computational-processing resources (i.e., the simplest) should be selected as optimal: 1, 2, 4 or 10.

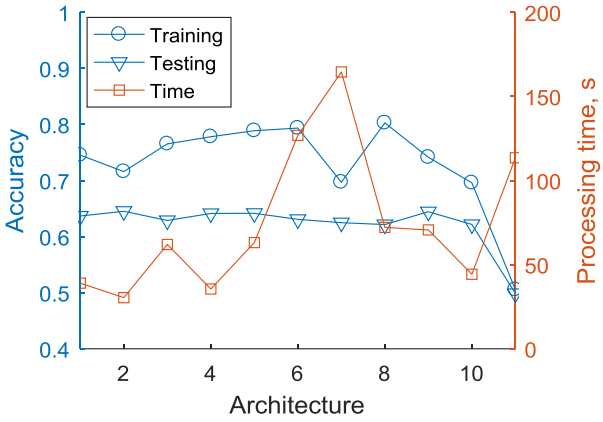


Figure 22. CNN architecture evaluation

4.6. CNN parameter tuning

CNNs are more complex, since they have more hyper-parameters than a standard multi-layer perceptron (MLP). However, the usual learning rates and regularization constants still apply. CNN training parameters, initial learning rate, momentum, batch size and the number of epochs must be tuned for the best performance. Since a 4D parameter grid based search is too resource intensive, a parameter range scanning approach was carried out to find the best accuracy of the given parameter value. The same initial CNN parameters as specified in section 4.5 were taken as base and changed during the tuning operation. The tuning process is described next.

The momentum value denotes the contribution for the next gradient value from the previous iteration in Stochastic Gradient Descent (SGD) method. Larger parameter values decrease the effectiveness of faster learning as shown in Figure 23. In tests, the values above 0.6 push the CNN to overfitting and thus decrease the generalization and testing accuracy. The value of zero for momentum is not recommended, since that invokes a loss of historical gradient learning information.

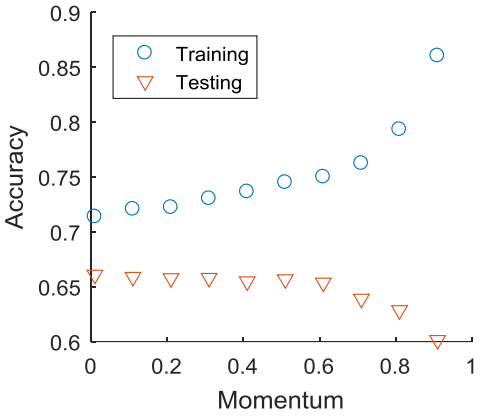


Figure 23. Momentum evaluation

The optimal number of training epochs ensures that the network learns and generalizes the provided features. The excessive epochs deteriorate the testing accuracy, since the network is overfitting. Figure 24 shows that the optimal count for training is 400–500 epochs.

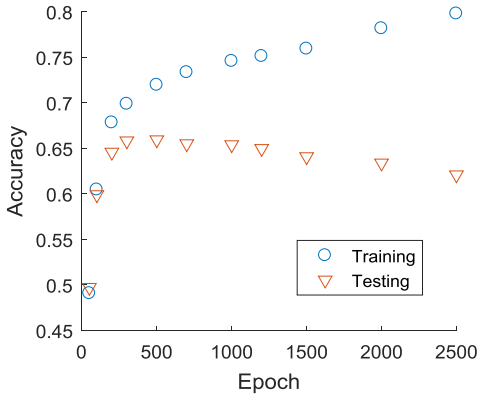


Figure 24. Epoch count evaluation

The batch size is the image count that is used for the single epoch training. It has a direct effect on the network learning quality as shown in Figure 25. The maximum batch size is the number of the total images, e.g., N = 288 in the experiments. The values lower than N/4 prevent the network from the fully maximizing learning efficiency: greater values only increase the computational costs at the price of no change in the testing accuracy.

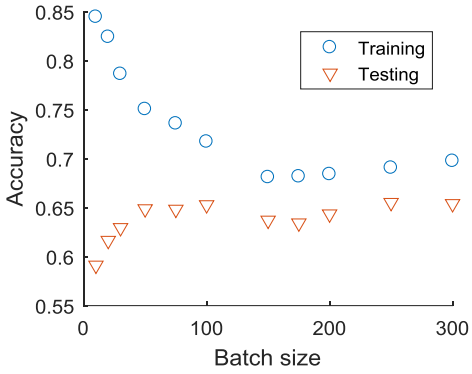


Figure 25. Batch size evaluation

The initial learning rate must be adopted for each problem. The experiments show that the value should not be bigger than 0.1, while the network testing accuracy peak is achieved with the values close to 0.01 as shown in Figure 26. Lower values allow learning fine grained features, while large ones have the tendency to overfit the network.

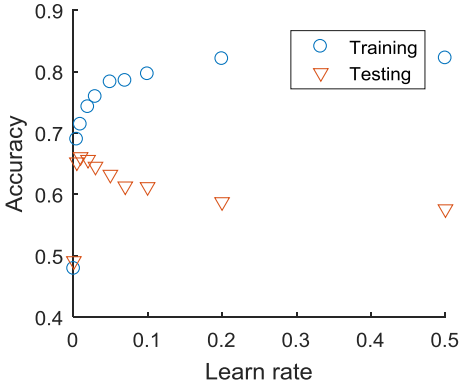


Figure 26. Initial learning rate evaluation

4.7. Feature map and filter dimensions

The problem is to find the right level of granularity in order to create data abstractions at the proper scale given a particular dataset. Different feature maps and filter sizes were analysed for the motor imagery problem. The dimensions from 8x8 to 64x64 of feature maps were tested. The test results are given in Figure 27.

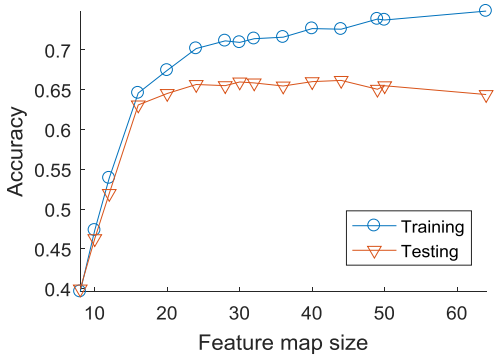


Figure 27. Feature map size evaluation

The plot shows that the optimal feature map size is 24x24 with the accuracy of 65%, even though a more accurate solution of 66% exists at size 44x44. Choosing a smaller size feature map ensures faster computation and processing speeds. Moreover, it should be noted that the accuracy convergence is reached when the feature map size is at least twice (15x15) the size as the convolution layer filter size (7x7 in the experiment). When the optimal size is reached, the further increase in dimension only introduces extra computational costs.

Convolution layer filter size limits the learning granularity by encompassing fixed size feature map regions. Ten different filter sizes were tested in the range [2; 11] for 22x22 feature maps. The test results are displayed in Figure 28.

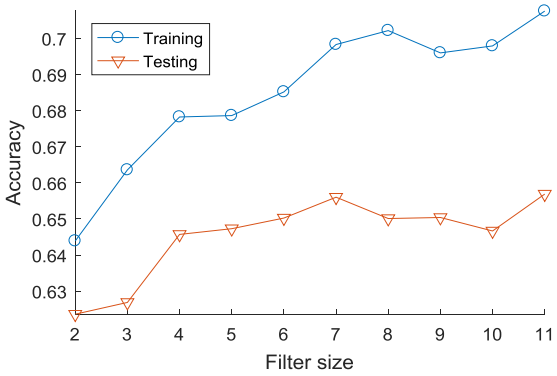


Figure 28. Convolution layer filter size evaluation

The optimal filter size, which gives the highest accuracy, is 7x7 and 11x11. Choosing the smaller filter size ensures faster processing speeds. The filters of size 2x2 and 3x3 exhibit too few weights to learn the details of the provided data fully.

4.8. Feature map generation

Many feature extraction methods form a single one-dimensional (1D) vector of coefficients known as the feature vector. A problem arises, since the CNNs are

designed to process only two-dimensional data (2D images). An alternative is to interpret the one-dimensional signal as a 2D single row image of dimensions *WxH*, where *W* is the width and *H* is the height (*H* = 1). However, the negative aspect of this approach is that only a single row CNN filters/kernels will be usable, and no row interdependence of the data will be taken into account.

Thus, a solution needs to be implemented to overcome this problem. This work reviews and proposes multiple feature vector adaptation techniques and algorithm, which allows transforming 1D feature vector data into 2D data that is viable for the use with CNNs. The adaptation step fits into the main EEG data classification chain as shown in Figure 29.

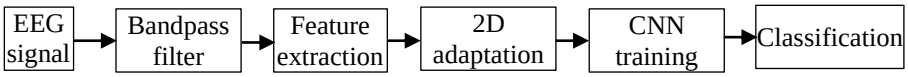


Figure 29. Adaptation in the EEG classification chain

Since CNNs typically take relatively small size (NxN) square/rectangular images (e.g., 7x7px, 22x22px, 44x44px, etc.) for the classification, there can be three possible outcomes when the size *N* is compared to the feature vector length *L*. Let's denote this comparison ratio as FVLIR (feature vector length and image ratio). The ratio definition is given in (42), and a more categorized view can be seen in Table 7.

$$FVLIR = {}^{L}/_{N}, \tag{42}$$

where *N* is the feature map (image) size, *L* is the feature vector length.

Table 7. Feature vector adaptation techniques

FVLIR	Technique	Options
< 1 (too short)	1. Duplication 2. Upsampling (interpolation/expansion) 3. 2D mapping	horizontal/vertical multiplier mapping function
1 (fits)	1. Direct packing 2. 2D mapping	horizontal/vertical mapping function
> 1 (too long)	1. Cutoff 2. Wrapping 3. Downsampling (reduction) 4. 2D mapping	length horizontal/vertical step size mapping function

Based on the FVLIR value (i.e., *L* and *N* comparison result), the adaptation technique can be selected. The proposed techniques will be detailed in the next sections.

4.8.1. Feature duplication

This approach is only viable for the feature vectors that are too short to be packed directly into the image ($L < N$). An example (Figure 30) of such transformation is to duplicate the feature vector in both directions to fill the feature map space.

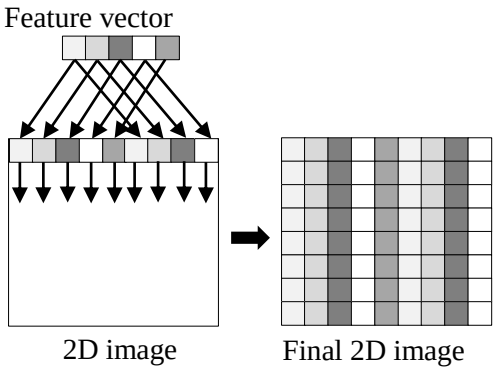


Figure 30. Horizontal and vertical vector duplication

Some additional filtering can be applied to the new repeated copies. An example of such feature map is given in Figure 31.

Figure 31. Feature map generated via vector duplication

4.8.2. Feature upsampling

Upsampling or interpolation allows to expand the feature vector and should be used when the feature vector length is less than the image width ($L < N$). This method will create additional n (multiplier) samples by interpolating between the existing data points (Figure 32).

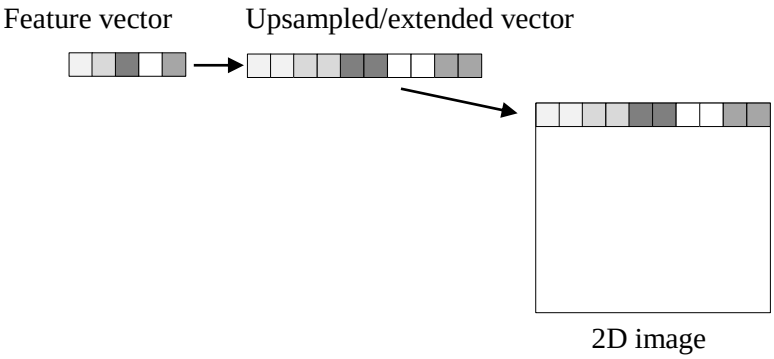


Figure 32. Upsampling feature vector

In the process, this will extend/lengthen the signal and shift it to the lower frequencies in the frequency domain because of the extended/lengthened signal oscillations.

If only one vector per single EEG trial is generated, the duplication technique will need to be used to replicate the extended vector vertically. If each row directly maps to each EEG channel; then, only direct packing will be necessary.

4.8.3. Direct packing

Direct packing technique places the feature vector directly into the image because the feature vector length and image size matches. Each vector is added to the row or column of the image (depending on the packing direction). The example of this approach can be seen in Figure 33.

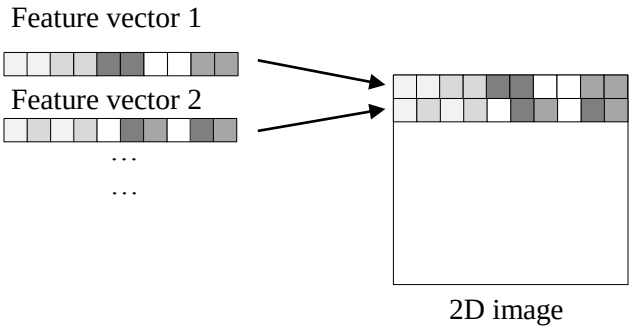


Figure 33. Horizontal direct feature vector packing

4.8.4. Feature cutoff

This technique can be used when the feature vector length is bigger than the image size ($L > N$). The vector tail that does not fit into the image ($> N$) is cut off (Figure 34), and only head forms a row in the image. The duplication or direct

packing can be used further to complete the whole image. It is not recommended to do this for long feature vectors as a significant amount of information can be missed.

