UNIVERSITY OF SHEFFIELD

AUTOMATIC CONTROL AND SYSTEMS ENGINEERING

DOCTOR OF PHILOSOPHY THESIS

Transfer Learning for Motor Imagery based Brain Computer Interfaces

Author

Mr. JOSHUA GILES

Supervisor

Dr. Mahnaz ARVANEH

Second supervisor

Dr. Kai Keng ANG

Third supervisor

Professor Lyudmila S MIHAYLOVA

April 8, 2023

Acknowledgment

This endeavour would not have been possible without my supervisor Dr. Mahnaz Arvaneh who has helped guide me throughout my PhD journey. Her guidance and advice helped me gain a better understanding of science and the world of research. Without her help and understanding, I do not believe I would have been able to reach this point.

I would also like to express my deepest gratitude to Dr. Kai Keng Ang my second supervisor for all his constructive comments and the advice he was always happy to provide. He has helped show me the world of research beyond academia and helped me settle into the Singaporean culture.

I am also grateful to Prof.Lyudmila Mihaylova who acted as my third supervisor for helping to guide me through the start of my PhD. As well as all my colleagues, both in Singapore and Sheffield, who were always available to discuss ideas or just help me de-stress.

I would also like to give my deepest thanks to all my friends and family who helped me when I was struggling. I am so thankful for all of you and all the support you gave me. In particular, I would like to thank my parents for always being there even when not convenient.

Thanks and appreciation should also go to the University PhD support staff for all their hard work and help. Finally, I would like to thank both the University of Sheffield and the Agency of Scientific, Technology and Research (A*STAR) in Singapore for allowing me the opportunity of completing this PhD and for providing funding.

Abstract

Current electroencephalogram (EEG) based brain-computer interface (BCI) systems have limited real-world practicality due to a number of issues, including the long calibration period required before each use. This thesis focuses on reducing the time required to calibrate the BCI system without sacrificing classification accuracy. To address this issue, previously collected EEG data could be potentially mined and reused in calibrating the BCI model for a new user/session. However, this is not a trivial task due to two key challenges. First, there are considerable non-stationarities between the current and previously collected EEG signals. Secondly, due to between-session variations, not all the previously collected EEG signals are helpful in training the new BCI model.

Initially, the thesis explored the application of distribution alignment techniques to reduce the effects of EEG non-stationarity. A novel multiclass data space alignment (MDSA) algorithm was proposed and evaluated. Our results showed that the proposed MDSA alignment algorithm successfully improved the classification accuracy and reduced the effects of non-stationarity.

The thesis then addressed the second challenge by developing a new framework. This framework utilised a new algorithm that identifies whether or not the new session would benefit from transfer learning. If so, a novel similarity measurement, called the Jensen-Shannon Ratio (JSR), was proposed to select one of the past session for training the BCI model. The proposed framework outperformed state-of-the-art algorithms when there were as few as five labelled trials per class available from the new session. Despite success to some extent the proposed framework was limited to a binary selection between only one of the past sessions and current data for training the BCI model.

Finally, the thesis utilised the findings of the previous research in order to address both challenges. A novel transfer learning framework was proposed for long-term BCI users. The proposed framework utilised regularisation, alignment and weighting to train a BCI which outperformed state-of-the-art algorithms even when only two trials per class from the new session were available.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Brain-computer interfaces (BCI) are systems designed to allow direct communication between the brain and a computer without the need for any additional input from the user [2]. This technology is beneficial for people who suffer from severe impairments with movement, such as locked-in patients, as the BCI can help to replace lost movements and communication [3]. Recently, BCI has been successfully used for stroke rehabilitation by activating the affected sensorimotor networks [4]. Motor imagery-based BCI (MI-BCI), using electroencephalogram (EEG) for brain signal acquisition, is the most common form of BCI. MI-BCI changes spontaneous EEG signals through the imagination of movement of different parts of the body as they can consistently be differentiated and are intuitive for the user to perform [2].

The main benefit of EEG is the high temporal resolution that can be obtained without requiring the patient to have any invasive operations [2]. Using EEG, BCI systems have been designed to control wheelchairs, robotic hands and a range of other devices to allow the users to communicate with their environment [5]. However, although the results are promising, BCI still requires a lot of improvement in order to be used reliably in daily life.

The significant challenges faced by EEG-BCIs are caused by the high dimensionality and non-stationary nature of the brain signals. In addition, the EEG signals being collected to control the BCI are affected by an extensive range of different factors including the emotions being felt, the users' fatigue level and several other factors [6, 7]. These changes in the EEG

affect the properties of the features extracted by the BCI.

Given EEG signals' high dimensionality and non-stationarity nature, MI-BCIs require a 20-30 minute calibration session before each use of the BCI [8]. The user is asked to perform the same actions repeatedly during this time. The collected data becomes the training data set made up of labelled trials. The collected data from one or more of these EEG recordings are then used to train the feature extraction model and the classifier. This extended calibration allows the system to identify the users' intention with relatively high accuracy in most cases, although some users still encounter BCI deficiency [2].

Transfer learning can be potentially used to reduce the BCI calibration time without compromising the BCI accuracy [9]. Transfer learning is a commonly employed technique in systems engineering when only a limited amount of data is available to train the model. Transfer learning compensates for the limited amount of labelled data available by extracting relevant information from other similar sources or domains to improve the classification model [9]. However, transfer learning in BCI is not a trivial task due to the unique structure of every brain and the non-stationary nature of brain signals. Furthermore, the properties of EEG signals often change considerably from session-to-session [10].

In some cases, data from a target BCI session has unique probability distributions, which are very different from the data distributions of other sessions (i.e. source sessions). In these cases, utilising source data can be detrimental to the classification accuracy of the target BCI session. Thus, it is critical to identify whether or not the source data will be detrimental and then apply some techniques to either reduce the effect of the detrimental source data on training the target BCI model or eradicate it.

To reduce the effects of the non-stationarities, various approaches have been explored and embedded in transfer learning algorithms proposed for BCI [11,12]. These approaches primarily focused on inter-subject transfer learning, evaluating the proposed solutions on datasets with only one or two sessions of data available for each subject. There is a research bias within BCI, with the majority of studies focusing on datasets where only one or two sessions are recorded from each subject [13]. A relatively small amount of literature focuses on inter-session transfer learning for long-term users. A greater understanding of inter-session transfer learn-

ing could be advantageous when developing BCIs for rehabilitation when the user is expected to use the BCI repeatedly over an extended time.

1.2 Motivation

Much work has been put into BCI increasing its accuracy and robustness. These research studies have led to several advancements in the field, making the BCI a more practical option for patients. Despite this, BCIs are not yet used in daily life due to the extended calibration times required at the beginning of each session, limiting their practicality and applications.

The motivation of this report is to develop novel transfer learning algorithms to improve the current BCI, reducing the calibration requirements while maintaining or improving accuracy. Our main ambition is to develop a BCI system for long-term BCI users, particularly stroke patients, that can operate accurately with almost zero training time. The focus of this thesis is on MI-BCI and EEG signals. As already discussed, due to its portability and high temporal accuracy, EEG can potentially be an accessible brain imaging technique with good practical applications [14].

1.3 Aims and Objectives

The aim of this research is to develop novel transfer learning frameworks for BCI users which reduce the calibration time required by as much as possible while maintaining effective levels of classification accuracy. Figure 1.1 shows the challenges related to this aim and how the objectives can address the challenges in order to fulfil the aim.

As shown in the figure there are two main challenges that need to be addressed in order to complete the aim of this project. The initial challenge is to reduce the non-stationarities between the source sessions, used for training data, and the target session, used for testing. These non-stationarities are one of the key issues that lead to long calibration times being required, particularly when the training data being used is from a different day or collected in a different setting. Reducing the effect of these non-stationarities can improve the classification accuracy without the need to completely retrain the BCI with the target session.

The second challenge is the large variations in the distribution of the data across the avail-

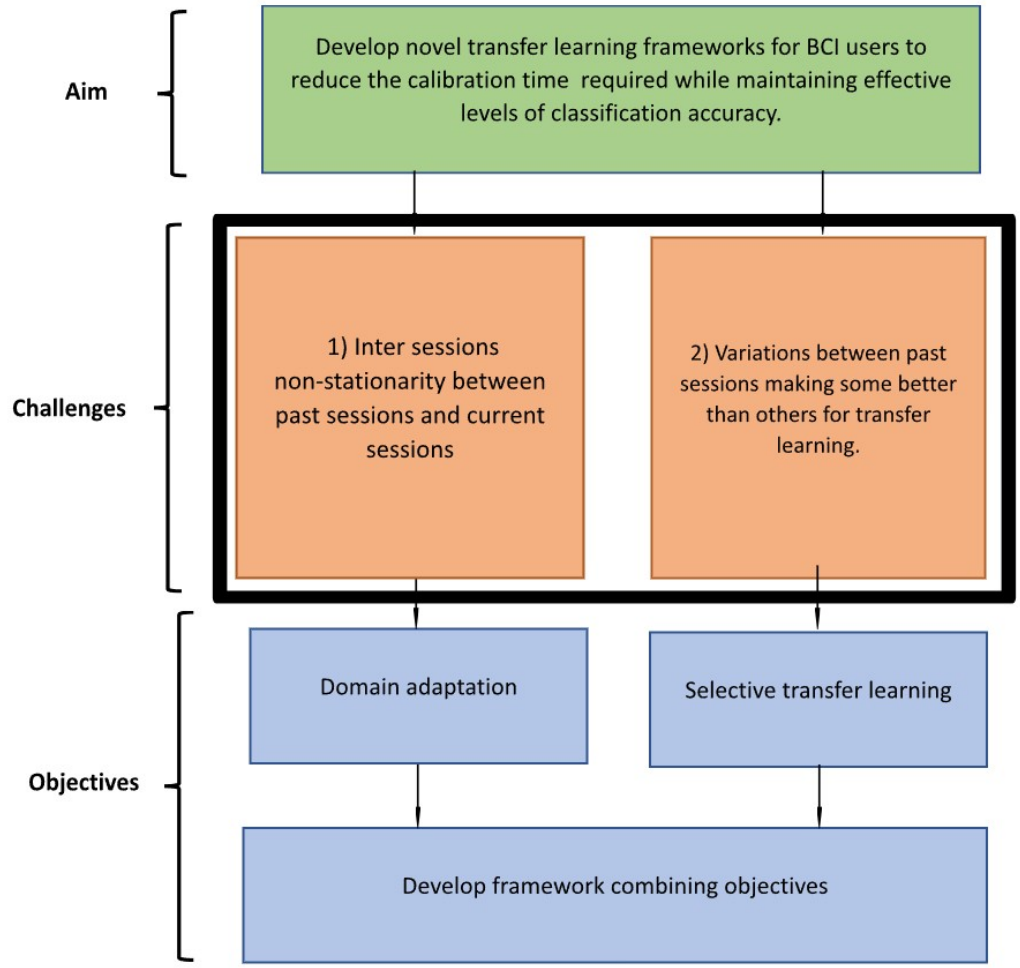

Figure 1.1: The aim, challenges that need to be addressed and objectives of this thesis.

able source sessions, leading some source sessions to be more suitable for training the BCI for the target session than others. In some cases, the source sessions distributions may be similar to the target sessions, while in others completely different. As such it is important to be able to measure these differences and then make sure that the source sessions used to train the BCI have similar distributions to the target session. While, those sessions which have high dissimilarities, and could potentially be detrimental to training the BCI, are given less weights or not used at all.

In some cases, this also leads to none of the source sessions being suitable for the target session. This can be due to none of the source sessions being similar to the target session. Equally, sometimes the target session may also have easily classifiable EEG data with high levels of dissimilarity between the session's two classes. For these sessions, a BCI with high

classification accuracy can be trained with only a few trials. In these cases the addition of more data can lead to under-fitting. Therefore limiting the effect or even avoiding transfer learning can lead to higher classification accuracy despite the lack of data from the target session.

The final challenge is combining the previous objectives to create a complete framework. For this, the final objective is to develop a novel framework. Taking advantage of both domain alignment and selective transfer learning to produce a reasonable level of classification accuracy with only a few target session trials.