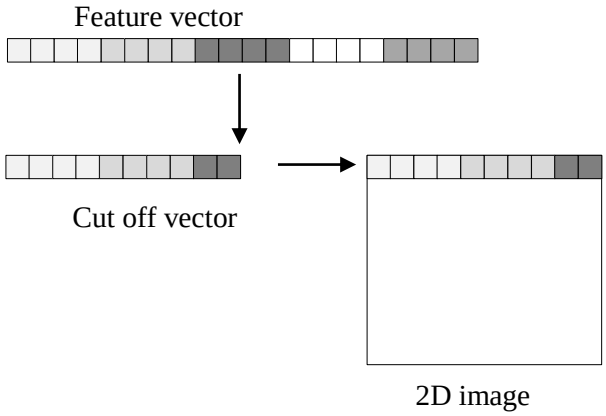


Figure 34. Feature vector cut-off

4.8.5. Wrapping

Wrapping packs a very long feature vector (Figure 35) into the image by chopping the vector into multiple (image size N) chunks. Each part forms a single row/column of the final image. The method should be employed when the feature vector length is close to the square of the image size ($L \approx N^2$) or larger.

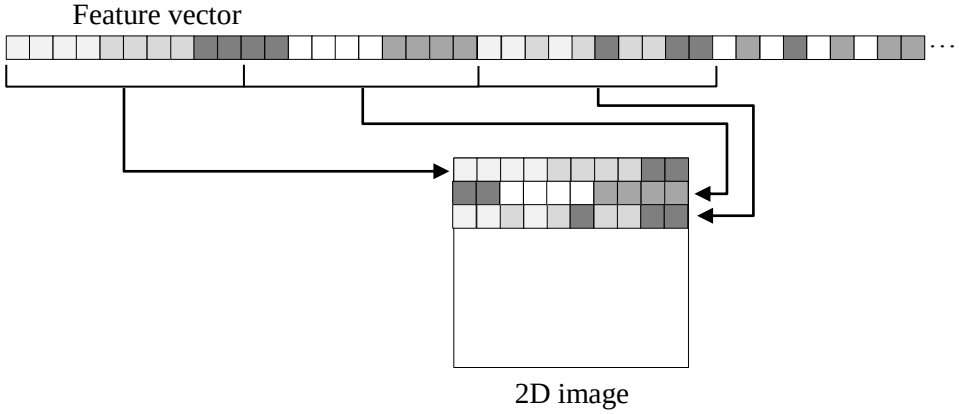


Figure 35. Wrapping feature vector

4.8.6. Downsampling

Downsampling is the opposite procedure of upsampling in which a feature vector longer than the image size ($L > N$) is shortened/reduced by removing every n-th (step) element from the initial vector (Figure 36):

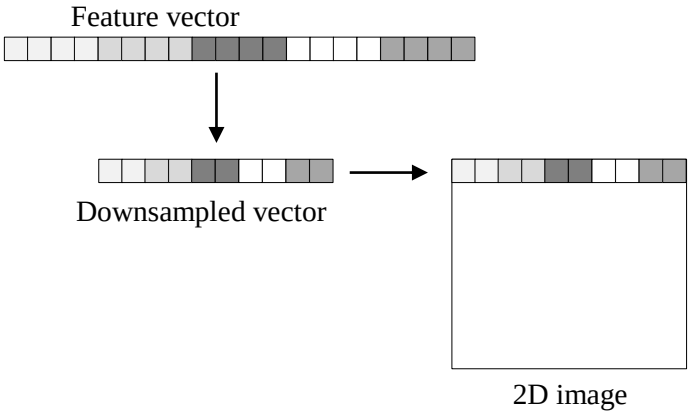


Figure 36. Downsampling feature vector

Once the vector size matches the image size ($L = W$), the direct packing or duplication will need to be done.

4.8.7. 2D mapping

In the mapping method, there is a need to select such a transformation H that allows converting 1D signal into 2D:

$$H: \mathbb{R} \to \mathbb{R}^2. \tag{43}$$

The example of such approach is to use function f that takes element pairs (x,y) from the feature vector and maps them to 2D Cartesian coordinate system as shown in Figure 37.

$$H: f(x_1, x_2) \to (x_1, x_2). \tag{44}$$

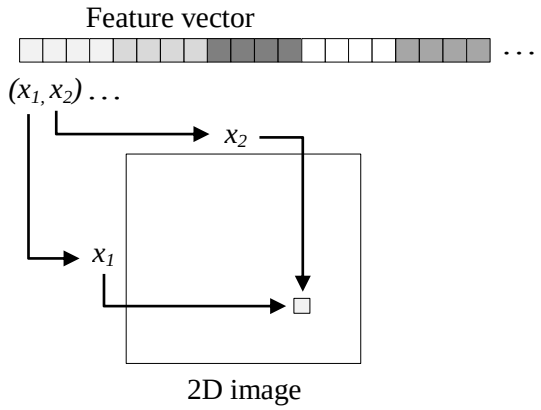


Figure 37. Feature vector element pair transformation to 2D map

In order to get a more meaningful feature representation on a 2D image, the feature vector element pairs can be constructed from the statistical measurements (e.g., x_1 = mean channel energy, x_2 = channel energy variance) and normalized in the range [-1; 1].

4.8.8. Adaptation algorithm

Given the feature vector length L and feature map/image size N, one can follow the algorithm depicted in Figure 38 to determine the recommended adaptation technique for the EEG trial feature vector transformation. The technique is selected in a way that it will try to preserve as much of the input data as possible and remove the least amount of features during 2D transformation.

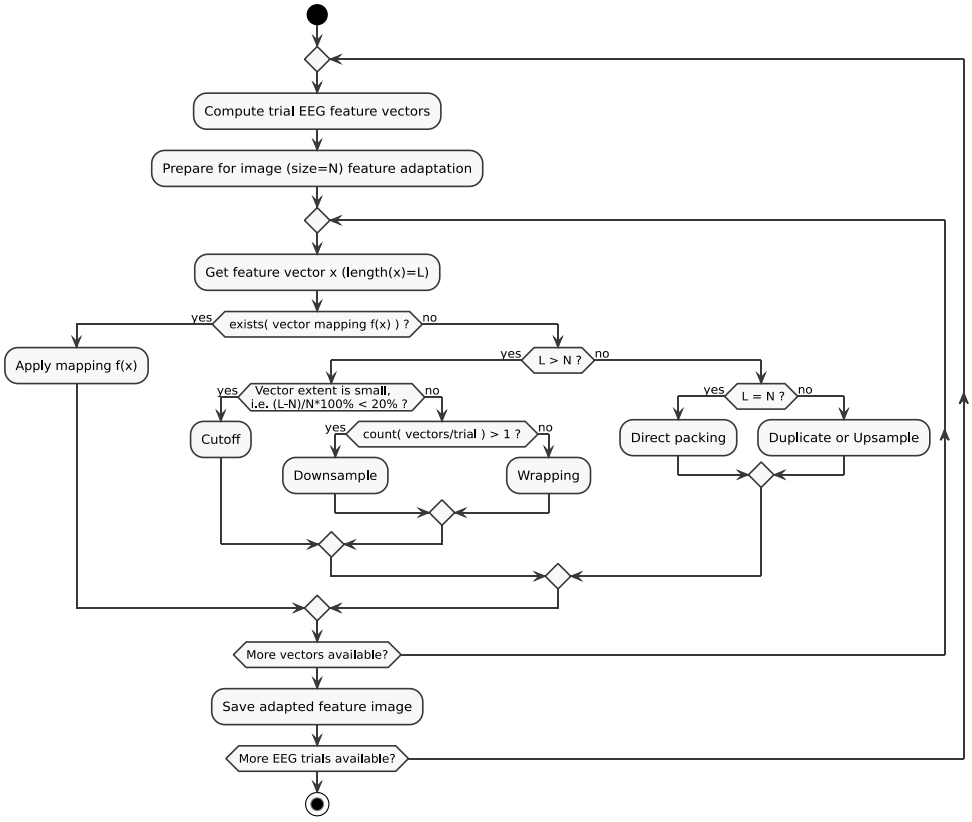


Figure 38. Feature vector adaptation algorithm

Based on the provided adaptation algorithm flow, the selected and used transformation techniques for feature extraction methods in experiments are shown in Table 8. The direct packing is used if the feature method generates 2D feature map or vector of the same size as image size N; otherwise, upsampling or downsampling is done for 2D data as well as duplication for 1D data. The default

vector length and image size baseline is set to match 22 EEG channels in order to be able to make square size training images easier for the CNN classifier.

Table 8. Adaptation method's proposed transformation techniques

Method	Features dimension (vector size L)	Determined 2D transformation		
		N > 22	N < 22	N = 22
MCE	1D (1x22)	duplication	cutoff/downsampling	direct packing
CV	1D (1x22)	duplication	cutoff/downsampling	direct packing
MWE	2D (22x22)	upsample	downsampling	direct packing
PCA	1D (1x22)	duplication	cutoff/downsampling	direct packing
BP	1D (1x22)	duplication	cutoff/downsampling	direct packing
CFFT	1D (1x22)	duplication	cutoff/downsampling	direct packing
DCT	2D (22x22)	upsample	downsampling	direct packing
TDP	1D (1x22)	duplication	cutoff/downsampling	direct packing
TKEO	1D (1x22)	duplication	cutoff/downsampling	direct packing
FFTEM	2D (22x22)	upsample	downsampling	direct packing
CWT	2D (22x22)	upsample	downsampling	direct packing
RAW	1D (1xSamples)	downsampling	downsampling	downsampling
SEM	1D (1xSamples)	downsampling	downsampling	downsampling

4.8.9. Effects of the feature map scaling

A baseline method and the simplest approach from all the feature extraction techniques is to classify raw EEG signal samples. The raw EEG data form factor of NxM, (where N is the number of channels, M is the number of samples, N<<M) restricts the direct use of it for the CNN feature images due to a large number of samples. Thus, it must be scaled down. Generally, a feature map of WxH size (where W is width and H is height) can be formed by down/up-scaling the raw EEG signal or extracted feature data. The technique of resizing can use bilinear or other type of filtering in order to prevent sharp data transitions, limit noise and smooth out the final feature map. An example of filtering that is applied to the raw EEG feature maps can be seen in Figure 39 (from left: nearest, bilinear, bicubic filtering).

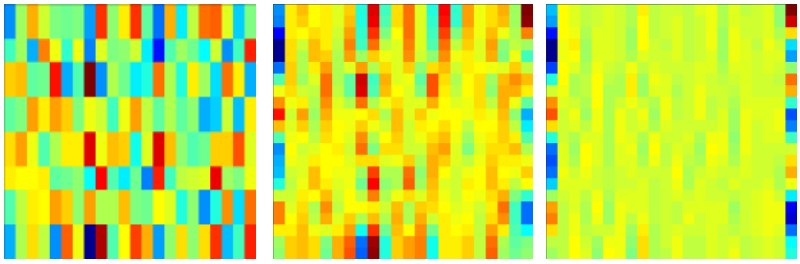


Figure 39. Example of 22x22 raw EEG feature maps

The initial testing results of the three different image filtering techniques for raw EEG signal classification are given in Table 9. The results show ~ 10% difference in classification accuracy when various filtering techniques are applied. It

can be seen that for raw EEG signal analysis, the nearest filtering method should be used in order to retain original signal details as much as possible. For other feature types, the effect could be the opposite.

Table 9. Raw EEG feature map resize filtering accuracy

Filter method	Training	Testing
Nearest	0.47 ± 0.14	0.43 ± 0.11
Bilinear	0.35 ± 0.11	0.33 ± 0.12
Bicubic	0.33 ± 0.10	0.32 ± 0.11

4.9. Experiments

The main purpose of the experimentation activities was to investigate the capabilities of the CNN classifier for four-class motor imagery classification problem as well as to analyse the influence of various CNN architectures, feature maps, filter sizes and other parameters to classification accuracy. The experiments were conducted in the analysis step (tuning the CNN network parameters) and the main motor imagery classification step (for each subject).

The experiment results were measured and evaluated by using normalized accuracy in the range [0; 1]. The CNN network parameters were tuned and verified before the final classification step. The tests were carried out by using ten-fold cross validation. Moreover, the ability of CNN to learn from feature data was validated visually by inspecting the learned filter/weight images.

The final classification results for each subject are provided in the results section further.

4.9.1. Dataset

The same BCI signal dataset 2a (contributed by Brunner et al. [44]) from the BCI IV competition held in 2008 and described earlier in Section 3.6.2 of this work was used for the classifier training and testing.

4.9.2. Details of implementation

The software code for the experiments was implemented in MATLAB 2016b/9.1 numerical computation environment. CNN is a new MATLAB functionality (starting from the 2016a/9.0 version), which uses GPU processor for parallel computations. Other alternatives for the convolutional neural networks exist, such as the open source MatConvNet library created by Vedaldi and Lenc [58]; however, due to the stability issues, the library was left as an option for future CNN evaluations. Parts of the open source BioSig library for biomedical signal processing and imaging were used in the EEG signal analysis.

CNN convolution layer's initial filter weights in all the tests were set to have a Gaussian distribution with a mean of 0 and a standard deviation of 0.01. The default for the initial bias was 0.

4.10. Results

The final classification results were obtained after the analysis, and CNN parameter fine-tuning step. A CNN with initial learning rate of 0.01, momentum of 0.1, batch size of 128, 200 epochs and architecture I(22x22)C(4x4,16)RPFSO was trained and tested for the final evaluation of all subjects. The results were verified by using 10-fold cross-validation scheme. The accuracies with their standard deviation values are displayed in Table 10, and a graphical view is given in Figure 40. From the results, it can be seen that the best performing approach (70% in training and 68% in testing) is the FFT energy map method. The second and third best methods in tests are the Channel variance (68%/61%) and Signal energy map (67%/61%) features. The lowest accuracy of (41%/31%) was achieved by the TDP feature method.