To address the above-mentioned aim the following objectives need to be addressed:

1. Develop novel domain alignment techniques to reduce non-stationary and mismatch between the source and target data. Source data refers to the data from past BCI sessions and target data refers to the few trials available from the current session.

2. Develop effective tools to investigate dissimilarities between two classes within the target data as well as dissimilarities between the target and the source data available.

    - Use the developed tools, to develop a decision making algorithm identifying whether or not transfer learning is beneficial for each user.
    - If transfer learning is beneficial, use the developed tools to select and/or weigh the source data for transfer learning.

3. Develop a comprehensive transfer learning framework that:

    - Minimises the nonstationarity between the source data and target data as obtained in objective 1,
    - Takes into account dissimilarity between the data of each source session and the target session, (i.e. objective 2)
    - Optimises the trade-off between the source and target models in the final BCI model.

The developed algorithms are analysed and evaluated through publicly available data as well as data collected from stroke patients throughout their rehabilitation. The advantages and disadvantages of these algorithms are explored along with possible future developments which may help improve them in the future.

1.4 Thesis Overview

This section describes the layout of this thesis and how the work completed in each of the chapters addresses the objectives mentioned above.

Chapter 2 provides an overview of the current literature in the field of BCI research. Initially, this chapter introduces different methods of signal acquisition that can be used for the BCI, the types of neurological signals that can be used to operate a BCI and current state of the art algorithms that are used for motor imagery-based BCIs. Following this the chapter explores the use of BCI for rehabilitation, particularly focusing on how it can be implemented to aid stroke patients with recovery. After defining some of the issues currently limiting the use of BCI for rehabilitation the chapter then introduces and defines a number of topics that may be effective at addressing these issues, namely adaptation and transfer learning. Then the current approaches to adaptation and transfer learning are reviewed and their challenges and limitations are explained.

Chapter 3 proposes a novel linear alignment algorithm that reduces the effects of non-stationarities between the past source sessions and the current target session. The proposed algorithm improves the classification accuracy by applying a linear transform to minimise the Kullback Leibler divergence between the sessions. The proposed alignment algorithm was evaluated using publicly available data and the results were published in the 40th Engineering in Medicine and Biology Conference (EMBC 2018).

Chapter 4 develops a novel selective transfer learning framework. The proposed framework only utilises transfer learning when it is beneficial for the target session and then only selects the best source data for transfer learning. This framework uses session-specific classification accuracy obtained from a small number of target session trials to identify whether or not the user would benefit from transfer learning. If so, it uses a novel measure to identify the source session which would produce the highest classification accuracy. This framework was again tested using a publicly available dataset and published in the 44th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2019).

Chapter 5 presents a novel framework which incorporates the novel linear alignment as well as selective transfer learning in order to produce usable levels of classification accuracy

when only a few target trials are available. The proposed framework is comprised of three parts, initially implementing the linear alignment from chapter 3 to align the source data to the target data, next combining the weighted aligned source data, and finally regularising the weighted aligned source data and the target data to create the BCI model. Combining the novel alignment method with a novel selective transfer learning algorithm proved to be very beneficial with both the approaches complementing each other, producing better classification accuracies with fewer target session trials than either approach alone. This framework was evaluated with a multi session dataset collected from stroke patients as well as a publicly available dataset and the results published in the Frontiers in Neuroergonomics.

Chapter 6 layouts the conclusions drawn from the work completed on this project, laying out the contributions that this work has provided as well as suggestions for the future work which could be completed based on this work.

1.5 Publications during this PhD

1-The conference paper "Data Space Adaptation for Multiclass Motor Imagery-based BCI" was published in the 40th Engineering in Medicine and Biology Conference (EMBC 2018) and presented through a poster.

2-The conference paper "A Subject-to-Subject Transfer Learning Framework based on Jensen-Shannon divergence for Improving Brain-computer Interface" was published in the 44th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2019) and presented through a poster.

3-The conference paper "Weighted Transfer Learning of Dynamic Time Warped Data for Motor Imagery based Brain Computer Interfaces" was published in the 42nd Engineering in Medicine and Biology Conference (EMBC 2020) and presented through a poster.

4-The journal paper "A Transfer Learning algorithm to Reduce Brain-computer Interface Calibration Time for Long-term Users" was published in the Frontiers in Neuroergonomics.

Chapter 2

Literature Review

2.1 Brain computer interfaces (BCI)

Brain-computer interfaces (BCIs) are systems which allow computers to interpret brain signals to control an output. Signals generated by the brain are collected either via either invasive devices, which are placed directly on the brain to collect signals, or non-invasive devices, which collect brain generated signals from the scalp. These BCI then use the collected signals to control a computer.

Brain-computer interfaces generally follow a set route from signal acquisition through preprocessing, feature extraction and classification. The classified signal is often used to control an actuator or provide user feedback. This method of communication is potentially beneficial for users with severe disability who cannot use the available human-machine interfaces (HMI), such as a computer and mouse, joystick etc. [15]. BCI can also be more intuitive than HMI in many applications, such as controlling a prosthetic, as motor imagery can be used directly instead of a non-intuitive controller. BCIs have been used to control a computer mouse, type messages onto a keyboard, control a wheelchair and more recently it has started to be implemented to assist in rehabilitation.

2.2 Types of signal acquisition methods

The imaging devices to acquire signals from the brain can be divided into invasive and non-invasive. Invasive devices tend to acquire less noisy signals from the brain but requires sensors

to be placed inside the body. On the other hand, non-invasive imaging devices acquire brain signals from outside the head, usually the scalp.

Invasive BCI: Invasive BCI is a broad term for any BCI systems requiring the sensors to be implanted either inside the brain or on the surface of the brain[15]. These sensors tend to obtain excellent signals due to their proximity to the brain. However, they require invasive surgery and a high risk of infection and post-surgery complications[15].

Non-invasive BCI: Non-invasive BCIs are much more common as they do not require any implantation, making them safer for the user and quicker to set up [15]. There are several forms of sensors which can be used for scanning the brain without being invasive; however, not all of these are suitable for BCI [15].

Functional magnetic resonance imaging (fMRI) produces a very clear image of the brain and is commonly used in medicine to produce high-quality real-time images of the brain. fMRI is not suitable for BCI, though, due to the requirements of the system itself. To produce the magnetic resonance, a giant electromagnet is required, which is too big to move, has a high cost for sustained use, and the magnet interferes with any metal in the room. This would mean that the BCI would not be affordable or portable and also struggle to interact with the user directly.

Magnetoencephalography (MEG) is another source of signal acquisition that can produce high-quality signals from the brain despite being non-invasive [16]. Compared to electroencephalogram (EEG), it has higher spatial resolution while maintaining high levels of temporal resolution[16]. MEG works by tracking the same electrical currents occurring within the brain as EEG; however, instead of directly measuring the current, the MEG measures the magnetic field created by this current[16]. The improved spatial accuracy is due to the magnetic field not being affected by the skull or scalp as the EEG does encounter this[16]. However, MEG lacks suitability for BCI due to its lack of portability. The magnetic signals collected from the brain are in the order of 10-1000 femtoteslas, which can not be distinguished from the background magnetic noise which is found in an urban setting, which tends to be in the order of 10^8 femtoteslas [17].

Near-infrared spectroscopy (NIRS) uses near-infrared radiation to track the changes in blood flow which occur within the brain [18].

The radiation is able to track the flow of blood very accurately within the brain with good temporal and spatial resolution [18]. Despite this, the NIRS does not work well in BCI. This is due to the blood flow not being a good way of tracking neuroactivity. The blood does flow towards activated cells within the brain, as they require more oxygen, but the blood does not tend to flow to the area until 8-10 seconds after the cell activation [18]. This delay makes a NIRS-based BCI too slow for many practical applications.

EEG is the most commonly used as it can be cheap, portable and have high temporal resolution despite the high level of noise it can encounter and low spatial resolution. More details on EEG are provided in the subsequent section.

2.2.1 Electroencephalogram (EEG)

EEG utilises electrodes placed onto the scalp to measure electrical activity originating from the brain. These electrical signals are conducted through the skull and the scalp where they can then be measured through the EEG [19]. In order to try and standardise the spatial resolution from these signals the international 10-20 system has been established to create standard electrode positions used by all researchers [20]. This electrical activity originates from the action potentials that occur when neurons fire. The action potentials produced by these neurons firing are tiny and hard to detect, being measured in millivolts [20]. As the EEG detects the voltage after it has passed through the dura matter, skull and scalp it can only detect voltage when millions of these neurons in the same area fire simultaneously. When this happens a current is produced that can be detected by the surface electrodes [19]. Due to this requirement, the EEG has a low spatial resolution and can only be used to measure and track major brain activities such as changes in emotion and motor imagery.

The electrodes that detect this surface current can be split into dry and wet electrodes depending on the material used to detect this surface current [21]. Wet electrodes utilise a conductive gel to conduct the electricity from the scalp. This gel ensures that the electrode has an excellent connection to the scalp even when there is hair between the two. Although the gel ensures a good connection it can also lead to a loss in spatial resolution if not applied correctly if it spreads across the scalp. The gel also can lead to an increase in the setup time of the experiment and the clear-up time. Dry electrodes are stuck directly onto the head without the

use of any conductive gel and tend not to perform as well as their wet counterpart in terms of the quality of the recorded signals [22]; This is due to the lack of stable contact between the electrode and scalp as air can get in the way. Despite this they benefit from a sorter setup and clear-up time. No gel has to be injected into the electrodes and the BCI user does not need to wash their air following the experiment.

The signals collected by the electrodes from the scalp, are amplified and then converted from analogue to digital signals so the computer can process them. These signals are then usually filtered to remove significant noise caused by muscle movements, mains electrical interference or other sources of noise [2].

The EEG signals are unique for every person due to the differences between each person's brain. In addition, the EEG signals are very non-stationary as they can be affected by changes in the users' emotions, concentration, fatigue, environment, external factors and illness. Despite each EEG being unique, there are trends, which can be detected [20]. Some of these trends can be identified through changes in certain rhythms in the brain. EEG signals are typically split into the delta, theta, alpha and beta rhythms, as shown below [21]. Sometimes the EEG is further split for more specific uses.

- Delta waves (<4 Hz): high amplitude waves prominent in a deep sleep.
- Theta waves (4-7 Hz): commonly correlated with concentration levels
- Alpha waves (7.5-13 Hz): these waves are affected by the motor cortex and are essential for motor imagery (MI) based BCI
- Beta waves (14-30 Hz): These tend to be weaker than the alpha waves but react to stimuli

The brain is highly complex, and different brain areas often process different information or fulfil different tasks [23]. As such location along with the frequency of the detected voltage is key to understanding the collected EEG. As such the electrode placement can cause many issues if not done correctly, if placed over a different area the previous placement of the signals can be very different. A good example of this can be seen over the motor cortex from the homunculus model, shown 2.1, which gives a vague approximation of which part of the motor cortex controls which part of the body. As such an international system was developed called

the 10-20 system which provides set points to cover the entire brain accurately and effectively on any patient [24].

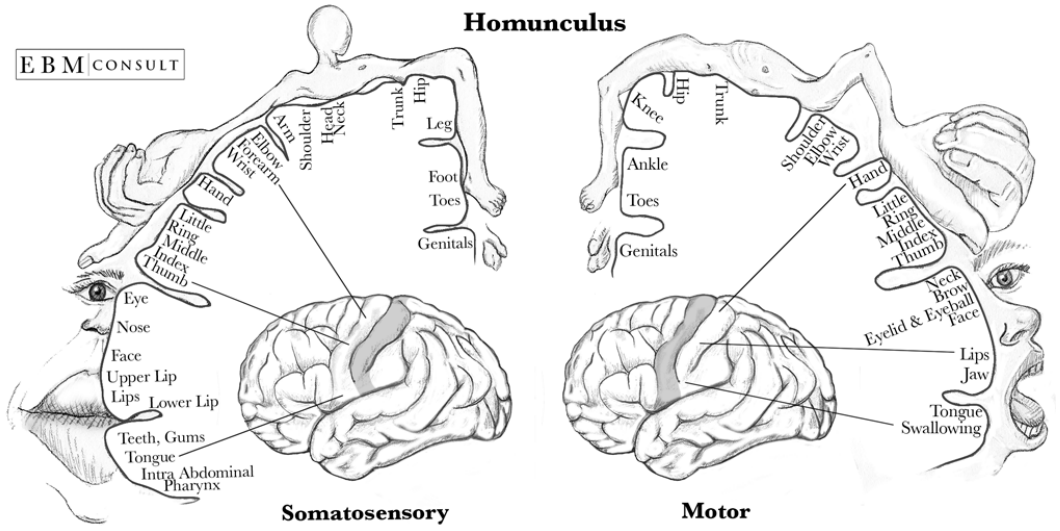

Figure 2.1: The different parts of the body can be roughly mapped to the motor cortex. This is oversimplified however hold true in general [1].

2.3 EEG Control signals in BCI

Once the signals have been acquired specific features have to be pulled by the BCI to use as inputs. These tend to be split into evoked signals, which are created in the brain in response to specific stimuli provided, and spontaneous signals which the user generates themselves [2]. The evoked signals can be generated by a range of stimuli including visual, touch and auditory stimulation, as this stimulation is provided by the researcher it can be easily times and is therefore easier to identify. Spontaneous signals occur naturally when the user of the BCI is focusing on something in particular

The two main evoked signals that are used by BCI are steady-state signals and P300 signals. Steady-state can be stimulated through both visual stimuli and auditory stimuli. By creating a repeating frequency, either with flashes or sound, which is observed by the subject that frequency is replicated within specific regions of the brain and can be detected. By providing multiple sources of stimuli, set to different frequencies, for the subject to focus on a choice can be made by the subject allowing communication [2].

P300 uses the brain's natural tendency to create a wave peak in activity 300 milliseconds after seeing an event on a place being focused on. This event can occur in different forms but is most commonly used by locked-in patients through a P300 speller. This uses flashing rows and columns of letters to allow the patient to spell out words [25].

The user generates spontaneous signals without external stimuli by concentrating on particular thoughts [20]. This can be in the form of understanding emotions but the most common signals used are sensorimotor signals or motor imagery.

Motor imagery signals occur when the subject imagines moving a particular part of their body, this signal created is very similar to the signal generated by the subject when they move their body. As such this has been used in stroke rehabilitation to help patients retrain their brains and improve their control over their bodies. These signals come primarily from the motor cortex and different movements can be differentiated by the location that the signal is originating from [2].

2.4 State-of-the-art algorithms in motor imagery-based BCI

The majority of the BCI systems can be split into pre-processing, feature extraction and classification. The classification is then often used to control some form of actuator or provide some form of feedback [2]. A few BCI are being developed through the use of neural networks that are capable of producing high levels of accuracy without going through these steps; however they require a large amount of data in order to train the system initially.

The signals acquired are initially pre-processed to remove any artefacts that have been generated. The artefacts are primarily electrical signals produced from the muscles around the head being used while some noise also comes from the environment, such as mains noise [26]. These electromyographic signals can cause issues because they are much stronger, generated closer to the surface of the skin, and can spread a large distance from their source. In most cases, this pre-processing is made up of a band pass filter that focuses on the frequencies required by the particular BCI, ignoring any irrelevant frequencies, and an artefact rejecter that removes any trials with very large swings in the amplitude detected. This artefact rejecter

works due to the difference in the magnitude of the electrical currents being detected from the brain, 10-100 micro-volts, and the artifacts [26].

The common spatial patterns technique (CSP) is the most popular form of feature extraction for MI-based BCI is the common spatial patterns technique (CSP). This separates two classes of multi-variate signals by providing weightings which maximize the difference in variance between the classes, leaving one class with high variance and the other with low variance [27]. This technique has been proved effective in a large number of experiments and has been adapted several times to improve the accuracy achieved in a range of scenarios. One of the most recent variations on the CSP is the filter-bank CSP (FBCSP) which splits the incoming signal into multiple components with different frequency bands. The FBCSP then produces weights for each band and selects the best features for the system to use by examining their mutual information [28].

Following the feature extraction the BCI needs to classify what the features mean, this is done through machine learning. Machine learning can be split into two key forms supervised and unsupervised. Unsupervised learning algorithms use data without any labels to learn rules. This is more commonly used for clustering and dimensionality reduction. While Supervised learning algorithms draw predictions and learn rules for new data based on existing labeled data. These can be used for the classification of the new data.

The most commonly chosen classifiers for BCIs are linear discriminant analysis (LDA) and the source vector machine (SVM) [29] [30]. Despite their popularity they do have some flaws which can become an issue due to the non-stationary nature of the EEG signals. Due to this more advanced forms have been developed such as the regularized LDA which has successfully been used to reduce the effects of overfitting [31].

2.5 BCI for stroke patients

BCI has been used in several medical applications to help users who have lost some control of their external environment, most commonly for users experiencing full paralysis with no control over their body [32]. Recently there has been an increase in researching the application of BCI for a number of other medical conditions, such as helping in the rehabilitation of stroke

patients [33].

A stroke is the occurrence of damage to the brain due to a lack of blood. This is a growing concern with more than 100,000 people in the UK having a stroke yearly [34]. The stroke can happen due to a blockage in the blood vessels in the brain, and ischaemic stroke, a blood vessel bursting, or haemorrhagic stroke, both of which can affect the way the body works [35]. These changes can affect the patients' cognitive functions, emotions and ability to control their muscles as damaged areas of the brain either struggle or cannot replicate the signals they were sending before [34].

BCI can be used to assist stroke patients in two key ways. The patient can use them to compensate for the user's lack of neuromuscular control, allowing them to communicate with their environment [33]. This is becoming increasingly popular for patients who have some form of muscle weakness limiting their ability to interact with the outside world effectively. However this application of the BCI is most practical for patients who have full paralysis and have no other option but to use the BCI as it can be uncomfortable when used for extended periods of time. The second application to assist stroke patients is to help provide rehabilitation. The rehabilitation of the stroke patient to improve motor function has traditionally been done through the application of physiotherapy, known as active motor training, to stimulate the activity in the patient's motor cortex; however, recently, more research has been done into BCI rehabilitation [36]. Motor imagery (MI) activates the neurons in the motor cortex, allowing even patients who lack motor control to stimulate the brain through their intentions. Several approaches have been researched using combinations of BCI, physiotherapy and exoskeletons for research [37]. According to motor relearning theory, such active participation is expected to facilitate motor rehabilitation and neuroplasticity in stroke patients [33].

These approaches are being increasingly researched with more compelling evidence of the BCI effectiveness being published. These innovative BCI are helping rehabilitate the patient; however they are far from being optimized, requiring long calibration times before each use and obtaining low levels of accuracy for many users.

2.6 Limitations of motor imagery-based BCI

While relying on motor imagery to control a BCI can be very beneficial, there are still several limitations with motor imagery-based BCI. One of the key limitations of using motor imagery to control and train a BCI is that it is a spontaneous signal instead of evoked [2]. As mentioned previously being a spontaneous signal makes the BCI more intuitive and removes the need for external stimuli. However along with these benefits there are several downsides. The main limitation of spontaneous signals is the reliance on the participant. When trying to collect signals to train the BCI it is impossible to know if the participant is correctly generating the desired signal. Many factors can lead to changes to the signal and lead to an inaccurate model. For example, if the user is distracted or not paying attention they may miss the cue to attempt the imagery or perform the imagery at a much later period than expected [38]. Equally if the participant is not motivated they may allow their mind to drift and not focus on generating the imagery correctly when prompted to. There is also the possibility of the participant misunderstanding the imagery that is required of them. For example, in some studies they have been instructed to imagine moving their hand; however as the exact movement is not specified different movements may be imagined each time [39]. Through careful planning while developing the trials, ensuring the reasons behind the experiment are understood and appreciated, giving the participants plenty of breaks to relax and re-focus and ensuring the imagery instructions are made clear, these limitations can be reduced but never completely removed.

All these limitations due to the reliance on the participant to accurately produce the desired signal can also lead to the need for a long calibration time. These calibration periods are required due to the non-stationary nature of the EEG signals being collected. This is particularly an issue with spontaneous signals where the previously mentioned reliance on the participant. Changes to participants' mental state, fatigue level and surroundings can lead to small changes in the EEG signals being produced and the participants' reaction time which in turn affects the timing of the signals [40]. Unlike evoked signals, such as SSVEP, where we are looking for a specific increase in one EEG frequency in a specific area, motor imagery can be activated differently for each participant. These lead to long periods of calibration being required to ensure a high level of classification accuracy [2].

Even with this calibration period the non-stationary nature of the EEG can cause variations both from session to session and even within a session itself. These inter and intra-session variations can be large enough to reduce the effectiveness of the trained BCI leading to a reduction in classification accuracy [41]. While an initial calibration period can reduce the effect of the inter-session non-stationarities it does not impact the intra-session non-stationarities. These are often minor early on but can become more dramatic as the BCI is used for a longer period [41].

Along with the effects of the non-stationary nature of the EEG signals motor imagery-based BCIs are also limited by the hardware currently developed to collect these signals. Collecting signals originating from the brain using electrodes on the scalp can be difficult with many challenges, such as combating the noise created by the signal passing through the skull and scalp and hair often stopping proper contact between the electrode and scalp [2]. The current best solution to this is the use of wet electrodes, which use conductive gel to ensure a good connection with the scalp; however despite the improvement in signal quality, this presents a new issue. The extra setup and clean-up time the wet electrodes cause practicality issues as it makes the process of using a BCI much longer. This increased setup and clean-up time the wet electrodes create reinforces the need for a reduced calibration time.

2.7 Reducing session to session variations and the long calibration time

2.7.1 Transfer Learning

To reduce the BCI's required calibration time, it is essential to develop accurate BCI models using as few trials as possible. The critical barrier to reducing BCI calibration time is the nature of the EEG signals described previously. As stated, the signals collected from EEG are unique, non-stationary and have high dimensionality [2].

As these signals are unique for each person, it is difficult to use any data collected from another user directly. Although there are often some similar features between users, there are also several differences which make identifying robust and accurate features difficult if

only a limited number of labelled trials are available from the new user. The non-stationary nature of the EEG leads to changes in the EEG both inter and intra-session. These changes in properties of EEG signals can be due to variations in the user's mental state, fatigue, and physical condition of the sensors among other reasons [40]. All these factors can often cause poor performance of the BCI when it is trained using data collected from the user previously.

Along with these factors, the EEG's high dimensionality makes it very difficult to precisely model the users' mental state using only a few trials [9]. It is common to have outliers in EEG signals, either due to muscular noise or the non-stationary nature of the EEG. These outliers can have a dreadful effect on the model, causing its accuracy to drop dramatically. Several studies have been carried out in order to try and address this issue. Utilising only a few labelled trials to try and develop a practical BCI. One of the critical approaches explored is the implementation of transfer learning [9].

Transfer learning is a commonly used machine learning technique, often implemented to train a system when there is only a limited amount of data. When only a limited number of labelled trials are available from one domain that can be used to train the system, transfer learning aims to take valuable data from other domains [42]. Unlike many other machine learning techniques, transfer learning does not presume that the training and test data are from the same feature space with a fixed distribution. This allows for a much more flexible application of the technique with data from similar sources training the model instead of new training data being required each time there is any change in the distribution. Negating the need for new training data after every shift of the distribution can reduce costs and save time. This has made the technique widely implemented in real-world applications such as computer-aided design (CAD), image recognition, Wifi localisation, software defect classification, tracking muscle fatigue and brain-computer interfaces (BCI) [9, 43–45]. In these areas, there are often changes in distribution when completing tasks which would require re-training the model with new data. The application of transfer learning has been particularly interesting in BCIs as the distribution often shifts due to the nature of the EEG signals [42].

When effective transfer learning can significantly reduce the calibration time required for BCIs and increase the overall classification accuracy. To explore this, many studies have ex-

plored different transfer learning approaches for BCI. These studies include applying a range of machine learning techniques to each of the critical components of the BCI. For example, some studies focus on improving the feature space of the target session using the data available in the source sessions, either transforming source space features to better fit with the available target sessions' features or trying to identify aspects common to all the feature domains [46]. Other studies explore using the available source data to improve the classifier, altering the classifier's parameters based on the source data [47]. Finally, some studies also explore instance-based transfer learning where the algorithm assumes the source data cannot be used directly, utilising some parts may provide valuable data [42].