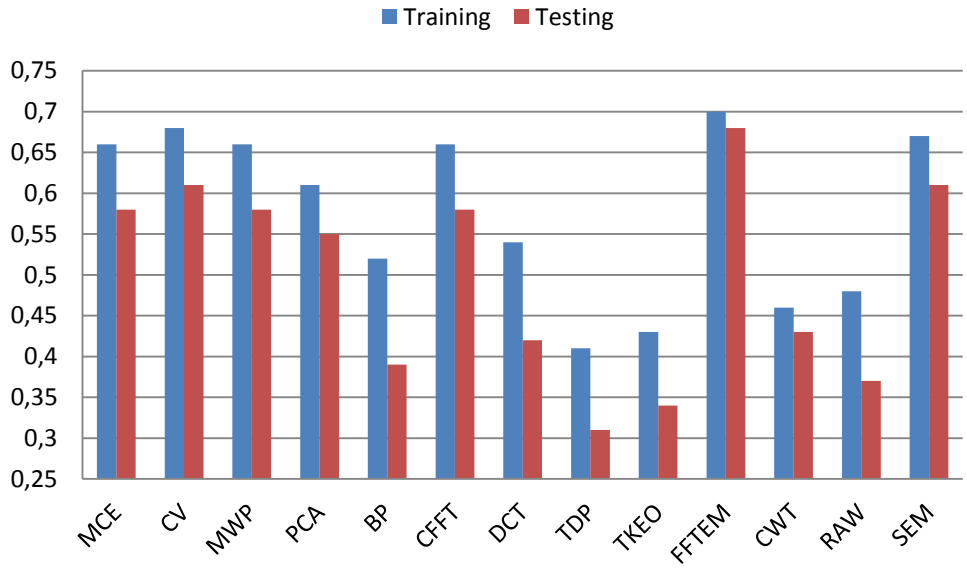


Figure 40. CNN classification results

Table 10. Classification results for the feature methods

Method	Training	Testing
MCE	0.66 ± 0.19	0.58 ± 0.20
CV	**0.68 ± 0.18**	**0.61 ± 0.22**
MWE	0.66 ± 0.19	0.58 ± 0.20
PCA	0.61 ± 0.16	0.55 ± 0.20
BP	0.52 ± 0.18	0.39 ± 0.11
CFFT	0.66 ± 0.19	0.58 ± 0.20
DCT	0.54 ± 0.17	0.42 ± 0.11
TDP	0.41 ± 0.11	0.31 ± 0.07
TKEO	0.43 ± 0.12	0.34 ± 0.05
FFTEM	**0.70 ± 0.18**	**0.68 ± 0.20**
CWT	0.46 ± 0.10	0.43 ± 0.13
RAW	0.48 ± 0.14	0.37 ± 0.11
SEM	**0.67 ± 0.18**	**0.61 ± 0.20**

The best FFTEM method accuracy has been compared to the other known methods (Table 11) found in the literature and against BCI IV finalists. The accuracy that was obtained is similar to the BCI IV competition winner and close to the other state-of-the-art methods.

Table 11. Comparison of the existing methods

Author	Features	Classifier	Kappa	Classification accuracy
Song et al. [69] (BCI IV 3rd place)	CSP	Ensemble multi-class	0.31	48%
Xygonakis et al. [24]	CSP	Ensemble model	0.46	59%
Guangquan et al. [69] (BCI IV 2nd place)	Log variance	LDA and Bayesian	0.52	64%
She et al. [25]	CSP	Non-linear Extreme Learning Machine	0.52	64%
Yang et al. [56] (BCI IV 1st place)	FBCSP	Naïve Bayes Parzen Window	0.57	68%
Proposed technique	**FFTEM**	**CNN**	**0.57**	**68%**
She et al. [11]	CSP	PPTSVM	0.63	72%
Ang et al. [68]	FBCSP	Naïve Bayes Parzen Window	0.66	75%

All experiment source code, test data and detailed results can be acquired from repository: https://github.com/tomazas/itc2017.

4.11. Conclusions

This chapter analysed Convolutional Neural Networks and their application to four-class motor-imagery problem. After an in-depth CNN analysis and parameter fine-tuning, promising results were achieved. After evaluating results, such conclusions were made:

1. The FFT energy map method demonstrated the best feature determination abilities and achieved 68% mean testing accuracy for all the BCI IV

competition 2a dataset subjects. The gained accuracy is slightly better than in the new techniques proposed by Tabar and Halici in [70] and similar to more complex state-of-the-art EEG analysis techniques by Yang et al. [56].

2. The introduced feature vector adaptation method allowed using feature extraction methods that produce 1D feature vectors for 2D feature map CNN classification. The implemented technique was successfully validated during the experiments.
3. The use of less complex feature extraction methods like FFT energy map shows high CNN method potential for the motor imagery EEG analysis and capability to discern complex brain signal oscillatory patterns.
4. It is enough to use non-complex CNN architectures (as ICRPFSO) to be able to achieve the best accuracy results, reduce computational resource usage and shorten the processing time.
5. CNN parameter fine-tuning is required to achieve higher accuracy results and efficient computational resource usage. A range based search approach is enough to tune parameter values without exceeding time and computational resource limits.
6. The best CNN filter size for the given MI problem was found to be 7x7 to learn the details of the provided features effectively. While the best feature map size has to be 24x24 pixels (and at least twice the size of the filter size). Larger dimensions introduce extra computational costs, while smaller produce too few weights.
7. CNN has proven to be a good choice for the EEG MI classification task. Further work can be continued in order to provide more efficient feature extraction methods adapted specifically to the CNNs and favouring processing speed and accuracy.

5. DESIGN OF EEG ACQUISITION SYSTEM[10]

5.1. Introduction

The increasing awareness of brain-computer interfaces (BCI) for brain signal analysis has sparked new interest in electroencephalogram (EEG) acquisition device development. Various rehabilitation [71], entertainment and even security [72] applications can be implemented by post-processing [73, 74, 75] such electrical signals recorded from the human scalp. However, the development of BCI is a challenging task due to the noisy and variable nature of the EEG signal itself. The lack of validation, design knowledge and analysis for such systems impede progress in this field. Even if there were adequate trials to use mobile devices for such a problem [76], professional high-quality and high-resolution analog front-ends are required to capture the non-stationary brain signals in microvolt ranges. With the introduction of dedicated EEG low-noise programmable analog-to-digital converters (ADCs), such as the ADS1298, such tasks can be achieved more easily. Professional and high-quality EEG capture systems are available from multiple vendors such as G.Tec, TMSi, etc. Due to their more than four thousand US dollar price (Table 12), these devices are not meant for general public or entry-level development and thus prevent wider BCI adoption and research. Furthermore, there is minimal knowledge of design or operational information on how these devices are actually validated and achieve their proclaimed specifications. Additionally, there are no compact EEG systems allowing the scaling and reconfiguration of hardware based on the problem requirements (up to 64 or more channels). Achieving this would help to manage and reduce complexity and minimize runtime costs.

Table 12. State-of-the-art brain-computer interface (BCI) systems

System	Sampling Speed, Hz	# of Channels	Accuracy	CPU	Electrodes	I/O	CMRR	Price €
g.tec [77] Nautilus	500	64	24-bit, <60 nV (LSB), <0.6 µV RMS	TI DSP	Active-dry/gel	Wireless 2.4 GHz/USB	>90 dB	>4.5 k
g.tec HIamp	38.4 k	256	24-bit, <60 nV (LSB), <0.5 µV RMS	TI DSP	Active-dry/gel	USB	>90 dB	>31 k
TMSi Mobita [78]	2000	32	24-bit, <24 nV	N/A	Passive dry	Wi-Fi IEEE 802.11 b/g	>100 dB	N/A
TMSi Porti [79]	2048	32	22-bit, <1 µV RMS	N/A	Active-shielding	Bluetooth/ optic fiber	>90 dB	N/A

[10] This chapter uses parts of the article [66] UKTVERIS, T., and V. JUSAS. Development of a Modular Board for EEG Signal Acquisition, Sensors (Basel, Switzerland), vol. 18,7 2140, 2018 (Sections 5.1–5.6).

TMSi Refa	2048	136	22-bit, <1 μV RMS	N/A	Active-shielding	Optic fiber	>90 dB	N/A

With respect to the previously mentioned problems, this work presents a new low-cost modular and vertically stackable development board that can be used for entry-level EEG signal acquisition. Furthermore, the proposed design allows the system to be easily scalable and adapted to various EEG tasks, while maintaining significant cost savings. Simple but effective validation methods are presented for the acquisition and overall design assessment.

The next sections of this dissertation give a more detailed review of the proposed solution. Section 2 reviews state-of-the-art designs and approaches found in the literature. Section 3 provides a system architecture view and discusses various technical decisions. The methods for system board evaluation and validation along with the experiments are described in Section 4. The review and comparison of results with other similar systems are discussed in Section 5. Final conclusions and directions for the future research are presented in the last Section 6.

5.2. Overview

More than a few papers exist that describe the developed prototypes of EEG acquisition systems. F. Pinho et al. [80] presented a computationally powerful, wearable system with 32 active dry electrodes (based on TLC272 precision op-amp) for long-term epileptic patient monitoring. The battery-powered design featured a 24-bit resolution analog-to-digital conversion unit ADS1299 that is capable of sampling up to 1 ksps (1k samples/s). The EEG data could be processed real-time on a dedicated 1 GHz ARM CPU or sent to a host PC over Wi-Fi 802.11 b/g for the analysis and post-processing. Even though the focus of the work was to create a standalone system with a higher performance CPU, the maximum battery life of 25 h was the main limitation while running under the maximum load. Since the device was not optimized for size, this required longer wires and use of active electrodes.

A similar approach was used by S. Feng [81] in designing their EEG acquisition system for solving a steady state, visually evoked potentials (SSVEP) problem. A 16-channel cape for a Beagle Bone Black development board (having an AM3358 ARM Cortex-A8 1 GHz CPU) has been developed with two ADS1299 ADCs and capable of sampling at the speed of 1 ksps. The authors claimed that their system was superior due to its provided embedded processing power and ability to work up to 12 h on two lithium batteries. However, while the produced cape consumed only 5% (101.2 mW) of the total required power under maximum load, the use of such system for portable battery powered applications is currently still a big challenge.

B. Senevirathna et al. [82] designed a low-cost 7-channel, small size and battery-powered EEG solution for long-term monitoring of schizophrenic patients. The board used a single ADS1299 ADC that was controlled by using SAM G55 microcontroller. The authors claimed their system captured the analog data at 250

Hz sample rate and sent it over Bluetooth by using 230.4 k baud. The power consumption of 69 mA was reported with all channels in active state. Similarly, T.T. Vo et al. [83] introduced a low-cost 8-channel EEG recording device for the BCI applications. Having an STM32F4 microcontroller, a single ADS1299, and capable of sending data over Bluetooth, the design was dedicated to favour small size and low power usage. A sampling speed of 250 Hz was used to record EEG via wet, gold-cup electrodes. Despite successful validation, both previously mentioned devices lack spatial resolution for the EEG, and the overall board expandability was not considered.

A new re-design for an ECG acquisition system featuring a 24-bit ADS1298 ADC was done by D. Campillo [84]. The author interfaced the 8-channel analog-to-digital converter to an MSP430F5529 microcontroller running at 12 MHz. The presented system board was capable of sampling at 500 Hz rate, the intrinsic channel noise (ICN) was 9 μV, and the common mode rejection ratio (CMRR) was 94 dB. The board was tested for more than 12 h of continuous use. The main limitation of the system for EEG use was the lack of channels for good spatial resolution.

M. Wild et al. [85] presented a tiny 4-channel in-ear proof of concept EEG acquisition device. Built upon OpenBCI project ideas, the authors designed a BCI board with ADS1299 ADC that was interfaced by using an Atmega328 microcontroller. The raw EEG data were sent over Bluetooth to remote a PC host for processing. Another 16-channel EEG recording device that is using dry electrodes has been developed by V. Nathan et al. [86] and tested with SSVEP, P300 speller, and motor imagery BCI tasks. The recorded raw EEG data were sent to the host PC via Bluetooth for final processing.

5.3. System architecture

This section gives a detailed overview of the main components along with the integration and communication mechanisms that were used to develop the system board.

5.3.1. Analog front-end

Designing a reliable, high-accuracy, precision analog front-end (AFE) is not a trivial task [87] that is why commercial, off-the-shelf solutions should be considered first. There exist multiple AFE devices in the market that are capable of discretizing the analog EEG signal. Since the main brain EEG oscillatory waves propagate in a low-frequency range of 0–40 Hz, a high-sampling performance AFE is not required. Thus, the main focus should be directed to the AFEs with a maximum number of supported channels, noise reduction capabilities and high acquisition resolution. E. Mastinu et al. [88] have compared two popular production grade AFEs, i.e., ADS1299 and RHA2216, and found that they provide similar results, although slightly better noise performances and higher myoelectric pattern recognition (MPR) accuracy was measured for ADS1299. D. Acharya et al. reviewed ADS1299

development board produced by Texas Instruments [89] for the EEG task. Based on the evaluation given in their paper, the ADS was recommended for the EEG acquisition due to low power use, low input referred noise (0.205 μVrms–6.5 μVrms) and the overall improvement over provided features in the same device segment.

In addition, the OpenBCI development board, which is popular among researchers [90] and entry-level enthusiasts, uses ADS1299 device for the AFE. According to M. Zieleniewska et al. who compared OpenBCI to the top-class EEG amplifier from TMSi [91], the signal quality was comparable to the commercial EEG amplifier and sufficient for research and advanced BCI applications, despite the board and electrode shielding problems.

Due to its extensive features, wide use in the industry and many applications, ADS1299 and alternative ADS1298 were selected as the AFE for the developed system. Free samples of ADS1298 were acquired from Texas Instruments.