Despite this range of focuses, two critical approaches have been explored to address this thesis's identified challenges. These are the application of adaptation to address the non-stationary nature between sessions, and selective transfer learning, to differentiate between the available source sessions.

2.7.2 Domain Adaptation

Domain adaptation is an approach utilised along with transfer learning where machine learning algorithms can be applied to alter the data from the source sessions or the target session to reduce the dissimilarity between the data. Within BCI, adaptation has often been employed to reduce the effect of the EEGs' non-stationary nature and uniqueness between sources that can lead to issues with the BCI model.

One application of adaptation is to deal with the covariate shift that can occur. This covariate shift the change in probability distribution within the data, which can occur due to the non-stationary nature of the EEG signals being used [48]. In [49], an adaptive BCI utilising a covariate shift monitoring algorithm was proposed to combat the covariate shifts that can occur due to the EEGs non-stationary nature. This utilised a covariate shift monitoring algorithm to identify when a shift occurred through an exponentially weighted moving average (EWMA) control chart, which combines current and historical data to detect any changes in the time series of data. After this algorithm has detected a shift, it is then validated with a multivariate Hotelling's T square statistical hypothesis test. If the p-value is below 0.05, the shift is validated, and the BCI adapts the classifier, re-training it using the previously correctly

classified trials. This approach proved very effective at identifying when covariate shifts occur in real-time, and the adaptations led to a statistically significant improvement in the classification accuracy. However, this continuous adaptation approach focused on an online algorithm with trials being labelled after each trial was classified. This is not practical in a real-world application as identifying if the classification of the trial was correct requires human intervention.

Similar approaches focus on addressing the covariate shift that can occur within transfer learning [48]. Focusing on countering the covariate shift between source sessions and the target session. For example, adaptive classifiers such as the Bagged Importance weighted LDA also utilises adaptation in a similar process. However, instead of utilising manually approved trials, it takes a small number of labelled trials from the target session to estimate the shift [48].

Another adaptation approach that has been explored is aligning the EEG trials available within the source data in the euclidean space [50]. Unlike the previous approach, where the adaptation was used to reduce the effect of the covariate shift, this alignment is done to identify robust features. By aligning the source data trials, it is possible to reduce the euclidean distance between the sessions in the data space domain. This approach significantly improved the classification accuracy achieved through transfer learning.

Despite the benefits of these approaches, they only address a few of the issues that occur within EEG-based MI BCIs. The use of adaptation can reduce the dissimilarity between different sessions of data, limit the effects of the EEGs non-stationary nature and identify robust features for training; however, it does not address the differences within the source data nor account for the range of variations that can be present in the target sessions. Even after adaptation has been applied, there are many cases where data in the source session is detrimental to training the BCI.

2.7.3 Selective transfer learning

One area of transfer learning which has been explored can be called selective transfer learning. With many types of data it has often been found that not all the available source data benefits a model, even if the data is collected from a very similar source. This can be particularly true for EEG data due to its non-stationary nature, which can cause changes in the data distribution

even when collected from the same subject on the same day. In order to reduce the effects of this detrimental data, selective transfer learning aims to remove any source data that could be detrimental or limit its impact. Previously this approach has been explored in a range of applications such as malware classification, image classification and financial fraud detection [51–53]. In these applications the transfer algorithms proposed attempt to reduce the effect of any detrimental data. This is done either through weighting the source data [54], excluding certain parts of the source data from being used [55] or regularising the source data using the available labelled trials from the target data [56].

Recently studies have started exploring these selective transfer learning approaches to improve the classification accuracy of BCIs. The application of weights to differentiate between source sessions based on their similarity to the target session has been examined multiple times with different measures applied to evaluate the similarity between the source sessions and the target sessions. For example, in [54], Azab et al. utilise Kullback-Leibler (KL) divergence to measure the similarity between the target session and source sessions. This approach led to a statistically significant increase in the classification accuracy when there were only a limited number of target session trials. In particular the proposed weighting approach benefited the target sessions where only poor classification accuracy could be achieved when trained with just the target data.

A different approach that can be used to adjust for the variations between the sessions is regularisation. In [56], Lotte et al. utilise regularisation parameters while combining the labelled target data and the available source sessions. By weighting the source sessions to account for the variations between them and including the target data, it became possible to produce a more stable covariance matrix. This performed than previously proposed state-of-the-art algorithms however failed to take into account the non-stationarities that can still occur between each session's data.

2.7.4 Challenges of transfer learning

Although many studies have now been completed exploring different possible applications of transfer learning to improve BCIs, there are still several limitations and challenges that have been identified throughout this chapter. One fundamental limitation which seems

familiar throughout these papers is the focus on inter-subject transfer learning and the lack of exploration of inter-session transfer learning. This focus is understandable as collecting data from a subject during multiple sessions over an extended time is much more complicated than collecting it from just one or two sessions. However, inter-session transfer learning is a key area which needs to be explored. In the majority of real-world applications for BCI technology, the users will be using the BCI for multiple sessions over an extended period. In these cases, it is essential to recognise the potential benefits that the data collected can provide. When data collected from one user is used to train a model for a new user, the classification accuracy is often low, yet studies have found that using the user's old data can be beneficial when enough is available [57]. This highlights the difference between the inter-session and inter-subject transfer learning methods that need to be explored. This issue is magnified when this is further examined to create a BCI applicable in the real world. One of the critical variables between inter-subject transfer learning for BCIs is the unique nature of the EEG signals created due to the unique make-up of each user's brain. In a large number of rehabilitation applications of the BCI, the users will have encountered some form of trauma to the brain. This trauma can permanently alter the brain and lead to activation patterns which differ significantly from the norm. In these cases, it can be expected that a large proportion, if not all, of the available inter-subject transfer learning data, will be detrimental to the training of the BCI model.

While reviewing the current transfer learning approaches that are being used to train BCI one of the common themes that seemed to occur was the requirement of a large number of labelled trials being required from the target session. The transfer learning techniques were able to take advantage of the valuable data available in the source sessions and improve the classification accuracy, but a large number of calibration trials did not reduce the calibration time required by the BCI. Reducing this calibration period is key to making the BCI a real-world practicality, as users often only have a limited time to access the BCI. The more time spent on re-calibrating the BCI to work with the current target session the less time there is for actual rehabilitation.

A final challenge that appears to be overlooked is exploring the issue of the BCI as a whole to allow the BCI to work in the real world. Many of the studies exploring selective trans-

fer learning proposed algorithms to address the variations present between source sessions and address them with different weights, while others proposed algorithms to account for the differences between the source and target sessions and even adaptation algorithms that help to reduce the effects of non-stationarity between sessions have been studied. Despite these highlighting potential growth areas, they often turn a blind eye to the other issues. Identifying which source sessions will work best for the new target session is crucial, but it is also important to try and reduce this difference. A lot has been developed to resolve the various limitations of BCI, but they tend to focus on one approach. Research into how these different methods of improving the BCI can work together and help each other account for issues the algorithms were not designed to handle on their own needs to be conducted. Instead of looking at the various issues individually, it is essential to address the situation as a whole to achieve a practical BCI that works well in the real world with little to no calibration between sessions.

2.8 Conclusion

This chapter has provided a general background and exploration of currently used approaches within BCI. Initially, a brief explanation of what BCIs are was provided, along with the brain signals commonly collected to control the BCI and the types of mental states which can be detected. From this research, EEG-based motor imagery BCIs were focused on due to the high temporal resolution of the EEG obtained non-invasively and the practicality offered through motor imagery (MI). Following this line, state-of-the-art algorithms currently implemented in MI-based BCIs were briefly examined. Then stroke rehabilitation, a possible application where MI-based BCI would be able to provide practical, real-world benefits to a large number of people, was explored. While this review found that the current EEG-based MI BCI provided many benefits, fundamental limitations must be addressed.

This chapter then explored transfer learning as a potential solution to many limitations. A definition of transfer learning was set out, and several approaches were examined. From these approaches, domain adaptations and selective transfer learning were identified as crucial areas and focused on, with various studies applying this technique being explored in detail. Finally, the chapter laid out some limitations of the current work that has been completed and found

vital areas that could be explored during the project.

Chapter 3

Multiclass Data Space Alignment to Reduce Session to Session Non-stationarity in Motor Imagery-based BCI

The findings presented in this chapter have been previously published in the 40th Engineering in Medicine and Biology Conference (EMBC 2018) and presented through a poster [58].

3.1 Introduction

As previously mentioned, one of the main issues affecting practicality of BCIs is the non-stationarites between different sessions [59]. These non-stationarities can cause the properties of the EEG signals to change considerably between sessions. Thus, a BCI trained with one session to classify a different session can have a very low classification accuracy, even if both sessions are from the same participant.

In order to reduce the mismatch between the source and target EEG data, caused by non-stationarities inherent in EEG data, various forms of adaptation and alignment have been researched [11, 12]. Using a limited number of trials from the target session, adaptation techniques reduce the effects of differences between the target and source data by adapting param-

eters of the trained BCI model. This can include adjusting the feature selection method or the classifier parameters based on the target session [60, 61]. Similarly, alignment algorithms use the trials from the target session to align the source and target data. Although aligning the BCI does require a calibration period the number of trials is much lower than the number of trials required to completely retrain the BCI.

A lot of research has been done to develop optimal adaptation or alignment techniques to improve the classification accuracy without requiring the whole BCI to be retrained. This includes research into adaptive feature extraction and classifiers such as implementing an adaptive linear discrimination analysis classifier which updates the global mean [62] referred to as pooled mean linear discrimination analysis. This research highlights the improvements of using adaptation with transfer learning over simply relying on naive transfer learning alone. This also allows the BCI to use previous data and not have to completely re-train the BCI each time it is implemented [8, 62–64]. Despite this, the focus of these alignment and adaptation techniques are often limited, tending to focus on binary class BCIs and being limited to some specific BCI models.

Recently there has been a shift from binary class towards multiclass BCI systems. This is due to the opportunity they would present by drastically increasing the information transfer rate (ITR). Multiclass BCIs have the potential to allow faster communication with the user as well as control of complex actuators, providing more degrees of freedom. Several research studies on multiclass BCIs focus on optimising different BCI components such as feature extraction techniques [65] and classifiers [64].

Some adaptation techniques, initially developed for binary BCIs, have been modified to be applicable in multiclass BCIs. For example, pooled mean linear discrimination analysis [62] has been altered to multiclass pooled mean linear discrimination analysis (MPMLDA)[66] allowing it to work within a multiclass setting. Another adaptive classifier, based on an enhanced Bayesian linear discrimination analysis [67], has also been developed for multiclass BCIs. These altered adaptation techniques have proven to be effective at reducing the fall in accuracy caused by the non-stationary nature of EEG. However, these adaptation techniques require to be implemented by building on a particular feature extraction technique or classifier.

This limits the BCI from future development as the adaptation method may not be applicable to a different feature extractor or classifier.

Data space alignment (DSA) is a method of changing the data distribution directly before it has gone through feature extraction or classification [46]. This method minimises the distribution difference between the BCI training data and the testing data using a linear transform. This means that DSA is not restricted by any particular feature extraction techniques or classifiers. In this chapter, DSA is modified so it can be applied to multiclass BCIs aligning the data with only a few labeled trials from the testing data. Delaying the test data but to a limited extent. The number of classes does not affect the previously proposed unsupervised DSA algorithm, however, the supervised DSA does require altering due to the change in the number of classes. In this chapter, the proposed multiclass data space alignment (MDSA) algorithm will be evaluated using BCI Competition IV dataset 2a [68]. The proposed MDSA will then be compared to two other state-of-the-art multicalss adaption methods, namely unsupervised DSA [46] and MPMLDA [66], providing an evaluation of the algorithm's ability to improve multiclass BCI performance.

3.2 Methodology

3.2.1 Proposed Multi-class Data Space Alignment (MDSA)

The proposed MDSA is an extension of the supervised binary data space adaptation (DSA) [46] allowing the alignment method to be incorporated into a multiclass BCI. Using a linear transform, MDSA alters the target EEG data (i.e. testing data), after it has been band-pass filtered, so it is as similar to the available source data (i.e. training data).

Assume the source data is defined as $\hat{\mathrm{D}} = (\hat{\mathbf{X}}_i, \hat{y}_i)_{i=1}^{\hat{n}}$, where $\hat{\mathbf{X}}_i \in \hat{\mathbf{X}} \subset R^{ch \times t}$ is the i^{th} recorded source trial with *ch* being the number of channels and *t* representing the time sample. $\hat{\mathrm{y}}_i \in \hat{\mathrm{Y}} \subset \mathbb{R}$ represents the corresponding class label. In order to perform the alignment a small number of trials are gathered from the target session and used to build an estimate of the target session probability distribution. The target data is presented as $D = (\mathbf{X}_i, y_i)_{i=1}^{m}$ where $\mathbf{X}_\mathrm{i} \in \mathbf{X} \subset R^{ch \times t}$ is the i^{th} recorded target trial and $y_i \in \mathbf{Y} \subset \mathbb{R}$ represents the class labels of the corresponding trials.

The proposed MDSA aims to use a linear transform, $\mathbf{V} \subset \mathbb{R}^{ch \times ch}$, to minimize the distribution difference between the source data and the target data. The ideal goal of $\mathbf{V}$ is to have the aligned target data, represented by $S(\mathbf{V^T X}, \mathbf{Y})$, to have the same data distribution as the training/source data. By matching the distributions, the feature extractor and classifier trained using the source data should still perform well on the target data.

In order to calculate the optimum $\mathbf{V}$ a few characteristics of the source and target data distributions must be known. The normalised covariance matrix of the EEG data can be estimated as shown in (3.1), where tr is the trace function, known as the sum of the diagonal elements of the matrix. Please note the mean of EEG trials is zero due to the EEG signals being band-passed. The distribution of the EEG data can be modelled as Gaussian based on the maximum entropy principle [69] with zero mean and the covariance matrix calculated as (3.1).

$$\mathbf{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{X}_i \mathbf{X}_i^T}{\mathrm{tr}(\mathbf{X_i X_i^T})} \tag{3.1}$$

Let us assume the Gaussian distributions of two datasets are presented as $N_0(\mu, \mathbf{\Sigma})$ and $N_1(\hat{\mu}, \hat{\mathbf{\Sigma}})$ with $\hat{\mu}$ and μ representing the means of the distribution while $\hat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$ co-variances. The difference between these two Gaussian distributions can then be calculated using the Kullback Leibler criteria [69] as shown in (3.2), where k is the dimension of the data and det refers to the determinant function.

$$KL[N_0 \parallel N_1] = \frac{1}{2}[(\hat{\mu} - \mu)^T \hat{\mathbf{\Sigma}}^{-1} (\hat{\mu} - \mu) + \mathrm{tr}(\hat{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma}) - \ln(\frac{\det(\mathbf{\Sigma})}{\det(\hat{\mathbf{\Sigma}})}) - \mathrm{k}] \tag{3.2}$$

To find the optimum $\mathbf{V}$ for supervised alignment, the KL divergence is calculated for the training and testing data of each class separately. In order to minimise the total loss function across all the classes, the differences are summed before the $\mathbf{V}$ is calculated. The transformed test data distribution is defined as $N_t(0, \mathbf{V}^T \mathbf{\Sigma}^c \mathbf{V})$ and training data distribution as $N_s(0, \hat{\mathbf{\Sigma}}^c)$ for class c in (3.3) while cl is used to represent the total number of classes in the BCI.

$$\min L(\mathbf{V}) = \min \sum_{c=1}^{cl} \frac{1}{2}[tr(\hat{\mathbf{\Sigma}}^{c-1} \mathbf{V}^T \mathbf{\Sigma}^c \mathbf{V}) - \ln(\frac{\det(\mathbf{V^T \Sigma^c V})}{\det(\hat{\mathbf{\Sigma}}^c)})] \tag{3.3}$$

To find the optimum $\mathbf{V}$ that minimises L given in (3.3), the first derivative of L is calculated with respect to $\mathbf{V}$ and set to zero, as shown in (5.6).

$$\frac{dL}{dV} = \sum_{c=1}^{cl} \frac{1}{2}[2\text{tr}(\hat{\mathbf{\Sigma}}^{c\,-1}\mathbf{\Sigma}^{c}\mathbf{V}) - 2\text{tr}(\mathbf{V}^{-1})] = 0 \tag{3.4}$$

$$\mathbf{V} = cl^{-0.5} \sum_{c=1}^{cl} (\hat{\mathbf{\Sigma}}^{c\,-1}\mathbf{\Sigma}^{c})^{\dagger 0.5} \tag{3.5}$$

In (3.5) † represents the pseudo-inverse. Using (3.5) the optimum $\mathbf{V}$ is calculated and then applied to the band-pass filtered test trials. The features of the linearly aligned test data is then calculated and classified using the feature extractor and the classifier previously trained using the training data.

3.3 Experiment

3.3.1 Dataset

The dataset used in this chapter is the publicly available dataset, BCI Competition IV dataset 2a [70]. This dataset contains EEG data from nine users who each completed two sessions, each containing six runs, on different days. Each run consists of 48 trials containing 12 trials from each of the four classes making a total of 288 trials from each session. The four classes are all variations of motor imagery with the user imagining the movement of their right hand, left hand, both feet or tongue.

3.3.2 Data processing

To examine the alignment capabilities of the different techniques, the first session was used to train the BCI model with common spatial patterns (CSP) for feature extraction and LDA for classification. Subsequently, the second session was used as the testing data. To evaluate the alignment and adaptation methods, the first 80 trials of the testing session were set aside for alignment and not included in the results. The same pre-processing and feature extraction was performed for each of the alignment methods. The training data was band pass filtered from 8Hz to 35Hz and a pair wise CSPs was trained for 6 class pairs was used for feature extraction.

In this study, MPMLDA [66] is one of the two methods used for comparison. This method adapts linear discriminate analysis (LDA) classifier used in the multiclass BCI by updating the

global mean, $\mu_{i,j}$, of each of the pair-wise LDAs as new trials are classified. The change caused by the new data, x, is weighted by the probability of the new data belonging to a relevant classes for the LDA as shown in (3.6). Here i and j represent the two classes that the LDA is classifying, $P_i(x)$ and $P_j(x)$ are the probabilities of the new data belonging to the class i and j respectively. β is the learning rate, set to 0.03 as suggested in [66]. The updated global mean, $\mu'_{i,j}$, is then utilised to recalculate the LDA before the next trial is classified.

$$\mu'_{i,j} = (1 - (P_i(x) + P_j(x))\beta)\mu_{i,j} + (P_i(x) + P_j(x))\beta x \tag{3.6}$$

The unsupervised DSA technique (DSA-US) does not require altering due to the fact it is independent from the classes, relying only on the distribution of the entire EEG data. As such it is used as a second comparison for the proposed MDSA. This method also uses a linear transform to align the test data to the trained data after it has been band pass filtered. However it does not split the data into its classes. DSA-US uses all the data at once to calculate the optimum linear transform as shown (3.7), where $\hat{\Sigma}$ and Σ respectively represent the average covariance of the available trials from the training and test sessions.

$$\mathbf{V}_{unsupervised} = \hat{\Sigma}^{-0.5}\Sigma^{0.5} \tag{3.7}$$

3.4 Results and Discussion

3.4.1 Classification Accuracy

Fig. 3.1 shows the increase in accuracy for each of the different alignment algorithms across the different number of trials used for alignment. As shown in Fig. 3.1, compared to the base BCI design without any alignment, all the three examined alignment algorithms improved the average classification accuracy of the test data. MPMLDA does not require any initial test trials to calculate the alignment parameters as it updates the global mean after every new trial added to the test data. This is a form of continuous adaptation thus, the MPMLDA accuracy presented in Fig. 3.1 is fixed across the x-axis.

Initially, when only ten trials per class are used for alignment, there is a very little difference

between the accuracies of the three alignment algorithms. The limited number of trials may have restricted the accurate estimation of the alignment parameters in both MDSA and DSA-US as the estimation could be easily distorted by artefact corrupted trials. By increasing the number of trials per class to 15, DSA-US slightly outperformed the MDSA algorithm. In this case estimation of covariance matrices of test data was based on 60 trials in DSA-US compared to 15 trials in the proposed MDSA. Having more trials for estimating covariance matrix in DSA-US could have led to a better estimation of alignment matrix.

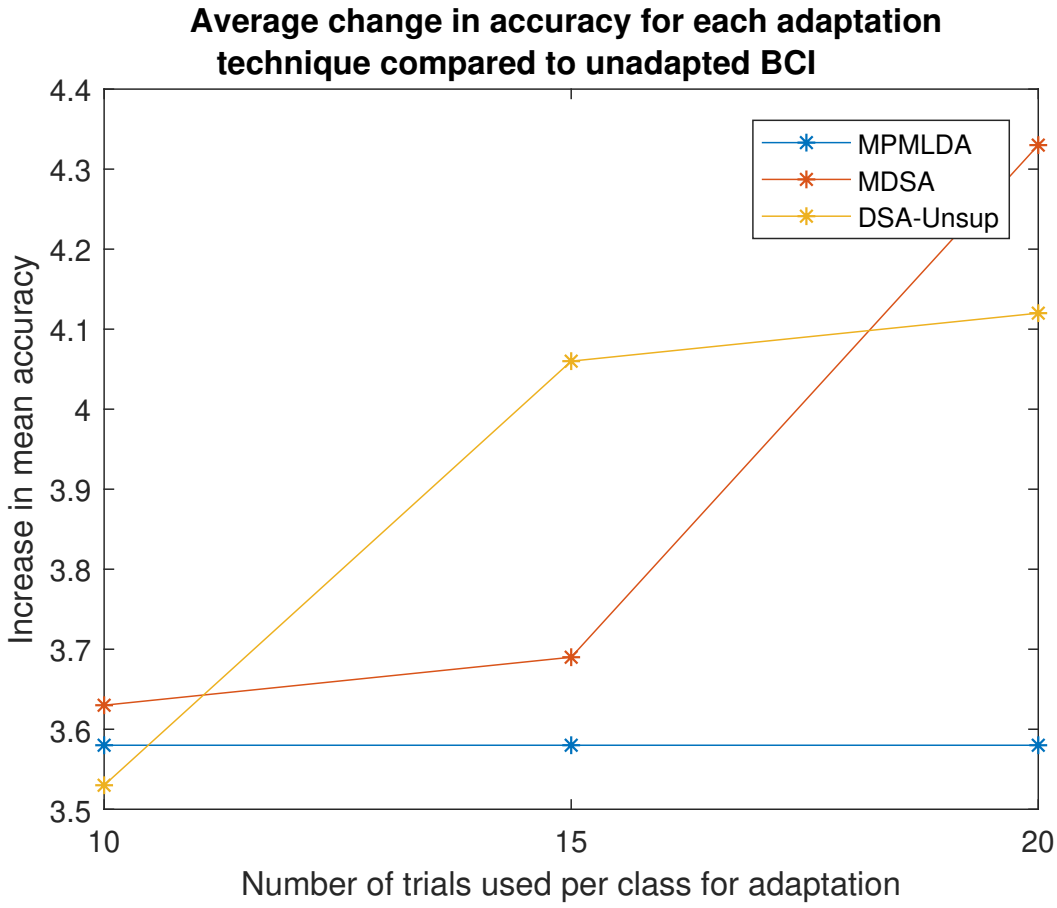

Figure 3.1: Average improvement in accuracy for DSA-US, MPMLDA and the proposed MDSA compared to no alignment, using different number of trials per class for adaption. When 20 trials per class are used for alignment the proposed MDSA outperforms the other techniques.

In the case of 20 trials per class being provided for alignment, MDSA outperformed DSA-US, possibly due to each class having enough trials to estimate an accurate distribution of the data. This does highlight one of the issues of MDSA, as it requires more trials than its unsupervised counterpart to produce its best level of accuracy. However, it can also produce

higher accuracy levels if a relatively large number of trials are available. This could be due to the MDSA creating representative data distributions for each class for the alignment while still recognising changes in the EEG signals relatively quickly unlike DSA-US. The DSA-US algorithm uses the 80 previous trials for each calculation of the linear transform, so if the user's EEG signals start to change, due to fatigue or changes in their mental state, it takes a while to be seen by DSA-US as the 79 other trials dilute the change. This problem is not very pronounced in MDSA due to the trials being split by class so the change is only diluted by 19 other trials per class.