ADS1298/9 is a device [92] for biopotential measurements and medical instrumentation (electrocardiogram (ECG), electromyogram (EMG) and EEG) with eight low-noise programmable gain amplifiers (PGAs) and eight high 24-bit resolution Delta-Sigma ADCs. The device has a self-test, temperature and lead-off detection mechanisms. Although similar and designed for the same application, the main differences between the devices are given in Table 13.

Table 13. Main differences of ADS1298/9 devices

Parameter	ADS1298	ADS1299
Sample rate (max), ksps	32	16
Input type	Differential Single-ended	Differential
Power consumption, mW	6	41
Min analog voltage, V	2.7	4.75
SNR, dB	112	121
Max programmable gain	12	24
CMRR, dB	-115	-110

5.3.2. Host microprocessor

There are no general solutions for choosing the host processor for interfacing AFE. In the literature [80–91], depending on the use case and required computational performance, the host processor ranges from microcontrollers to embedded microprocessors with 1 GHz or higher frequency. It is inappropriate to choose high-performing CPU for such battery-powered EEG recording devices. All intensive computations, such as machine learning should be carried out remotely on a host PC. The CPU in this work was chosen in order to reach the required maximum analog front-end sampling speed of 1 ksps (1k samples/s) and the communication with wireless device modules speed.

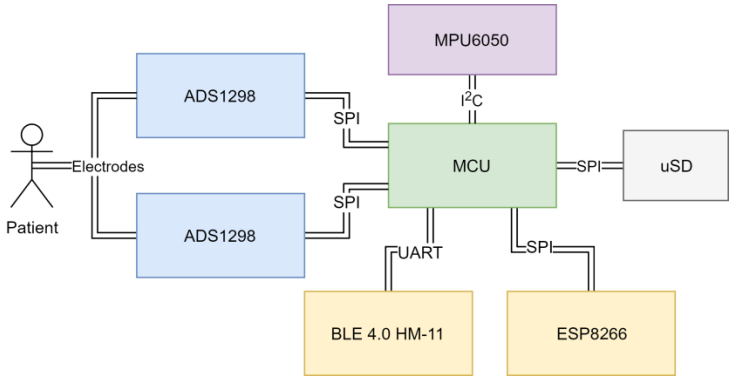
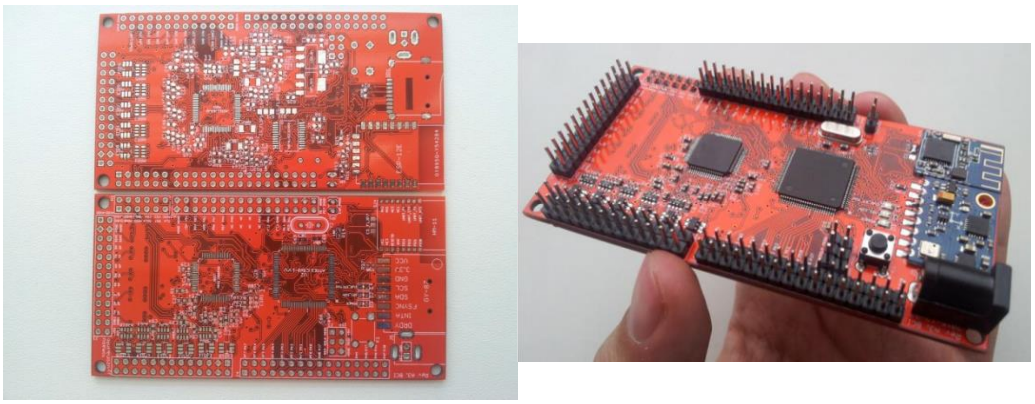


Figure 41. EEG system board component integration view

For the initial system version (Figure 41), an Atmega2560 microcontroller (MCU) has been selected running at 16 MHz. Interrupt based serial peripheral interface (SPI) communication for the ADC sampling and data transmission over wireless connection has been implemented. Two ADS1298 AFEs were tightly packed (top and bottom) on a single 4-layer printed circuit board (PCB) (Figure 42a) giving in total 16-channel EEG in the standalone system. Additional general purpose inputs-outputs (GPIOs) were broken out by two headers. For wireless communications, two add-on boards were used, i.e., Bluetooth 4.0 Low Energy HM-11 (top) and a popular ESP8266 Wi-Fi module (bottom). Additionally, an accelerometer and gyroscope MPU 6050 module controlled over I2C bus was added into the system. Local data storage was implemented by using a micro SD card slot. The image of a finished initial system board version can be seen in Figure 42b. The dimensions of the credit card sized board are 10 cm × 5 cm.



(**a**) (**b**)

Figure 42. Designed PCB: (a) Top and bottom of the PCB board, (b) Initial version of the finished system PCB board

In order to achieve expandable and modular architecture, a SPI header was exposed to the PCB for stacking additional boards up to the total count of four, thus, reaching in total 64 EEG channels. All ADS1298 devices were connected by using

the cascaded configuration mode (Figure 43). The other supported "Daisy-Chain" configuration type that was not acceptable due to the limitation-inability to read and write each ADS registers and was not used in this work.

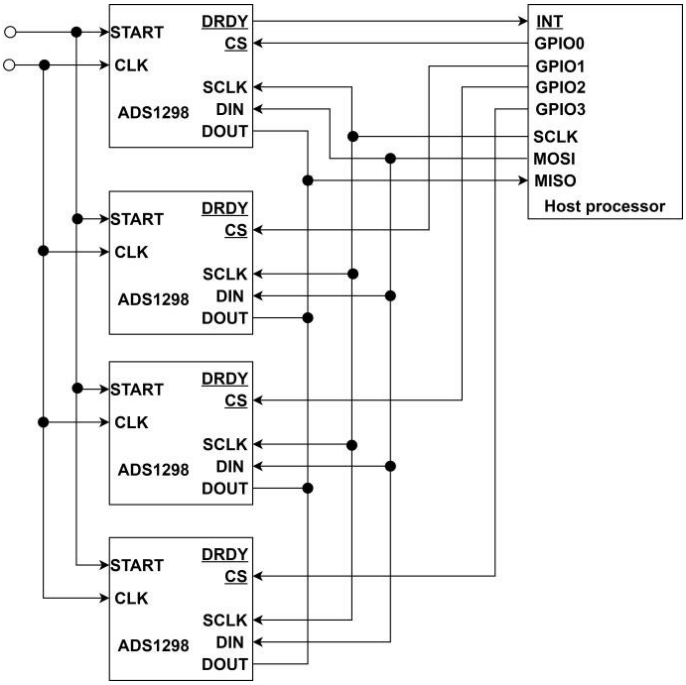


Figure 43. Cascaded view of multiple connected ADS1298 devices

5.3.3. Wireless communication

In order to decouple the system board from various AC and other noise sources, the EEG data must be sent to the host PC over a wireless connection. Furthermore, wireless transfer is the main solution for replacing long electrode cable braids and limiting the cable swing introduced signal noises and artifacts [93]. Multiple alternatives exist for such task. The most common approach is to send data over Bluetooth due to the very low power consumption of such technological devices. However, the short connection range and low data rates (baud) are the main bottlenecks of this technique when the higher sample rate or higher count of the EEG channels are used. Another approach is to use higher bandwidth communication technologies [94] such as Wi-Fi 802.11. By employing Wi-Fi, the bottleneck changes, and the limiting factor is the speed of the MCU.

Both technological approaches were used in the proposed EEG system. The Bluetooth component was implemented by using mini HM-11 BLE 4.0 module that is limited to a maximum baud of 230,400. The Wi-Fi component was implemented by using ESP8266-12E module via universal asynchronous receiver-transmitter (UART) and SPI interfaces that are limited to the maximum baud of 921,600 and MCU speed, respectively. The required baud rate B (Table 14) in bits for sending

uncompressed EEG data of different sampling speed F_s and the number of EEG channels N_{ch} can be computed by using Equation (45):

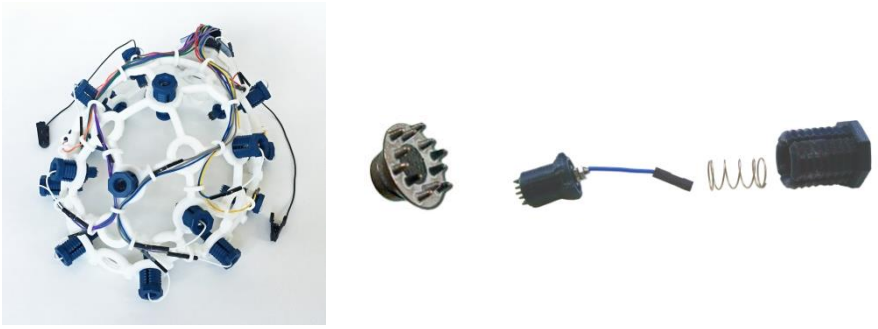
$$B = 24 \cdot (N_{ch} + 1) \cdot F_s. \quad (45)$$

Table 14. Bandwidth requirements for the raw EEG data

# of EEG Boards	Sampling Speed, Hz	# of Channels	BAUD, BPS	BLE 4.0 UART	ESP8266 UART/SPI
1	250	16	96,000	Yes	Yes/Yes
2	250	32	192,000	Yes	Yes/Yes
2	500	32	384,000	No	Yes/Yes
2	1000	32	768,000	No	Yes/Yes
4	250	64	384,000	No	Yes/Yes
4	500	64	768,000	No	Yes/Yes
4	1000	64	1,536,000	No	No/Yes

5.3.4. Electrode system and head cap

A prototype acrylonitrile butadiene styrene (ABS) plastic head cap (Figure 44a) based on the popular open-source Ultracortex OpenBCI model was printed by using a 3D printer and used in tests. The placement of electrodes in the head cap conforms to the international 10–20 electrode system.



(**a**) (**b**) (**c**)

Figure 44. Headcap components: (a) 3D Printed plastic head cap used for tests, (b) Electrode cap from Florida Research Institute, (c) Plastic electrode holder with a spring system

Since gel-based electrodes require the application of conductive paste and tend to dry out when used for prolonged times, a reusable dry type EEG electrodes (Figure 44b) from Florida Research Institute were tested instead. Dry electrodes must have good contact with skin to limit resistance to 10 kΩ or less [95]. Pressing the electrode against the skin surface tends to improve the contact with the skin. To prevent skin-electrode contact degradation (and thus impedance increase) and due to the advances in 3D printed part usage for EEG [96], a spring tension system (Figure 44c) for each electrode was used in screwable socket type holders to hold the electrode in place.

5.3.5. Accelerometer

The addition of an accelerometer and gyroscope into the acquisition system allowed the detection of artifacts in the EEG signal (as in Reference [97]) that were introduced due to the movement of the patient. It is not always possible for the subject to stay still for long periods of time. Due to the high component integration, it was optimal to use a pre-existing MPU-6050 module package for the initial version of the system PCB.

5.3.6. Cost of parts

The developed system consists of easily obtainable hardware parts. The initial goal was to design the PCB only from the essential pieces that are required for an EEG acquisition board. Table 15 shows the parts that were used, their prices and a possible source for building a single board (16-channels) without including manufacturing cost. In order to build a system with 64 channels, four such boards must be produced. The total price for a single board is about 114€ at the time when the author was writing this dissertation. This opens more possibilities for researchers and the general public to experiment with BCI. These significant savings come at the expense of performance and have no professional support for hardware and software.

Table 15. Bill of materials of a single EEG board

Part#	Item	Usage	Source	Count	Price/Pcs, Eur	Total, Eur
1	4 Layer PCB board	Base for mounting SMT devices	Seeed	1	8.00	8.00
2	ADS1298IPAG	Analog front-end chip	Mouser	2	31.00	62.00
3	ESP8266-12E	Wi-Fi module	eBay	1	1.40	1.40
4	MPU6050 GY-521	Accelerometer+gyro module	eBay	1	0.92	0.92
5	Atmega2560	Main CPU	eBay	1	4.20	4.20
6	HM-11	Bluetooth 4.0 module	eBay	1	1.34	1.34
7	SN74LVCC3245	TTL to 3V3 level shifter	Mouser	1	0.98	0.98
8	LM2664	Voltage inverter	Mouser	1	0.73	0.73
9	MCP1825S	5 V LDO/0.5 A	Mouser	1	0.50	0.50
10	MCP1825S-3V3	3.3 V LDO/0.5 A	Mouser	1	0.50	0.50
11	MIC5219-2.5	2.5 V LDO/0.5 A	Mouser	1	0.88	0.88
12	TPS72325	−2.5 V LDO/0.2 A	Mouser	1	2.23	2.23
13	Other components	Capacitors, resistors, diodes, buttons, pin headers, sockets	Mouser	1	30.00	30.00
				Total (€):	**82.68**	**113.68**

5.4. Evaluation

This section presents the EEG acquisition system board evaluation techniques and tests done to validate the operational correctness. While there are methods to verify the system by using high-priced third-party test equipment [98], simpler techniques exist to assess the system. The proposed methods are detailed in further sections. Stacked, four board system was validated with 64 electrodes. The validation tests were done by using a high-resolution (HR) mode with a PGA gain of one and sample rate of 500 Hz, while 1 kHz sampling rate was used for Wi-Fi bandwidth evaluation.

5.4.1. Internal ADC tests

ADS1298 analog front-end device contains several internal operation modes for validating the internal ADCs. The validation and calibration of ADCs are crucial for correct EEG recordings. ADS registers (CONFIG1, CONFIG2 and CONFIG3) were programmed to connect the internal test signal output to each channel ADC input (INT_TEST = 1). If channel ADCs are working correctly, the corresponding signal would be seen on each channel output. Three different signal generation modes were tested: slow 1 Hz square wave (TEST_FREQ = 0), fast 2 Hz square wave (TEST_FREQ = 1) and "DC" mode, which allowed constant high voltage (V_{CC}) to be set for each channel.