3.4.2 Improvement in Classification Accuracy across Different User Groups

The plots presented in Fig. 3.2 compare the classification accuracies of the proposed MDSA, DSA-US and MPMLD. As shown in Fig.3.2, the proposed MDSA and DSA-US both outperformed MPMLDA when implemented with users who were able to achieve levels of accuracy above 70%. Users 1, 3, 7 and 8 all achieved better accuracies when DSA-US or MDSA were applied compared to MPMLDA. The only user who achieved accuracy higher than 70% and performed best with MPMLDA was user 9. Conversely, the users with low levels of accuracy found MPMLDA most effective at improving their accuracy in all cases except for user 5. Fig.3.2 also displays that the MDSA outperformed the DSA-US in 66% of users, excluding users with less than 1% difference between the two algorithms. Suggesting that although all the alignment algorithms are capable of improving the average accuracy of the BCI, MDSA and MPMLDA outperformed DSA-US when used with users encountering BCI deficiency.

In Table 3.1, the users were grouped into two groups based on their accuracy without alignment, i.e. either above 70% accuracy or below 70%. As shown in Table 3.1, on average MPMLDA and MDSA perform similarly for subjects with accuracies less than 70%, while DSA-US is less effective for this group. This suggests that MDSA is useful as MPMLDA when implemented to reduce BCI deficiency. High accuracy users see little improvement from MPMLDA while DSA-US and MDSA both perform equally well. Suggesting that MDSA has a good overall increase in accuracy for users who obtain high levels of accuracy and those encountering BCI deficiency.

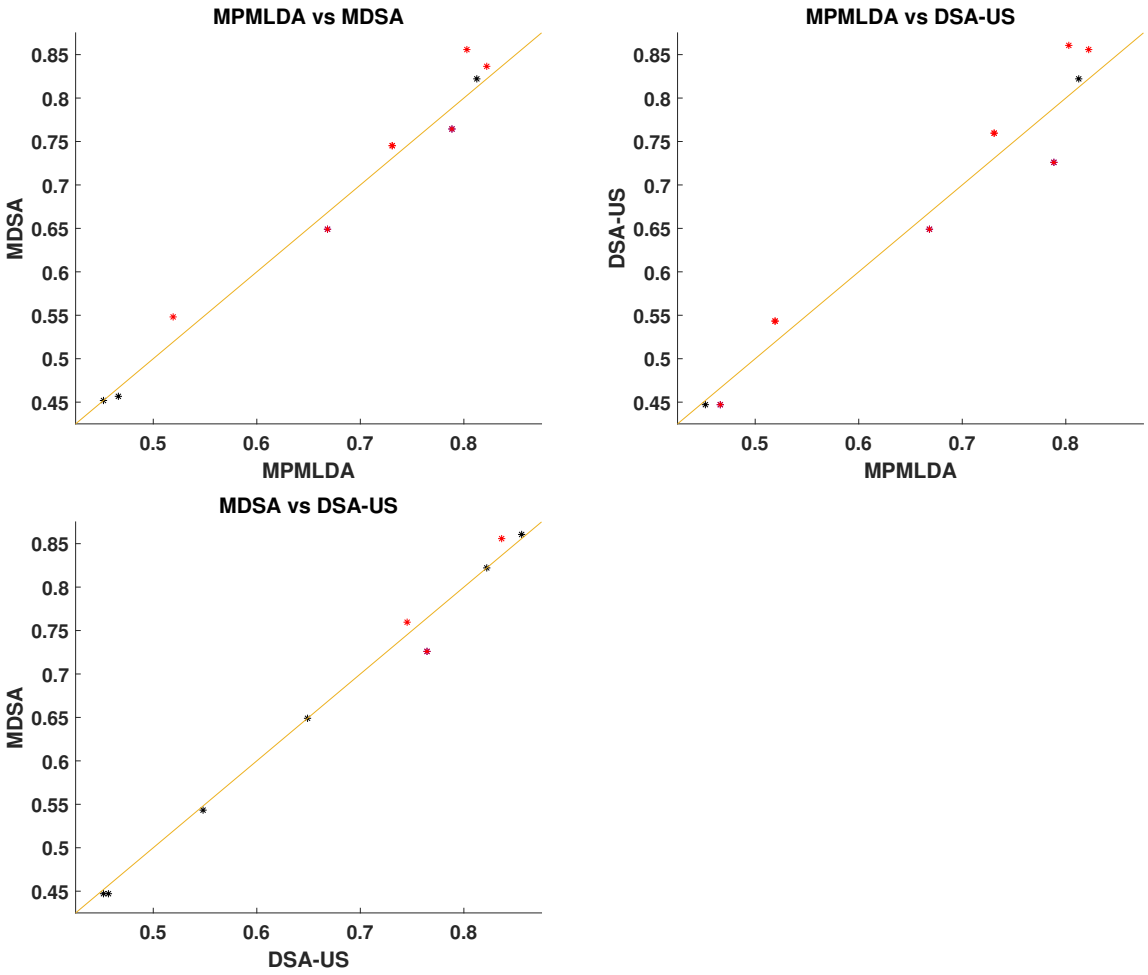

Figure 3.2: Scatter plots comparing the classification accuracies of the different alignment algorithms; Each subject is presented with a dot with black dots being used if the difference between the techniques being less than 1%. Having the dot on the left hand side of the line means the alignment technique on the y axis works better for the corresponding subject. 20 trials per class were used for estimating transformation matrix for DSA-US and MDSA.

3.5 Conclusion

The objective for this chapter was to develop a novel domain alignment technique to reduce non-stationary and mismatch between the source and target data. This alignment method had to be flexible, allowing for varying numbers of classes and not being reliant on a particular feature extractor or classifier, so that it could be easily incorporated into the project's future research. The proposed MDSA occurs in the data space making it independent from the feature selection and classification used by BCI. It has shown to be effective in improving the

<70%			>70%		
MPMLDA	DSA-US	MDSA	MPMLDA	DSA-US	MDSA
4.44%	3.96%	4.44%	2.89%	4.23%	4.23%

Table 3.1: Average accuracy improvement for each technique for users encountering BCI deficiency (Below 70% Without alignment) and users with good accuracy (Above 70% without alignment)

accuracy of multiclass BCIs, capable of outperforming MPMLDA and DSA-US when enough data is provided. However, the improvement is not statistically significant. Despite the proposed MDSA not showing significant improvement over the other algorithms, it did improve accuracy for both users who were proficient with BCIs and users encountering BCI deficiency, unlike the DSA-US or MPMLDA. This range of effectiveness suggests that although the overall accuracy improvement was not statistically significant, the MDSA alignment could apply to a wide range of sessions. Thus, combining this alignment method with other transfer learning techniques could easily be explored.

As transfer learning often utilises multiple source sessions, it would be more practical to align the source sessions to the target session instead of aligning the target session to source sessions. As the target session would only be able to align to either a single source session or an average of the available source sessions, it would be possible to align each source session to the target session. This could reduce the number of calibration trials required while maintaining and increasing the classification accuracy.

A second possible improvement to this algorithm would be adjusting it so that it works on a continuous manner similar to the MPMLDA used for comparison. Through updating the alignment with every new trial it would allow for the proposed MDSA to try and limit the effects of intra-session non-stationarity along with the inter-session non-stationarity it currently addresses. Implementing this would also remove the short period of collecting trials at the start of the session for the alignment. This would require much more computational power as the alignment would have to be adjusted in real time as the new data is received but could further reduce the mis-match between the model and target data.

Interestingly while exploring these different transfer learning algorithms it became clear that for some of the users using other subjects data was detrimental. Some users were able to achieve a much higher classification accuracy using a model trained with only a couple of trials of their own data than one trained with any of the available source data.

Chapter 4

A Subject-to-Subject Selective Transfer Learning Framework for Improving BCI

The findings presented in this chapter have been previously published in the 44th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2019) and presented through a poster [71].

4.1 Introduction

In order to improve the accuracy, research has been carried out to improve the components within the BCI [65]. These range from feature extractors, such as the filter-bank common spatial patterns algorithm [72], to classifiers, such as the adaptive linear discriminant analysis (aLDA) which updates the classifier parameters when new trials are available [62]. A range of other adaptation methods have also been explored in BCI to further improve the classification accuracy. An example is the multiclass data space alignment (MDSA) algorithm, proposed in the previous chapter. MDSA reduces the difference between the training data and test data through a linear transform [73][58]. Despite these techniques improving accuracy, they often require a large number of calibration trials to provide a significant improvement.

To reduce the need for the long calibration time, transfer learning between sessions and

subjects has been investigated. Transfer learning often refers to a procedure of using a data set from a different but related task to improve the accuracy of a new task [74]. When used for BCI, the data sets are often from the same task but different users.

One form of transfer learning is through identifying features which are stationary across multiple subjects, known as domain adaptation. This area has been explored by Lotte and Guan [75] and Kang et al [76] [61] with some success. The other form of transfer learning is called rule adaptation which attempts to find the framework of classification rules. The rule adaptation-based transfer learning attempts to select the most appropriate feature extraction and classification rules from a pool of available components [74]. This area of transfer learning has has been recently explored by He and Wu [77].

These various approaches to transfer learning have been effective for a large proportion of the sessions examined however many actually see a decrease in the classification achieved. Increasing the pool of data being used often allows the trained model to avoid being over-specialised however for some sessions this extra data actually has a detrimental impact []. Some sessions perform best when all the source data comes from the same participant. Meanwhile other times the session performs best when trained with data from the session itself even when the amount of this data is very limited. Due to this in order to maximise the classification accuracy achieved by the algorithm it's important to identify these sessions.

In this chapter we propose a framework to make use of selective transfer learning and rule adaption transfer learning. First, a method to identify sessions that will not benefit from transfer learning is explored. After the sessions which don't require transfer learning are identified their own data is used to train the model. Otherwise, if the sessions would benefit from transfer learning a new measurement of similarity, named the Jensen-Shannon ratio (JSR), is implemented. This measure is used to compare calibration trials with existing data sets for transfer learning. Finally, the data set with the highest similarity to the trials of the target user is selected, from previously recorded data, for training a BCI model for the target user.

The proposed framework will be evaluated using the publicly available BCI Competition IV data set 2a [68]. The algorithm will then be compared to utilizing only the Kullback Leibler (KL) divergence for data set selection and a framework previously proposed by Lotte using

other subjects data to alter the co-variance and mean of the training data set [78].

4.2 Methodology

4.2.1 Proposed Selective Subject to Subject Transfer learning Framework

The proposed framework consists of two steps as follows:

4.2.1.1 Identifying Participants Who Would Benefit from Transfer Learning

Users who encounter BCI deficiency can benefit substantially from the application of transfer learning. While for other users, who easily obtain high classification accuracy, the transfer learning can be detrimental. To counter this the proposed framework identifies the users who can benefit from subject to subject transfer learning then selects the best previously recorded data set for these users to train their BCI models. To identify the subjects requiring transfer learning the leave-one-out validation (LOOV) accuracy is applied on the few subject-specific target trials. If the average accuracy for those subject-specific trials is below 70% they are identified as BCI deficient. For user who are found to encounter BCI deficiency the proposed JSR was then used to select an appropriate data set for training the BCI.

4.2.1.2 Proposed Jensen Shannon Ratio to Select source data

The proposed JSR measures the difference of the average EEG signals from the same class between users and the opposing classes using the Jensen Shannon divergence. The JSR is then used to select an appropriate signal for training, where the same classes are similar and opposing classes are far apart. The Jensen Shannon divergence is based on the Kullback Leibler (KL) divergence with some useful differences.

$$KL[N_j \parallel D_j] = \frac{1}{2}[(\overline{\mu}_j - \mu_j)^T \overline{\mathbf{\Sigma}_j}^{-1}(\overline{\mu}_j - \mu_j) + \mathrm{tr}(\overline{\mathbf{\Sigma}_\mathrm{j}}^{-1}\mathbf{\Sigma}_\mathrm{j}) - \ln(\frac{\det(\mathbf{\Sigma}_\mathrm{j})}{\det(\overline{\mathbf{\Sigma}_\mathrm{j}})}) - \mathrm{k}] \tag{4.1}$$

The band pass filtered EEG signals can be modelled as Gaussian distributions. The similarity between two Gaussian distributions can be measured through the KL divergence, as shown in (4.1). For this equation $N_j(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $D_j(\overline{\boldsymbol{\mu}}, \overline{\mathbf{\Sigma}})$ are used to represent the distributions

of class j from the target subject N and training subject D. $\overline{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}$ represent the means of the distribution, and $\overline{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}$ denote covariances.