For each signal type, several second recordings have been captured by using 500 Hz sample rate. An example of 8-channel data from each test is shown in Figure 45. The recordings presented typical 1 Hz and 2 Hz square waves and DC pattern. This allowed concluding that ADCs were working properly. In order to validate the system integrity, each time ADS1298 was started, the same signals were used for device calibration.

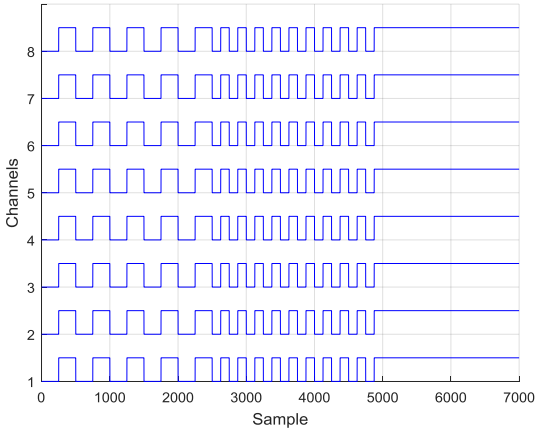


Figure 45. 1 Hz, 2 Hz and DC of recorded internal ADC test signals

5.4.2. Lead-off detection

Lead-off detection allowed validating ADS function to recognize the addition or removal of electrodes from the human scalp properly. This function ensures that electrodes will have contact with the scalp skin before any EEG recording is made. Lead-off detection has been validated by enabling lead-off detection for each of the EEG channels in ADS registers (LOFF_SENSP = 0xFF). The electrodes were placed on the subject's scalp, and the status of LOFF_STATP register was checked. The value of 0xFF for the register was expected for proper subject skin contact and value of 0x00 if all the electrode leads were removed. Additionally, each individual channel was checked by using the same routine. The device passed the lead-off detection test for all channels successfully.

5.4.3. EEG capture software

An open-source OpenBCI graphical user interface (GUI) was modified (Figure 46) to support the board that was developed in this work. The GUI was used for monitoring, recording and testing purposes. All raw EEG signal filtering (low-pass, high-pass and notch filters for 50/60 Hz [99]) were implemented in the software. Support for up to 64 channels has been introduced along with the accelerometer data visualization.

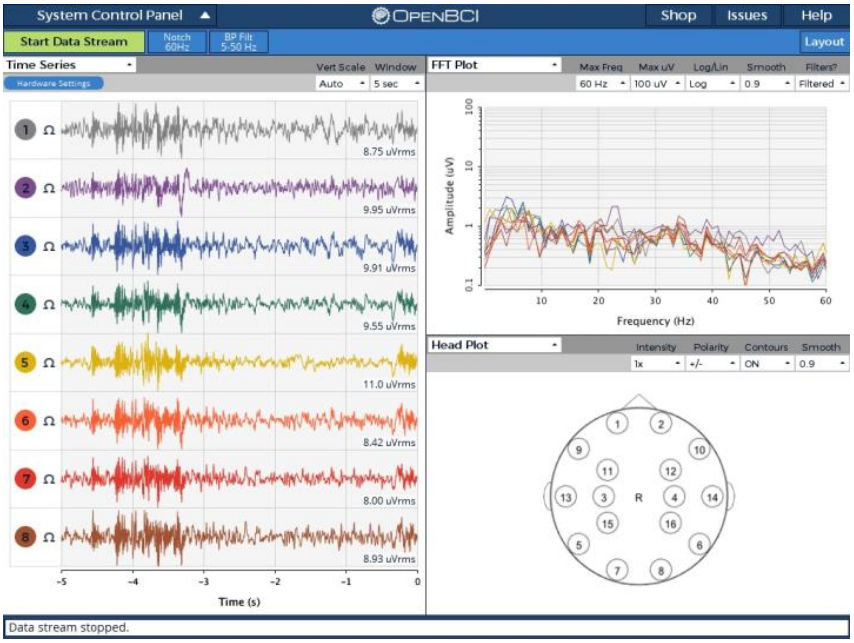


Figure 46. OpenBCI GUI used for validating each ADS1298

5.4.4. Teeth clenching and eye blinks

One type of the EEG signal artifacts that can be easily captured during the recording session is muscle induced teeth clenching and eye blinks [100]. The existence of these unwanted artifacts allows validating the sensitivity of the analog front-end. The EEG recording session has been initiated to see the artifact influence on the system. For this reason, eight electrodes were placed on the subject's scalp (based on the electrode placement system 10–20), and two different states were recorded, i.e., teeth clenching and eye blinks. The resultant EEG trace of the experiment is shown in Figure 47. Clenching artifacts are clearly visible (samples 270–450), while harder to recognize eye blinks have notable periodic behaviour (samples 525–700). The recording shows that the analog front-end is susceptible to muscle movement artifacts and confirms the sensitivity of the system.

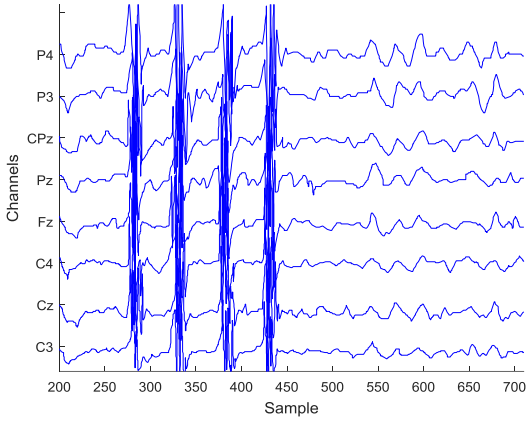


Figure 47. Teeth clenching and eye-blink test EEG signals

5.4.5. Alpha waves

Another common technique for the EEG recording system validation is to analyse alpha waves [101]. These waves can be recorded in a wakeful human subject during relaxation, when the subject's eyes are closed.

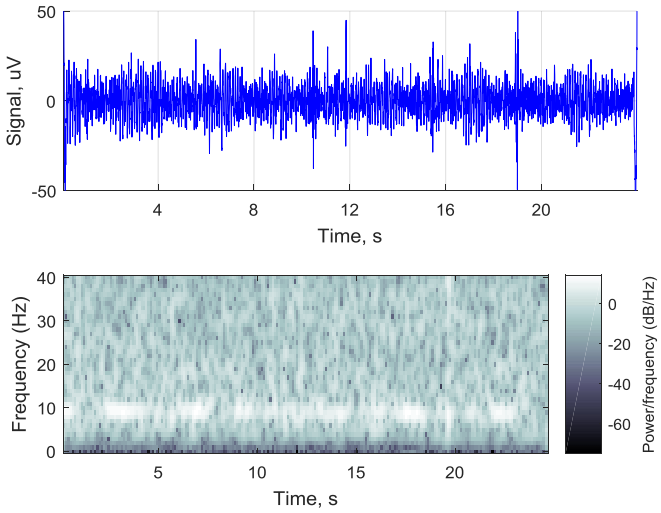


Figure 48. Alpha waves detection of the recorded EEG signal

Detection and recording of the alpha waves were tested by connecting the electrodes (O1, Oz, O2 from 10–20 electrode placement system) to the subject's scalp and asking them to relax, open their eyes for 30 s and stay relaxed with closed eyes for one minute. During the closed eyes interval, an increased activity in 7.5–12.5 Hz region in frequency domain showed a typical alpha wave signal (Figure 48) of brain's occipital lobe area. The acquired results proved that the system was able to successfully record EEG signal of such phenomenon.

5.4.6. ECG signal detection

One of simpler tests that can be initiated to validate any instrumental ADC is to record the activity of the heart (electrocardiogram or ECG). A healthy patient ECG was recorded by using three leads. An example of a 78 bpm ECG diagram is shown in Figure 49. Typical periodic QRS complexes are visible in 2 mV peak signal, which denotes proper functioning of the signal capture front-end.

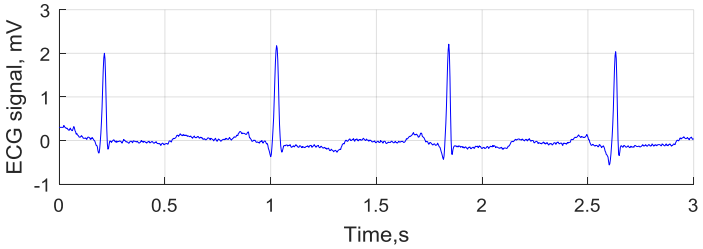


Figure 49. ECG signal recorded by using the developed board

5.4.7. Input referred noise

Input referred noise is each channel's characteristic showing the noise generated by internal ADS1298 chip circuitry and ADCs. The noise level for each channel was checked by sorting all channels inputs via ADS1298 register CHnSET = 1 (where *n* = 1 to 8) configuration and recording the noise floor for 10 s to a micro SD card by using different PGA and sampling rate settings. The averages of each channel noise are given in Table 16.

An example of noise floor recording for a single channel is given in Figure 50. The average channel input-referred noise value for this signal was 6.59 μV_{pp}.

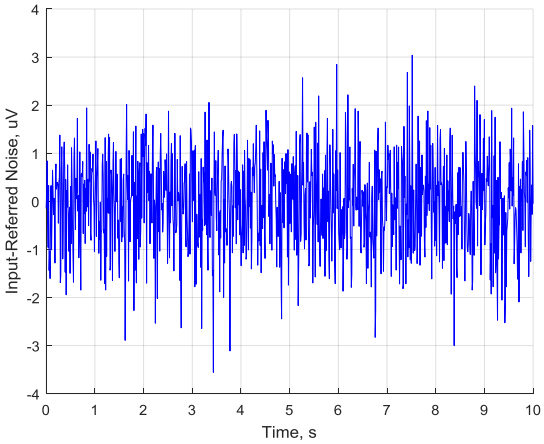


Figure 50. Channel input-referred noise signal (Fs = 1000 Hz, PGA = 3)

Table 16. Average channel input-referred noise μVPP

Sampling frequency Fs, Hz	PGA						
	1	**2**	**3**	**4**	**6**	**8**	**12**
32000	2876	1883	937	753	617	357	283
16000	710	285	152	152	101	66	48.83
8000	118	43.90	33.62	33.09	21.65	15.66	11.54
4000	47.49	27.70	15.41	11.49	11.08	10.72	8.94
2000	31.70	13.88	10.25	10.25	6.32	5.70	5.74
1000	16.56	8.67	7.82	6.04	4.55	5.23	3.38
500	14.85	5.92	4.92	4.77	3.94	3.13	2.87

It can be seen that noise is effectively cancelled as the sampling rate is decreasing (due to the averaging done by ADS). The maximum noise decrease of around 1883/5.92 = 318 times has been observed for PGA = 2. High sampling frequency is not required for the EEG applications; thus, the rates up to 1 kHz are more than enough to capture brain oscillations in 7–30 Hz range confidently. Further, the higher gain value allows reducing the input-referred noise. The maximum noise reduction of around 710/48.83 = 14.5 times has been seen in experiments for F_s = 16 kHz.

5.4.8. SNR and precision

Signal-to-noise ratio (SNR) shows the ability of the system to discern effective signals from the background noise. When working with EEG signals, it is a requirement to have as high SNR as possible, since valuable signal is in the same micro-volt range as the noise. The SNR in decibel scale is defined as shown in Equation (46):

$$SNR = 20\,log_{10}\left(\frac{A_{signal}}{A_{noise}}\right), \tag{46}$$

where A_{signal} and A_{noise} are the root-mean-square (RMS) amplitudes of the signal and noise. In order to evaluate the SNR of the designed system, the EEG recording experiment was conducted. First, a noise signal of 1 min length was recorded on all ADS1298 channels by using 250 Hz sampling speed, and the average noise RMS amplitude was calculated from all data. Next, a known amplitude effective 10 Hz sine signal was generated as input on each channel, and the same length recordings were taken. These acquired signals were used to calculate the average RMS amplitude and, finally, the SNR value. The input sine signal amplitude was scaled from 0 dB (100% VCC) to −100 dB (0.001% VCC) to capture system behaviour for very large and very small signals fully. The same technique was repeated for 500 Hz sample rate. The results of the experiment are shown in Figure 51. It can be noted that over 100 dB SNR is reached for the input signals whose amplitude is greater than −12 dB (>25% VCC). For typical 10–100 µV (−100 dB to −80 dB) EEG signals, the SNR value varies from 12 dB to 35 dB. A lower sampling rate gives higher SNR due to higher ADS1298 signal averaging/oversampling (noise cancellation).

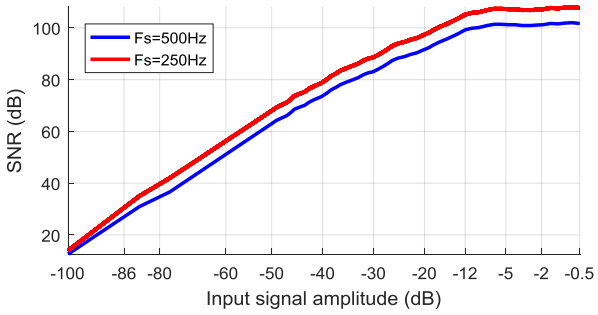


Figure 51. Signal-to-noise ratio (SNR) evaluation results

ADC precision evaluation used the same previously recorded data. Each of the recorded signal samples was compared with the original sine input signal values to find the conversion error. The results of this experiment can be seen in Figure 52. The average obtained error was 0.07% with a 0.22% standard deviation. The obtained results show a good match with the ones published in the official TI ADS1298 datasheet and allow qualifying the system as a properly working device.