Jensen Shannon divergence is an extension of the KL divergence. This extension provides a symmetric and finite value for the similarity by measuring to a middle point providing, as shown in (4.2). The middle point $M_{ji}(\boldsymbol{\mu}_{ji}, \boldsymbol{\Sigma}_{\mathbf{ji}})$ is calculated from the average of the two distributions being compared, with $\boldsymbol{\mu}_{ji} = 0.5(\boldsymbol{\mu}_j + \overline{\boldsymbol{\mu}}_i)$ and $\boldsymbol{\Sigma}_{ji} = 0.5(\boldsymbol{\Sigma}_j + \overline{\boldsymbol{\Sigma}}_i)$. This Jensen Shannon divergence is then used to calculate the JSR and select the best data sets for the test data.

$$\mathrm{JS}[N_j \parallel D_i] = \frac{1}{2}(\mathrm{KL}[N_j \parallel M_{ji}] + \mathrm{KL}[D_i \parallel M_{ji}]) \tag{4.2}$$

Through knowing the differences between the EEG signals for each subject and class the JSR can be calculated. This aims to select a data set which has similar distributions for the same class while ensuring that the opposing classes are not similar. This is done through equation (4.3), with C representing the number of classes. The JSR aims to minimize the dissimilarity between the classes of two data sets while simultaneously maximizing the dissimilarity between different classes.

$$JSR = \frac{\sum_{j=1}^{C} \mathrm{JS}[N_j \parallel D_j]}{\sum_{i=1 i \neq j}^{C} (\mathrm{JS}[N_j \parallel D_i])} \tag{4.3}$$

When using the JSR for BCI subject to subject transfer learning the band-pass filtered EEG signals are used. As such $D_i(0, \overline{\boldsymbol{\Sigma}}_i)$ can be used to represent the distribution of one of the training data sets. While $N_j(0, \boldsymbol{\Sigma}_j)$ represents the distribution of the few subject specific trials we have from the user for each j class. In each of these distributions the normalized co-variance is estimated through the signal values x as shown in (4.4), with N number of trials.

$$\boldsymbol{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} \frac{\mathbf{x_i} \mathbf{x_i}^T}{tr(\mathbf{x_i} \mathbf{x_i}^T)} \tag{4.4}$$

As the band pass filtered EEG has a zero mean, equation (4.3) can be simplified to equation (4.5). Once the JSR has been calculated between the distribution of the subject specific trials and each of the possible training data distributions, the training data with the lowest JSR value is then selected. This data set is used to train the CSP and LDA of the BCI.

$$\text{JSR} = \sum_{j=1}^{C} \sum_{i \neq c} \frac{(\text{tr}(\mathbf{M})^{-1}\overline{\mathbf{\Sigma}}_j + (\mathbf{M})^{-1}\mathbf{\Sigma}_j) - \ln(\frac{\det(\overline{\mathbf{\Sigma}}_j)}{\det(\mathbf{\Sigma}_j)})}{C(\text{tr}(\mathbf{M})^{-1}\overline{\mathbf{\Sigma}}_i + (\mathbf{M})^{-1}\mathbf{\Sigma}_j) - \ln(\frac{\det(\overline{\mathbf{\Sigma}}_i)}{\det(\mathbf{\Sigma}_j)})} \tag{4.5}$$

4.2.2 Experiment

4.2.2.1 Dataset

To evaluate the framework proposed in this chapter the BCI Competition IV dataset 2a [70] dataset described in the previous chapter was used.

4.2.2.2 Data Processing

As before the EEG data for each user was split into the two sessions available. One being labeled as the source sessions and the other as the target session. The first 40 trials of each of the target sessions were removed and used by the proposed algorithm, as well as the comparison algorithms, to identify the best source sessions or source session weightings. This number was selected heuristically following examining different numbers of trials and finding that most classification accuracies leveled off after this point.After the source data is selected was band pass filtered from 8Hz to 35Hz, a pair wise CSPs was trained for 6 class pairs and then these features were used to create and train 6 pair wise LDAs.

To evaluate the effectiveness of the proposed JSR transfer learning, its results are compared with the accuracies obtained using a KL based similarity measure. Moreover, the proposed framework is compared to the algorithm previously suggested by Lotte utilizing other subjects training data [78]. These were also compared to training the BCI with the available subject specific trials provided to highlight the improvement in accuracy achieved by providing the additional training trials.

KL divergence is a long established method of calculating the difference between two Gaussian distributions. As such it is used for comparison against the JSR as a mean of transfer learning in the data domain. Equation (5.1) displays the calculations required to calculate the KL divergence. In this the data set which has the lowest summation of KL divergence between the test and target subjects classes is used for the BCI training.

Lotte and Guan previously developed an algorithm for BCI which used other subjects data

to reduce the need for calibration trials [78]. This evaluates the training data by training a BCI using each of the training data sets available. The subject specific trails are then used to evaluate the data sets. The selected data sets are then weighted, with λ, then used to estimate a new co-variance and mean in the feature domain as shown (4.7) and (4.6). For these equations μ and $\mathbf{\Sigma}$ are the mean feature vector and co-variance of the target subject while $\overline{\boldsymbol{\mu}}$ and $\overline{\mathbf{\Sigma}}$ are the mean feature vector and co-variance of the training subset. s is the number of selected training data sets.

$$\mathbf{\Sigma} = (1-\lambda)\mathbf{\Sigma} + \lambda\frac{1}{s}\sum_{i=1}^{s}\overline{\mathbf{\Sigma}}_{\mathbf{s}} \tag{4.6}$$

$$\boldsymbol{\mu} = (1-\lambda)\boldsymbol{\mu} + \lambda\frac{1}{s}\sum_{i=1}^{s}\overline{\boldsymbol{\mu}}_{s} \tag{4.7}$$

$$\lambda = \frac{DatasetAccuracy - SubjectSpecificAccuracy}{100 - ChanceAccuracy} \tag{4.8}$$

The weighting of λ is calculated through comparing the leave-one-out validation (LOOV) accuracy that is achieved by the subject specific trials and the accuracy achieved when the other data sets are used for training. If the leave one out validation outperforms the other data sets it is used for training the BCI, while if it is less than chance the trials are not used at all. If the LOOV accuracy is between the chance level and the accuracy achieved by the other data sets then they are weighted as shown in (4.8).

4.3 Results and Discussion

4.3.1 Improvement for BCI Deficient Users

Initially the proposed JSR is compared to the other transfer learning algorithms. The JSR allows BCI deficient users to achieve higher accuracy then any of the other algorithms. This can be seen in figure 4.1 which shows the accuracy achieved by each of the algorithms when 8 subject specific trials are available. For the users who achieved less than 70% accuracy with their subject specific trials the average improvement was 8% with JSR. In comparison the algorithm proposed by Lotte improved the accuracy for these subjects by just 3% and the KL divergence

caused a fall in accuracy. Subjects 1 and 5 in particular had a large increase in classification accuracy when the JSR was applied. While the average accuracy across all the subjects is not improved by JSR, compared to the standard BCI. This could be improved with a larger data base with more subjects to select from.

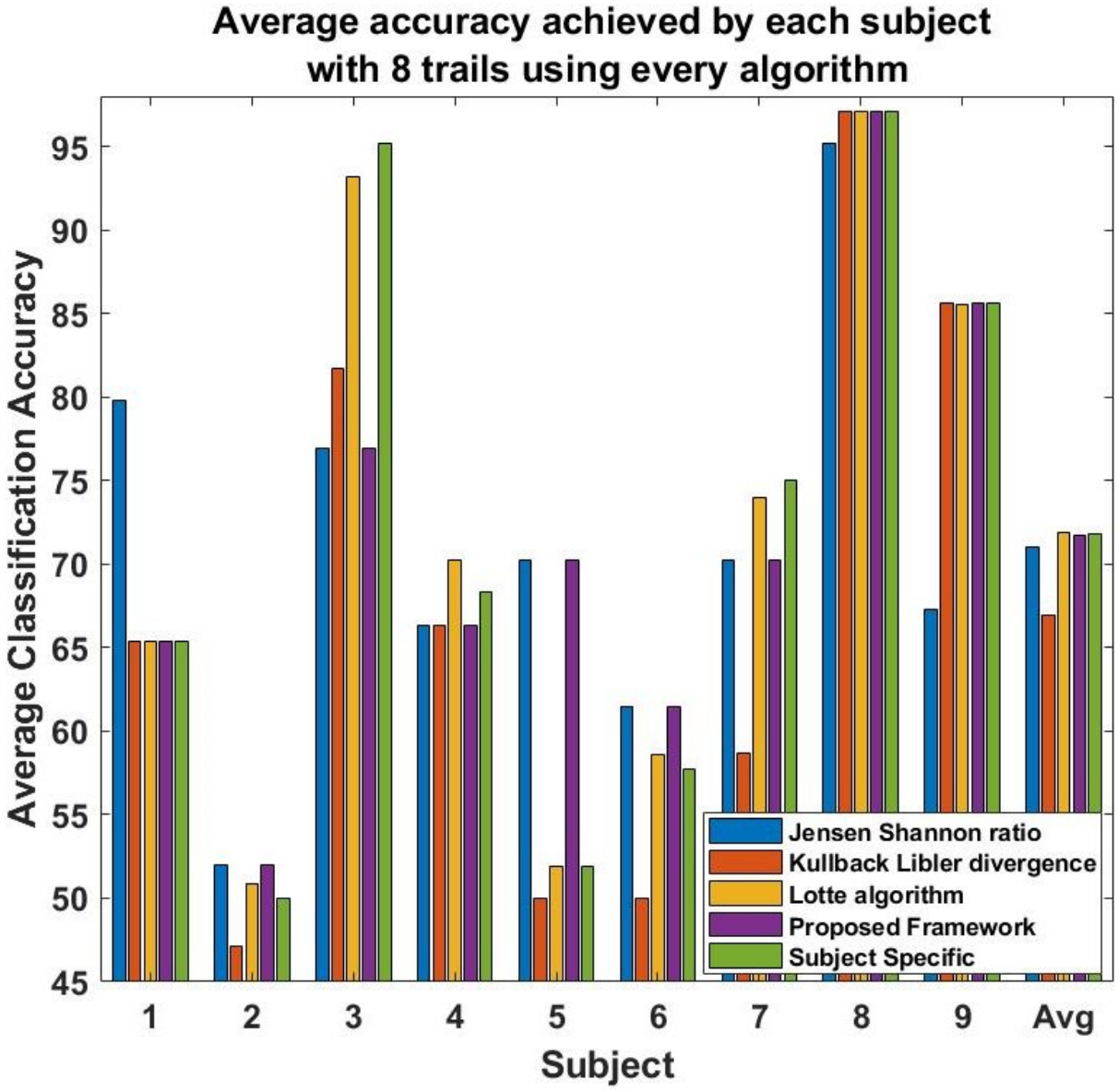

Figure 4.1: The accuracy achieved by each algorithm for every subject in the data set when only 8 trials are available for either training or calibration.

When examining the average classification accuracy across all the subjects, when 8 subject specific trials are available, the algorithm proposed by Lotte outperforms the JSR 0.9%. This may be due to Lotte's algorithm only using others data for users encountering BCI deficiency,

who require assistance, while JSR was used for all subjects. The subjects able to achieve high levels of accuracy with only a few subject specific trials lose accuracy with any of the other data sets available in the database. As such it is important to differentiate between the subjects who will achieve high accuracy and the subjects who will encounter BCI deficiency.

To identify these BCI deficient subjects the proposed framework incorporates the LOOV accuracy as a quick way to estimate the users competency in controlling EEG based BCI. Through this the users are classified as either BCI deficient or sufficient. The users will then either use the JSR to select the best training set for them or use the subject specific trials as the training set for the BCI.