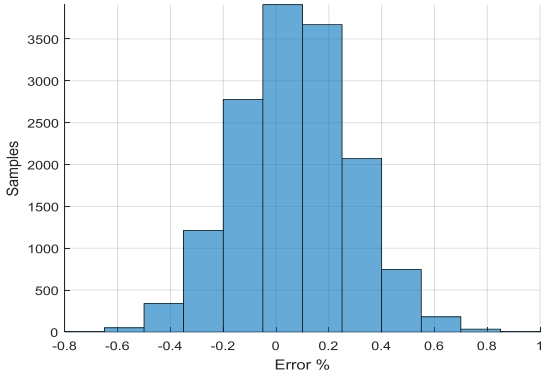


Figure 52. Signal precision for 60 sec recording (Fs = 250 Hz, PGA = 6)

5.4.9. Common Mode Rejection Ratio (CMRR)

The ability to reject common mode signal is crucial for the EEG recording systems. A higher common mode rejection ratio (CMRR) ensures that less common-mode signals will appear in the measurements. CMRR is a property of a differential amplifier [102]. The output of such an amplifier can be modelled as a sum of differential and common mode components as shown in Equation (47):

$$V_{out} = A_d V_{in} + A_{cm} V_{cm}, \tag{47}$$

where A_d is the differential and A_{cm} is common mode gains expressed as Equations (48) and (49):

$$A_d \approx V_{out}/V_{in}, \tag{48}$$

$$A_{cm} \approx V_{out}/V_{cm}, \tag{49}$$

where

$$V_{in} = (V_p - V_n), \tag{50}$$

$$V_{cm} = (V_p + V_n)/2, \tag{51}$$

with V_p being the voltage on positive input and V_n being the voltage on negative input. While $A_{cm} = 0$, and $A_d \to \infty$ for the ideal amplifiers. In real applications, $A_{cm} \neq 0$, and $A_{cm} \ll A_d$. Equation (49) equality holds only when the same common mode signal is fed to both amplifier inputs as the differential gain component is eliminated due to $V_{in} \to 0$. CMRR can be calculated by evaluating Equation (52):

$$CMRR = 20 \log\left(\frac{A_d}{A_{cm}}\right). \tag{52}$$

The test was performed by connecting each channel differential inputs INxP, INxN (for $x = \overline{1,8}$) and generating an external fixed frequency sine input common mode signal. The voltages were measured, and the average CMRR was computed by using Equation (52). The test was done for different gain values 1, 2, 4, 12 and frequencies from 1 Hz to 1 kHz. The results of the CMRR evaluation can be seen in Figure 53.

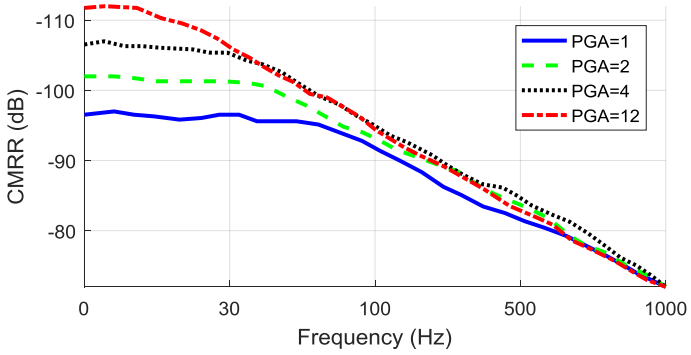


Figure 53. Common-mode rejection ratio measurement results

The test results show that the increase in gain from 1 to 12 allow ~ 15 dB better rejection ratio for 1–10 Hz signals to be achieved, despite a much quicker decline seen from 15 Hz to 1 kHz. It should be noted that the smallest gain value of one provides a stable CMRR of ~ 97 dB for signals up to ~ 70 Hz. In order to achieve the highest stable CMRR for EEG signals in 7–30 Hz frequency range, a gain value of four should be used.

5.5. Discussion

The implemented system exposes similar characteristics to the other state-of-the-art implementations while introducing new expandability features. The summary of functionality and comparison between the other systems found in the literature is given in Table 17.

It should be noted that most of the systems are designed to be non-expandable/modular from the start. Having a configurable system allows scale on demand and control to be achieved, minimizing the complexity for each problem. Thus, modularity has been taken into account while designing the proposed system. Since ADS1298 can be easily cascaded, such chip property was exploited.

Furthermore, the EEG applications require a moderate number of electrodes to capture brain oscillations reliably. High-end commercial systems are capable of recording 256 channels EEG. However, this creates a significant complexity and is harder to analyse and process later. The proposed design incorporates a configurable number of available channels, up to 64 (four stacked boards), while starting from eight channels (single sided board). Such an electrode count is commonly used in literature for capturing the EEG data with adequate head scalp coverage.

Some of the compared systems use active electrodes instead of passive or active shielding to remove cable swing/movement induced artifacts. The 3D printed head cap allows the minimization of the length of cables and fix them into position, thus, limiting movement-related artifacts. In order to fully suppress such artifacts, a switch from passive to active electrodes should be made. Since ADS1298 is not designed to work with the external pre-amplification stage, other AFE solutions will be necessary.

Table 17. Comparison of the existing systems

Property	**Proposed system**	**Pinho et al. [80]**	**Campillo et al. [84]**	**Boquete et al. [103]**	**Myung et al. [94]**
Modular	Yes	No	No	No	No
Channels	64	32	8	8	16
Sampling frequency, Hz	1000	1000	500	400	512
Electrodes	Passive dry	Active dry	Passive dry	Ag/AgCl adhesive	Wet gel
Resolution, bits	24	24	12	12	24
I/O	BLE 4.0, WiFi 802.11 b/g/n	WiFi 802.11 b/g/n	UART	Zig-bee 802.15.4	WiFi 802.11 d
CPU	Atmega 2560	DM3730	MSP430	Atmega 2560	STM32 F103
Clock frequency, MHz	16	1000	12	16	72
CMRR, dB	-110	-115	-94	-	-
Max gain	12	24	12	10k	-
Local processing	No	Yes	No	No	No
Power, mAh	250	500	-	100	80

The maximum sampling resolution is denoted by the AFE used for each device. Currently, a lot of the devices use a 24-bit AFE to record fine details of the EEG signal. However, it is hard to reach such high discretization resolution due to various PCB designs and physical issues. Thus, the real resolution is usually much lower due to the noise in the least significant bits (LSBs).

Depending on the application, the EEG signal processing can be done online or offline. Since mobile EEG devices usually run on battery power (to avoid additional common-mode and other noise from power sources), the power usage must be minimized. The selection of the low-power microprocessor, such as Atmega2560, running at 16 MHz and drawing ~ 30 mA@5V on full load still allowed the required 1 ksps (1k samples/s) bandwidth to be handled from 8x ADS1298. The power

consumption of 8x ADS1298 was ~ 48 mA@5V and MPU 6050 accelerometer ~ 2 mA@5V, while the most energy was wasted for the Wi-Fi connection ~ 170 mA@3.3V. This added up to 250 mAh for the whole four board stack running at maximum load (see Figure 54). With a typical 3000 mAh LiPo smartphone battery, the system can function up to 12 h.

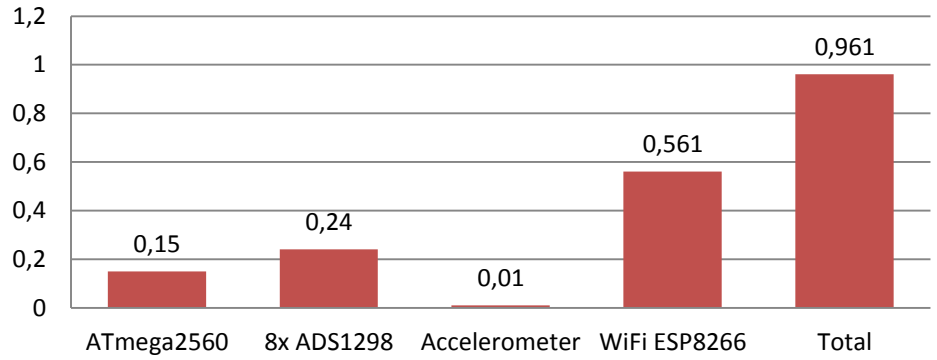


Figure 54. Power use in W/h

Maintaining a cost-efficient solution, while providing sufficient quality, is an important topic worth discussing. Since the first-ever EEG devices were made, the most critical analog system part shrunk and was embedded inside the silicon chip of the ADC, such as ADS1298. By doing this, higher quality for noise suppression and other parameter controls were achieved. In addition, the integration part got simpler since the microprocessor only needs to interface with the ADC chip. The bill of materials (BOM) has shrunk, and the most expensive part of the EEG system is the ADC chip itself. With the increasing integration level, the future of the EEG systems could evolve into a single programmable chip. With this extreme level of integration, further finer control of acquisition system properties could be achieved.

5.6. Conclusions

This chapter presented a modular biopotential acquisition system design that is capable of recording up to 64 EEG channels by exploiting a novel, stackable configuration.

1. Full board tests have been completed, and results showed correct working behaviour of each of the system components. The selected system architecture and ADC chip for the EEG acquisition proved to be a successful choice for building a compact and modern system.
2. The proposed simple evaluation techniques allowed verifying the system's ability to capture EEG signal correctly and effectively while providing the needed feedback for further development of the board.
3. Passed internal ADC tests were the initial step for verifying device correctness. The in-range to the official ADS1298 datasheet input referred to

the noise value of 6.59 μV_{pp}, and average CMRR of −97 dB in 0–70 Hz band was received in other performed tests.

4. The ability to capture EEG alpha waves phenomenon or ECG correctly as well signalled that the system was working as expected. During the experiments, 12–35 dB SNR for 10–100 μV EEG signals and greater than 100 dB SNR for signals with amplitudes bigger than 25% VCC were measured.
5. SNR and precision were found to match the proclaimed device characteristics closely as stated in the official Texas Instruments datasheet.
6. With a maximum power consumption of ~ 250 mAh on the full load and more than 10 times lower manufacturing cost (compared to commercial devices), the proposed system can be a portable device for cEEG or ECG acquisition and monitoring.
7. The system comparison with other developed boards found in the literature showed similar or better performance. However, compared to the commercial grade hardware, the system lacked better noise suppression, and further improvements are needed.
8. Since enhanced noise suppression is required for such high-resolution AFEs, further research and development can be directed towards active electrode implementation and shielding.

6. CONCLUSIONS

1. After the analysis of scientific papers on the topic of four-class motor-imagery (MI) classification, it has been determined that deep learning classification methods have not been widely analysed and used for the specified problem.
2. After the experimental evaluation of common feature extraction and classification algorithms for four-class motor imagery problem, it has been determined that new feature extraction and classification methods are necessary.
3. Solutions to the MI problem have been proposed and validated experimentally:
    a) new filtering/feature extraction Channel difference method has been introduced. It has been shown that the Channel difference method (54% accuracy) compares similarly to the CSP method (58% accuracy) and outperforms other tested feature extraction methods by 6–16% when not using CSP filtering.
    b) Convolutional Neural Networks (CNN) were adapted for the MI task. It has been shown that CNN method is effective, achieves 68% accuracy when using FFT energy maps for feature extraction and compares similarly (69% accuracy) to the other state-of-the-art classification methods found in the literature.
    c) proposed new feature vector adaptation technique allowed using feature extraction methods that produce 1D feature vectors for 2D feature map CNN classification successfully. The technique has been validated during the experiments, and promising classification results were achieved.
4. Due to the lack of four-class MI test datasets, low-cost solutions for scientific EEG signal recording devices and lack of design information in literature, a new EEG system capable of recording 64 channels of EEG MI signals has been successfully designed and manufactured.
5. The produced EEG system has been validated experimentally by using proposed simple but effective methods. The comparison with existing devices showcased similar or better performance:
    a) unique modular/stackable design with sampling speed up to 1 kHz;
    b) high CMRR of -97dB in 0-70 Hz has been measured;
    c) typical 12–35dB SNR for 10–100μV signals;
    d) low power consumption < 250mAh.

LITERATURE

1. UKTVERIS, T., and V. JUSAS. Application of Convolutional Neural Networks to Four-Class Motor Imagery Classification Problem. *Information Technology And Control*, vol. 46 (2), 2017, 260-273, http://dx.doi.org/10.5755/j01.itc.46.2.17528
2. MAHMOOD, A., N. MAHMOOD, N. IQBAL, Z. HAFEEZ. Methodology for EEG Based System Development to Detect Objective Pain in Human Body, *International Journal of Scientific and Engineering Research*, 2012, vol. 3(11), ISSN 2229-5518
3. YI, W., S. QIU, H. QI, L. ZHANG, B. WAN, D. MING. EEG feature comparison and classification of simple and compound limb motor imagery. *Journal of NeuroEngineering and Rehabilitation*, 2013, vol. 10(1), https://doi.org/10.1186/1743-0003-10-106
4. YI, W., S. QIU, K. WANG, H. QI, F. HE, P. ZHOU, L. ZHANG, D. MING. EEG oscillatory patterns and classification of sequential compound limb motor imagery. *Journal of NeuroEngineering and Rehabilitation*, 2016, vol. 13(1), https://doi.org/10.1186/s12984-016-0119-8
5. ÚBEDA, A., J. M. AZORÍN, R. CHAVARRIAGA. Classification of upper limb center-out reaching tasks by means of EEG-based continuous decoding techniques, *Journal of NeuroEngineering and Rehabilitation*, 2017, vol. 14(1), https://doi.org/10.1186/s12984-017-0219-0
6. ANGULO-SHERMAN, I. N., M. RODRÍGUEZ-UGARTE, N. SCIACCA, E. IÁÑEZ. Effect of tDCS stimulation of motor cortex and cerebellum on EEG classification of motor imagery and sensorimotor band power, *Journal of NeuroEngineering and Rehabilitation*, 2017, vol. 14(1), https://doi.org/10.1186/s12984-017-0242-1
7. ANDRADE, J., J. CECÍLIO, M. SIMÕES, F. SALES and M. CASTELO-BRANCO. Separability of motor imagery of the self from interpretation of motor intentions of others at the single trial level: an EEG study, *Journal of NeuroEngineering and Rehabilitation*, 2017, vol. 14(1), https://doi.org/10.1186/s12984-017-0276-4
8. WANG, K., Z. WANG, Y. GUO, F. HE, H. QI, M. XU and D. MING. A brain-computer interface driven by imagining different force loads on a single hand: an online feasibility study, *Journal of NeuroEngineering and Rehabilitation*, 2017, vol. 14(1), https://doi.org/10.1186/s12984-017-0307-1
9. GEORGIADIS, K., N. LASKARIS, S. NIKOLOPOULOS and I. KOMPATSIARIS. Exploiting the heightened phase synchrony in patients with neuromuscular disease for the establishment of efficient motor imagery BCIs, *Journal of NeuroEngineering and Rehabilitation*, 2018, vol. 15(1), https://doi.org/10.1186/s12984-018-0431-6
10. PFURTSCHELLER, G., P. LINORTNER, R. WINKLER, G. KORISEK, and G. MÜLLER-PUTZ. Discrimination of Motor Imagery-Induced EEG Patterns in Patients with Complete Spinal Cord Injury, *Computational Intelligence and Neuroscience*, 2009. https://doi.org/10.1155/2009/104180
11. SHE, Q., Y. MA, M. MENG, and Z. LUO. Multiclass Posterior Probability Twin SVM for Motor Imagery EEG Classification, *Computational Intelligence and Neuroscience*, 2015. https://doi.org/10.1155/2015/251945
12. XU, B., Y. FU, G. SHI, X. YIN, Z. WANG, H. LI, and C. JIANG. Enhanced Performance by Time-Frequency-Phase Feature for EEG-Based BCI Systems, *The Scientific World Journal*, 2014. https://doi.org/10.1155/2014/420561

13. HIGASHI, H. and T. TANAKA. Common Spatio-Time-Frequency Patterns for Motor Imagery-Based Brain Machine Interfaces, *Computational Intelligence and Neuroscience*, 2013. https://doi.org/10.1155/2013/537218
14. ZHANG, R., P. XU, T. LIU, Y. ZHANG, L. GUO, P. LI, and D. YAO. Local Temporal Correlation Common Spatial Patterns for Single Trial EEG Classification during Motor Imagery, *Computational and Mathematical Methods in Medicine*, 2013. https://doi.org/10.1155/2013/591216
15. ASTIGARRAGA, A., A. ARRUTI, J. MUGUERZA, R. SANTANA, J. I. MARTIN, and B. SIERRA. User Adapted Motor-Imaginary Brain-Computer Interface by means of EEG Channel Selection Based on Estimation of Distributed Algorithms, *Mathematical Problems in Engineering*, 2016. https://doi.org/10.1155/2016/1435321
16. BATULA, A. M., Y. E. KIM, and H. AYAZ. Virtual and Actual Humanoid Robot Control with Four-Class Motor-Imagery-Based Optical Brain-Computer Interface, *BioMed Research International*, 2017. https://doi.org/10.1155/2017/1463512
17. KIM, Y., J. RYU, K. K. KIM, C. C. TOOK, D. P. MANDIC, and C. PARK. Motor Imagery Classification Using Mu and Beta Rhythms of EEG with Strong Uncorrelating Transform Based Complex Common Spatial Patterns, *Computational Intelligence and Neuroscience*, 2016. https://doi.org/10.1155/2016/1489692
18. LIU, R., Z. ZHANG, F. DUAN, X. ZHOU, and Z. MENG. Identification of Anisomerous Motor Imagery EEG Signals Based on Complex Algorithms, *Computational Intelligence and Neuroscience*, 2017. https://doi.org/10.1155/2017/2727856
19. MONDINI, V., A. L. MANGIA, and A. CAPPELLO. EEG-Based BCI System Using Adaptive Features Extraction and Classification Procedures, *Computational Intelligence and Neuroscience*, 2016. https://doi.org/10.1155/2016/4562601
20. GUAN, S., K. ZHAO, and S. YANG. Motor Imagery EEG Classification Based on Decision Tree Framework and Riemannian Geometry, *Computational Intelligence and Neuroscience*, 2019. https://doi.org/10.1155/2019/5627156
21. JI, H., J. LI, R. LU, R. GU, L. CAO, and X. GONG. EEG Classification for Hybrid Brain-Computer Interface Using a Tensor Based Multiclass Multimodal Analysis Scheme, *Computational Intelligence and Neuroscience*, 2016. https://doi.org/10.1155/2016/1732836
22. KO, L. W., S. S. K. RANGA, O. KOMAROV, and C. C. CHEN. Development of Single-Channel Hybrid BCI System Using Motor Imagery and SSVEP, *Journal of Healthcare Engineering*, 2017. https://doi.org/10.1155/2017/3789386
23. MARTIŠIUS, I. and R. DAMAŠEVIČIUS. A Prototype SSVEP Based Real Time BCI Gaming System, *Computational Intelligence and Neuroscience*, 2016. https://doi.org/10.1155/2016/3861425
24. XYGONAKIS, I., A. ATHANASIOU, N. PANDRIA, D. KUGIUMTZIS, and P. D. BAMIDIS. Decoding Motor Imagery through Common Spatial Pattern Filters at the EEG Source Space, *Computational Intelligence and Neuroscience*, 2018. https://doi.org/10.1155/2018/7957408
25. SHE, Q., K. CHEN, Y. MA, T. NGUYEN, and Y. ZHANG. Sparse Representation-Based Extreme Learning Machine for Motor Imagery EEG Classification. *Computational Intelligence and Neuroscience*, 2018. https://doi.org/10.1155/2018/9593682
26. DAI, M., D. ZHENG, S. LIU, and P. ZHANG. Transfer Kernel Common Spatial Patterns for Motor Imagery Brain-Computer Interface Classification, *Computational and Mathematical Methods in Medicine*, 2018. https://doi.org/10.1155/2018/9871603

27. OZMEN, N. G., L. GUMUSEL, and Y. YANG. A Biologically Inspired Approach to Frequency Domain Feature Extraction for EEG Classification, *Computational and Mathematical Methods in Medicine*, 2018. https://doi.org/10.1155/2018/9890132
28. LI, M., X. LUO, J. YANG, and Y. SUN. Applying a Locally Linear Embedding Algorithm for Feature Extraction and Visualization of MI-EEG, *Journal of Sensors*, 2016. https://doi.org/10.1155/2016/7481946
29. BATRES-MENDOZA, P., M. A. IBARRA-MANZANO, E. I. GUERRA-HERNANDEZ, D. L. ALMANZA-OJEDA, C. R. MONTORO-SANJOSE, R. J. ROMERO-TRONCOSO, and H. ROSTRO-GONZALEZ. Improving EEG-Based Motor Imagery Classification for Real-Time Applications Using the QSA Method, *Computational Intelligence and Neuroscience*, 2017. https://doi.org/10.1155/2017/9817305
30. KHALID, S., T. KHALIL and S. NASREEN. A survey of feature selection and feature extraction techniques in machine learning, 2014 Science and Information Conference, 2014, pp. 372-378. https://doi.org/10.1109/SAI.2014.6918213
31. BRODU, N., et al. Comparative study of band-power extraction techniques for motor imagery classification. *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, IEEE Symposium on, 2011, pp. 1–6, https://doi.org/10.1109/CCMB.2011.5952105
32. PFURTSCHELLER, G., et al., Graz-Brain-Computer Interface: State of Research, MIT Press, 2007, pp. 65–84
33. KAISER, JF. On a simple algorithm to calculate the energy of a signal. IEEE *Int. Conf. Acoustic Speech Signal Process*; 1990, Albuquerque, NM. https://doi.org/10.1109/ICASSP.1990.115702
34. MARTIŠIUS, I., et al. Using Higher Order Nonlinear Operators for SVM Classification of EEG Data, *Elektronika ir Elektrotechnika*, 2012, vol. 119(3), pp. 99–102, http://dx.doi.org/10.5755/j01.eee.119.3.1373
35. DOLEZAL, J., V., CERNY, J., STASTNY. Online motor-imagery based BCI, *Applied Electronics (AE)*, International Conference on, 2012, pp.65,68, 5-7
36. TANDONNET, C., B., BURLE, T., HASBROUCQ, F., VIDAL. Spatial enhancement of EEG traces by surface Laplacian estimation: comparison between local and global methods. *Clinical Neurophysiology*, 2005, 116, 18-24, https://doi.org/10.1016/j.clinph.2004.07.021
37. QIN, L., B. HE. A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications. *Journal of Neural Engineering*, 2005, vol. 2, pp. 65–72, https://doi.org/10.1088/1741-2560/2/4/001
38. MÜLLER-GERKING, J. et al., Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical Neurophysiology*. 1999, 110(5):787-98. https://doi.org/10.1016/S1388-2457(98)00038-8
39. THANG, L. Q., C., TEMIYASATHIT. Increase performance of four-class classification for motor-imagery based brain-computer interface. *Computer, Information and Telecommunication Systems* (CITS), 2014 International Conference on, 2014, pp.1,5, 7-9, http://dx.doi.org/10.1109/CITS.2014.6878959
40. FISHER, R. A. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 1936, vol. 7, pp. 179–188, https://doi.org/10.1111/j.1469-1809.1936.tb02137.x
41. BEN-HUR, A., J., WESTON. A User's Guide to Support Vector Machines, *Data Mining Techniques for the Life Sciences, Methods in Molecular Biology*, 2010, vol. 609, pp. 223-239, https://doi.org/10.1007/978-1-60327-241-4_13

42. SZACHEWICZ, P. Classification of Motor Imagery for Brain-Computer Interfaces, *Master's thesis, Poznan University of Technology*, 2013, Poznan
43. HSU, C., et al., A Practical Guide to Support Vector Classification, *National Taiwan University*, 2010, Taiwan
44. BRUNNER, C., et al. BCI Competition 2008 – Graz data set A, 2008, online: https://www.bbci.de/competition/iv/desc_2a.pdf
45. SCHLOGL, A., et al., Evaluation criteria in BCI research. In: G. Dornhege, J. del R. Millan, T. Hinterberger, D. J. McFarland, K.-R. Muller (Eds.). *Toward brain-computer interfacing*, MIT Press, 2007, pp. 327–342
46. ANG, K., K., et al., Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface, *Neural Networks*, 2008. IJCNN 2008. IEEE World Congress on Computational Intelligence, 2008, pp.2390,2397, 1-8, https://doi.org/10.1109/IJCNN.2008.4634130
47. MIROWSKI, P., Y. LECUN, D. MADHAVAN, R. KUZNIECKY. Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG, *IEEE Workshop Machine Learning Signal Processing*, 2008, https://doi.org/10.1109/MLSP.2008.4685487
48. CECOTTI, H., A. GRÄSER. Convolutional neural network with embedded fourier transform for EEG classification, *Proceedings of the 19th International Conference on PatternRecognition (ICPR '08),* 2008, 1–4, https://doi.org/10.1109/ICPR.2008.4761638
49. BEVILACQUA, V. et al. A novel BCI-SSVEP based approach for control of walking in Virtual Environment using a Convolutional Neural Network, *International Joint Conference on Neural Networks (IJCNN)*, 2014, 4121-4128, https://doi.org/10.1109/IJCNN.2014.6889955
50. CECOTTI, H., A. GRÄSER. Convolutional neural networks for P300 detection with application to brain–computer interfaces, I*EEE Trans. Pattern Anal. Mach. Intell.*, 2011, vol. 33(3), 433-445, https://doi.org/10.1109/TPAMI.2010.125
51. MANOR, R., A. B. GEVA. Convolutional Neural Network for Multi-Category Rapid Serial Visual Presentation BCI, *Frontiers Media S.A*, 2015, https://dx.doi.org/10.3389%2Ffncom.2015.00146
52. QIN, L., B. HE. A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications, *J. Neural Eng.*, 2005, Vol. 2, 65-72, https://doi.org/10.1088/1741-2560/2/4/001
53. XIAO, D., Z. D. MU, J. F. HU. Classification of motor imagery EEG signals based on energy entropy, 2009 *International Symposium on Intelligent Ubiquitous Computing and Education*, 2009, 61-64, https://doi.org/10.1109/IUCE.2009.57
54. ZHOU, B., WU, X., ZHANG, L., LV, Z. and GUO, X. Robust Spatial Filters on Three-Class Motor Imagery EEG Data Using Independent Component Analysis. *Journal of Biosciences and Medicines*, 2014, vol. 2, 43-49, http://dx.doi.org/10.4236/jbm.2014.22007
55. BAI, X., X. WANG, S. ZHENG, M. YU. The offline feature extraction of four-class motor imagery EEG based on ICA and Wavelet-CSP, *Control Conference (CCC)*, 2014, 7189-7194, https://doi.org/10.1109/ChiCC.2014.6896188
56. YANG, H., S. SAKHAVI, K. K. ANG, C. GUAN. On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification, *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, 2620-2623, https://doi.org/10.1109/EMBC.2015.7318929

57. LECUN, Y., L. BOTTOU, Y. BENGIO, P. HAFFNER. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11), 2278–2324, https://doi.org/10.1109/5.726791
58. VEDALDI, A., K. LENC. MatConvNet - Convolutional Neural Networks for MATLAB, *Proc. of the ACM Int. Conf. on Multimedia*, 2015, https://arxiv.org/abs/1412.4564
59. NAEEM, M., C. BRUNNER, G. PFURTSCHELLER. Dimensionality Reduction and Channel Selection of Motor Imagery Electroencephalographic Data, *Computational Intelligence and Neuroscience*, 2009, http://dx.doi.org/10.1155/2009/537504
60. WANG, Y., S. GAO, X. GAO. Common Spatial Pattern Method for Channel Selection in Motor Imagery Based Brain-Computer Interface, *IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005, 5392-5395, https://doi.org/10.1109/IEMBS.2005.1615701
61. UKTVERIS, T., V. JUSAS. Comparison of Feature Extraction Methods for EEG BCI Classification, *Information and Software Technologies: 21st International Conference*, 2015, 81-92, https://doi.org/10.1007/978-3-319-24770-0_8
62. BOX, G. E. P., D. R. COX. An analysis of transformations, *Journal of the Royal Statistical Society*, Series B, 1964, vol. 26(2), 211-252, http://www.jstor.org/stable/2984418
63. WANG, D., D. MIAO, G. BLOHM. Multi-Class Motor Imagery EEG Decoding for Brain-Computer Interfaces. *Frontiers in Neuroscience*, 2012, https://doi.org/10.3389/fnins.2012.00151
64. BIRVINSKAS, D., V. JUSAS, I. MARTIŠIUS, R. DAMAŠEVIČIUS. EEG dataset reduction and feature extraction using discrete cosine transform. *UKSim-AMSS EMS 2012: 6th European Modelling Symposium on Mathematical Modeling and Computer Simulation 2012*, 186-191, https://doi.org/10.1109/EMS.2012.88
65. HU, J., Z. MU, D. XIAO. Application of Energy Entropy in Motor Imagery EEG Classification. *JDCTA*, 2009, vol. 3, 83-90, http://dblp.dagstuhl.de/rec/bib/journals/jdcta/HuXM09
66. UKTVERIS, T., and V. JUSAS. Development of a Modular Board for EEG Signal Acquisition, *Sensors (Basel, Switzerland)*, vol. 18,7 2140, 2018, https://doi.org/10.3390/s18072140
67. LE VAN QUYEN, M., J. FOUCHER, J. LACHAUX, E. RODRIGUEZ, A. LUTZ, J. MARTINERIE, F. J. VARELA. Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. *Journal of Neuroscience Methods*, 2001, vol. 111, 83-98, https://doi.org/10.1016/S0165-0270(01)00372-7
68. ANG, K. K., Z. Y. CHIN, C.WANG, C. GUAN, and H. ZHANG. Filter bank common spatial pattern algorithmon BCI competition IV datasets 2a and 2b, *Frontiers in Neuroscience*, 2012, vol. 6. https://dx.doi.org/10.3389%2Ffnins.2012.00039
69. BCI IV competition final results. Available online: http://bbci.de/competition/iv/results/ (accessed on 8 June 2019)
70. TABAR, Y. R., U. HALICI. A novel deep learning approach for classification of EEG motor imagery signals, *Journal of Neural Engineering*, 2016, vol. 14(1), https://doi.org/10.1088/1741-2560/14/1/016003
71. RASHID, N., J., IQBAL, A., JAVED, M.I., TIWANA, U. S., KHAN. Design of Embedded System for Multivariate Classification of Finger and Thumb Movements Using EEG Signals for Control of Upper Limb Prosthesis, *BioMed Research International*, 2018. https://doi.org/10.1155/2018/2695106

72. DAMAŠEVIČIUS, R., R., MASKELIŪNAS, E., KAZANAVIČIUS, M., WOŹNIAK. Combining Cryptography with EEG Biometrics, *Computational Intelligence and Neuroscience*, 2018. https://doi.org/10.1155/2018/1867548
73. LIU, A., K., CHEN, Q., LIU, Q., AI, Y., XIE, A., CHEN. Feature Selection for Motor Imagery EEG Classification Based on Firefly Algorithm and Learning Automata. *Sensors* 2017, 17, 2576. https://doi.org/10.3390/s17112576
74. PLANELLES, D., E., HORTAL, Á., COSTA, A., ÚBEDA, E., IÁEZ, J.M., AZORÍN. Evaluating Classifiers to Detect Arm Movement Intention from EEG Signals. *Sensors* 2014, *14*, 18172-18186. https://doi.org/10.3390/s141018172
75. TANG, Z., S., SUN, S., ZHANG, Y., CHEN, C., LI, S., CHEN. A Brain-Machine Interface Based on ERD/ERS for an Upper-Limb Exoskeleton Control. *Sensors* 2016, *16*, 2050. https://doi.org/10.3390/s16122050
76. BLUM, S., S., DEBENER, R., EMKES, N., VOLKENING, S., FUDICKAR, M. G., BLEICHNER. EEG Recording and Online Signal Processing on Android: A Multiapp Framework for Brain-Computer Interfaces on Smartphone, *BioMed Research International*, 2017. https://doi.org/10.1155/2017/3072870
77. G.TEC MEDICAL ENGINEERING GMBH, Products. Available online: http://www.gtec.at/Products (accessed on 26 June 2018).
78. ASKAMP, J., M. J. A. M. VAN PUTTEN. Mobile EEG in epilepsy. *Int. J. Psychophysiol.* 2014, *91*, 30–35, https://doi.org/10.1016/j.ijpsycho.2013.09.002.
79. TWENTE MEDICAL SYSTEMS INTERNATIONAL, Products. Available online: https://www.tmsi.com/products (accessed on 26 June 2018).
80. PINHO, F., J. H., CORREIA, N. J., SOUSA, J. J., CERQUEIRA, AND N. S., DIAS. Wireless and wearable eeg acquisition platform for ambulatory monitoring. *IEEE 3nd International Conference on Serious Games and Applications for Health (SeGAH)*, Rio de Janeiro, 2014, pp. 1-7, https://doi.org/10.1109/SeGAH.2014.7067078
81. FENG, S., M., TANG, F., QUIVIRA, T., DYSON; F., CUCKOV, AND G., SCHIRNER. EEGu2: an embedded device for brain/body signal acquisition and processing. *International Symposium on Rapid System Prototyping (RSP)*, Pittsburgh, PA, 2016, pp. 1-7. https://doi.org/10.1145/2990299.2990304
82. SENEVIRATHNA, B., L., BERMAN, N., BERTONI, F., PARESCHI, M., MANGIA, R., ROVATTI, G., SETTI, J., SIMON, P., ABSHIRE. Low cost mobile EEG for characterization of cortical auditory responses. 2016 *IEEE International Symposium on Circuits and Systems (ISCAS)*, Montreal, QC, 2016, pp. 1102-1105. https://doi.org/10.1109/ISCAS.2016.7527437
83. VO, T. T., N. P., NGUYEN, T., VO VAN. WEEGEE: Wireless 8-Channel EEG Recording Device. 6th International Conference on the Development of Biomedical Engineering in Vietnam (BME6). *BME 2017. IFMBE Proceedings*, Singapore, 2017, vol 63, pp. 621-625. https://doi.org/10.1007/978-981-10-4361-1_106
84. CAMPILLO, D., R., GUARDARRAMA, R., GONZÁLEZ, J., RODRÍGUEZ, D., JIMÉNEZ. A real time ECG preprocessing system based on ADS1298. *Computing in Cardiology*, Zaragoza, 2013, pp. 947-950.
85. WILD, M., R., PEGAN. Wearable Bluetooth brain-computer interface for detection and analysis of ear-EEG Signals, 2015.
86. NATHAN, V., J., WU, C., ZONG, Y., ZOU, O., DEHZANGI, M., REAGOR, R., JAFARI. A 16-channel bluetooth enabled wearable EEG platform with dry-contact electrodes for brain computer interface. *In Proceedings of the 4th Conference on Wireless Health (WH '13)*, ACM, New York, NY, USA, 2013. https://doi.org/10.1145/2534088.2534098

87. CONSUL-PACAREU, S., R., MAHAJAN, M. J., ABU-SAUDE, B. I., MORSHED. NeuroMonitor: a low-power, wireless, wearable EEG device with DRL-less AFE. *IET Circuits, Devices & Systems*, vol. 11(5), pp. 471-477, 9 2017. https://doi.org/10.1049/iet-cds.2016.0256
88. MASTINU, E., M., ORTIZ-CATALAN, B., HÅKANSSON. Analog front-ends comparison in the way of a portable, low-power and low-cost EMG controller based on pattern recognition. *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, 2015, pp. 2111-2114. https://doi.org/10.1109/EMBC.2015.7318805
89. ACHARYA, D., A., RANI, S., AGARWAL. EEG data acquisition circuit system based on ADS1299EEG FE. *4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, Noida, 2015, pp. 1-5. https://doi.org/10.1109/ICRITO.2015.7359346
90. SPICER, R., J., ANGLIN, D. M., KRUM, S. L., LIEW. REINVENT: A low-cost, virtual reality brain-computer interface for severe stroke upper limb motor recovery. *2017 IEEE Virtual Reality (VR)*, Los Angeles, CA, 2017, pp. 385-386. https://doi.org/10.1109/VR.2017.7892338
91. ZIELENIEWSKA, M., A., CHABUDA, M., BIESAGA, R., KUŚ, P., DURKA. ADS 1299-based open hardware amplifier from OpenBCI.com: signal quality for EEG registration and SSVEP-based BCI, *Neurotechnix*, 2015
92. TEXAS INSTRUMENTS. ADS129x Low-Power, 8-Channel, 24-Bit Analog Front-End for Biopotential Measurements. Accessed 2018-05-11, http://www.ti.com/lit/ds/symlink/ads1298.pdf
93. SYMEONIDOU, E. R., A. D., NORDIN, W. D., HAIRSTON, D. P., FERRIS. Effects of Cable Sway, Electrode Surface Area, and Electrode Mass on Electroencephalography Signal Quality during Motion, *Sensors 2018*, 18, 1073. https://doi.org/10.3390/s18041073
94. MYUNG, B., S., YOO. Development of 16-channels Compact EEG System Using Real-time High-speed Wireless Transmission. *Engineering*, vol. 5, p. 93, 2013. https://doi.org/10.4236/eng.2013.55B019
95. LOPEZ-GORDO, M. A., D., SANCHEZ-MORILLO, F. P., VALLE. Dry EEG Electrodes. *Sensors* 2014, 14, 12847-12870. https://doi.org/10.3390/s140712847
96. KRACHUNOV, S., A. J., CASSON. 3D Printed Dry EEG Electrodes. *Sensors* 2016, 16, 1635. https://doi.org/10.3390/s16101635
97. MASSAI, S., S., ROUTHU, D., WRIGHT, K.S., MOON, Y., OZTURK, S. Q., LEE. A Wireless Visual Attention Brain Signal Monitoring System. *MATEC Web of Conferences*, vol. 32, pp. 04005, 2015. https://doi.org/10.1051/matecconf/20153204005
98. TORESANO, L. O. H. Z., S. K., WIJAYA, PRAWITO, A., SUDARMAJI, C., BADRI. Data acquisition system of 16-channel EEG based on ATSAM3X8E ARM Cortex-M3 32-bit microcontroller and ADS1299. *AIP Conference Proceedings*, vol. 1862, pp. 030149, 2017. https://doi.org/10.1063/1.4991253
99. DEHZANGI, O., M., FAROOQ. Portable Brain-Computer Interface for the Intensive Care Unit Patient Communication Using Subject-Dependent SSVEP Identification. *BioMed Research International*, 2018. https://doi.org/10.1155/2018/9796238
100. LIAO, L. D., S. L., WU, C. H., LIOU, S. W., LU, S. A., CHEN, L. W., KO, C. T., LIN. A Novel 16-Channel Wireless System for Electroencephalography Measurements With Dry Spring-Loaded Sensors. *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 6, pp. 1545-1555, 2014. https://doi.org/10.1109/TIM.2013.2293222
101. CHEN, J., X., LI, X., MI, S., PAN. A high precision EEG acquisition system based on the CompactPCI platform. 2014 *7th International Conference on Biomedical*

Engineering and Informatics, Dalian, 2014, pp. 511-516. https://doi.org/10.1109/BMEI.2014.7002828
102. ANALOG DEVICES. Op Amp Common-Mode Rejection Ratio. Accessed 2018-05-11, http://www.analog.com/media/en/training-seminars/tutorials/MT-042.pdf
103. BOQUETE, L., J. M. R., ASCARIZ, J., CANTOS, R., BAREA, J. M., MIGUEL, S., ORTEGA, N., PEIXOTO. A portable wireless biometric multi-channel system. Measurement, vol. 45, pp. 1587-1598, 2012. https://doi.org/10.1016/j.measurement.2012.02.018

LIST OF PUBLICATIONS

Articles in *Web of Science* database publications having citation index:

1. Uktveris, T., Jusas, V. Application of Convolutional Neural Networks to Four-Class Motor Imagery Classification Problem. *Information Technology and Control*. 2017, 46 (2), pp. 260–273.
2. Uktveris, T., Jusas, V. Development of a Modular Board for EEG Signal Acquisition. *Sensors*. 2018, 18, 2140.

Articles in *Web of Science* database publications without citation index:

1. Uktveris, T., Jusas, V. Comparison of Feature Extraction Methods for EEG BCI Classification. *Information and Software Technologies: 21st International Conference, ICIST 2015*, 2015, pp. 81-92.
2. Uktveris T., Jusas, V. Convolutional Neural Networks for Four-Class Motor Imagery Data Classification. *Intelligent Distributed Computing XI. IDC 2017*, 737, 2018, pp. 185-197.
3. Uktveris T., Jusas, V. Development of a Modular Board for EEG Signal Acquisition. *Fifth International Conference on Mathematics and Computers in Sciences and in Industry (MCSI)*, 2018.

UDK 004.5+616.831-073] (043.3)

SL344. 2019-07-02, 12,75 leidyb. apsk.l., tiražas 14, užsakymas 150.
Išleido Kauno technologijos universitetas, K. Donelaičio g. 73, 44249 Kaunas
Spausdino leidyklos „Technologija“ spaustuvė, Studentų g. 54, 51424 Kaunas