Using the framework to select the subjects requiring transfer learning, before applying the JSR, improves the average accuracy across all the subjects. This is shown in figure 4.2 where the proposed framework is able to achieve 77% accuracy when 40 subject specific trials are available. When only the subject specific trials are used for training the average accuracy is only 74.5%. The proposed framework consistently outperforms the standard BCI although it does not perform optimally initially and experiences a small decrease in accuracy when 28 trials are available. The drop in accuracy which occurs when there are 28 subject specific trials is due to subjects 2 and 8 both experiencing a fall in accuracy. These subjects are both correctly identified as BCI deficient and sufficient respectively however still lose accuracy due to a few inconsistent trials. These trials causes the JSR to select a bad data set for subject 2 and leading to a fall in accuracy of 2%. This shows that the framework could benefit from an algorithm to evaluate and remove trials that are outliers.

4.3.2 Average Improvement from Proposed Framework

As mentioned the framework is not able to improve the average accuracy when only 8 subject specific trials are available. The LOOV misidentifies subjects 1 and 3 lowering the average accuracy by 0.1% compared to the standard BCI. The initially low accuracy of the proposed framework highlights one of the main problems which is its ineffectiveness in noticing BCI deficient users quickly. Using the LOOV accuracy is able to produce a fairly accurate prediction of the users capabilities when enough trials are available however a few outlying trials can affect the results. These outliers are not necessarily just trials that produce low levels of

accuracy but can also produce uncharacteristically high levels of accuracy which lead to the subjects being miss-classified by the LOOV and the framework under performing. The LOOV does perform well when there are enough trials provided to the validation and the framework does still improve on the standard BCI when 10 or more trials are available.

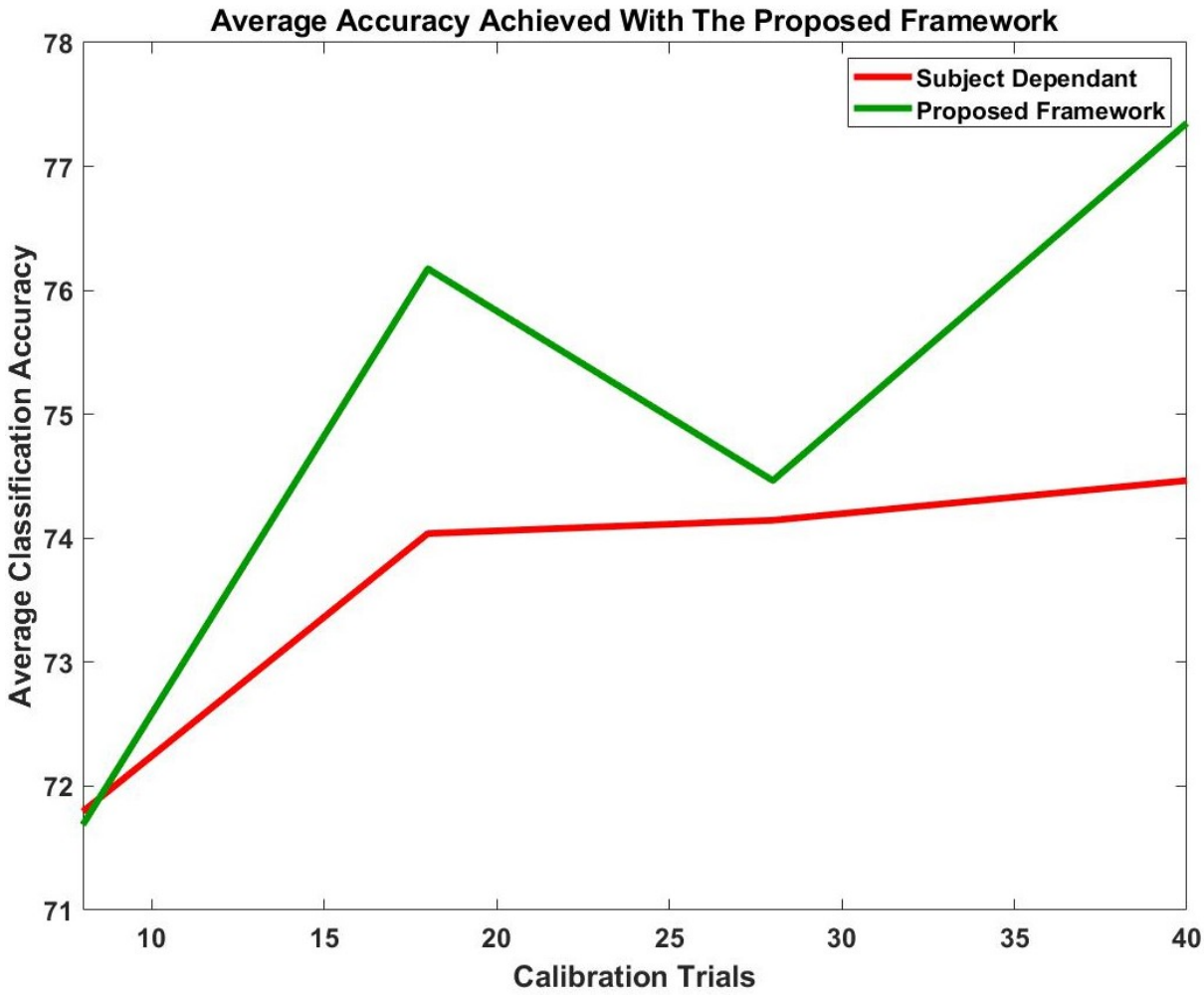

Figure 4.2: The average accuracy achieved by the standard BCI and framework improves as the number of trials increase.

Table 4.1 highlights this failing of the LOOV accuracy as a measurement of BCI competency. The proposed framework improves upon standard BCI however there is still a lot of room for improvement. If a better selection method was available this could further improve the accuracy of the framework. This increase in accuracy could be up to 3% if the correct subjects are selected to utilize the JSR. The current framework is able to improve the accuracy for BCI deficient subjects by over 5.5% when only 8 trials are available and by up to 6.5% when 40

Table 4.1: Average accuracy for BCI deficient users

Trials	Subject Specific	Proposed framework	JSR
8	57	**62.5**	**62.5**
40	58.65	65.1	**66.1**

trials are available.

4.4 Conclusion

Overall an improvement in the classification accuracy was consistently achieved for users encountering BCI deficiency by the proposed Jensen Shannon ratio data selection. When the "leave one out" method was used to select the users who required alternative training trials the average accuracy of the system outperformed the standard BCI by 3%. It is also important to remember that this was conducted using a publicly available data set with only nine subjects, providing a relatively small amount of training data to select from. A larger data set with more subjects may be able to find more appropriate data sets for each deficient subject. A number of users who were miss-classified could have benefited from using a different users data. As such to progress this work a key area to focus on will be in selecting a better predictor of classification accuracy going forward. This could potentially improve the systems allowing it to achieve an accuracy of 77% with only 8 trials. As the framework improves it can be applied to assist stroke patients with rehabilitation.

Chapter 5

A Transfer Learning algorithm to Reduce BCI Calibration Time for Long-term Users

The findings presented in this chapter have been published in the 42nd Engineering in Medicine and Biology Conference (EMBC 2020)[79] and in the "Frontiers of Neuroergonomics" [80].

5.1 Introduction

To reduce the BCI calibration time transfer learning can be used [9]. Transfer learning is a commonly employed technique in systems engineering when only a limited amount of data is available to train the model. Transfer learning compensates for the limited amount of labelled data available by extracting relevant information from other sources or domains to improve the classification model [9]. However, transfer learning in BCI is not a trivial task due to the non-stationary nature of brain signals. The properties of EEG signals often change considerably from session to session [10].

To reduce the affects of the non-stationarities, a range of approaches have been explored and embedded in transfer learning algorithms proposed for BCI [11, 12]. For example, some transfer learning algorithms applied alignment of the EEG distributions between the source and target sessions [77, 81–83] or weighted the source sessions according to their similarities

with the target session [84]. However these research studies focused on inter-subject transfer learning, evaluating the proposed solutions on datasets with only one or two sessions of data available for each subject. There is a research bias within BCI, with the majority of studies focusing on datasets where only one or two sessions are recorded from each subject [13]. There is a relatively small amount of literature focused on long-term users and inter-session transfer learning. One of the main causes of inter-subject variations in EEG signals is the varying brain morphology across the users. This issue is particularly significant for stroke users whose brain is altered by lesions that vary in size and location from user to user. Inter-session transfer learning is not affected by this issue with the majority of the non-stationarities coming from other causes.

Despite the limited number of studies on long-term BCI users, the potential benefits of inter-session transfer learning to reduce the calibration period is clear. Arvaneh et al. found that when 11 previous sessions, with 60 trials of each class in each session, were combined in the form of a "naive transfer learning", potentially invariant BCI features could be identified [57]. These sessions were collected over a period of one month and showed the potential of inter-session transfer learning to reduce the need for a calibration session. However, this approach is very limited. A lot of data is required before these so-called invariant features can be extracted, while some users still continue to perform better with the BCI model trained only using the data collected from the new session, called the "session-specific model".

One of the key sources of literature on inter-session transfer learning for long-term BCI users is from the teams competing in the Cybathlon BCI event [85]. The BCI race at the Cybathlon competitions has been held every 4 years where teams with tetraplegic pilots competed to control an avatar through a race track using BCI [86]. This required the teams to develop BCIs that could detect three different mental commands and to train a pilot to use the system. The user training period ranged from a month to over a year for some teams, allowing an in-depth exploration of BCI for long-term use and the potential of inter-session transfer learning.

The team led by Hehenberger et al. utilised inter-session transfer learning and intra-session adaptation for their BCI model developed for the Cybathlon competition race [87]. This team

worked with their pilot for 14 months collecting 26 sessions, each containing 120 EEG trials. Using the collected dataset, in an offline analysis, the authors highlighted the benefits of inter-session transfer learning over session-specific BCIs. They combined the data from the past five sessions and the new data to train the BCI model in the form of an inter-session naive transfer learning. Although it was successful to some extent, no optimisation was performed when combining the new and previous data to reduce the inter-session non-stationarities. Another Cybathlon team led by Benaroch et al. explored using both inter-session alignment and intra-session adaptation for their BCI model [88]. Over a period of 3 months this team collected 20 sessions, with the length of the sessions varying between an hour and two hours. To reduce the calibration time and reduce inter-session variability, they applied an alignment method projecting the Riemannian mean spatial covariance matrix from each session to a common reference point (i.e. the identity matrix). This alignment proved effective at reducing the non-stationarities and improving the inter-session transfer learning. However, even with this alignment still some of the source sessions were detrimental to the BCI model. Other researchers suggested that the use of selective transfer learning would reduce the effects of detrimental source sessions by weighting the source sessions based on their similarities with the target session [89].

Recently, Cao et al. has explored weighting the source sessions to improve the inter-session transfer learning in long-term BCI users performing BCI-based stroke rehabilitation [90]. For this purpose, they utilised a previously proposed inter-subject transfer learning algorithm [84]. The proposed transfer learning algorithm added a regularisation parameter to the objective function of the BCI classifier, aiming at minimising the dissimilarity between the classification parameters of the new session and the past sessions while maximising the two class separations. Importantly, the proposed algorithm gave different weights to different source classifiers based on the similarity between their features and the features of the target session. Cao et al. validated the utilised inter-session transfer learning algorithm on a BCI dataset from seven stroke patients. The dataset consists of 12 BCI sessions per stroke patient, each session having 180 trials performed in a randomised order. The proposed inter-session transfer learning algorithm significantly increased the classification accuracy of stroke subjects encountering BCI

deficiency. However the improvement in BCI accuracy was not statistically significant for all users. Many of the stroke participants performed better when relying on the standard naive transfer learning; with all source sessions having the same weight.

This chapter focuses on inter-session transfer learning to reduce the required calibration time for stroke patients who use BCI for rehabilitation. The proposed algorithm reduces this calibration time by combining previously recorded data from the same user with a limited number of data recorded from the current session, reducing the need for an extended calibration session. The proposed algorithm called regularised Kullback-Leibler weighted data space alignment (r-KLwDSA), consists of three steps to make effective use of the inter-session source sessions. Initially, the algorithm uses linear alignment to reduce non-stationarity between the current target session and the previous source sessions. The source sessions are then weighted to minimise the effects of any detrimental source data. Finally, the algorithm utilises regularisation to incorporate the target data and the weighted aligned source data into the BCI model.

The proposed r-KLwDSA algorithm is evaluated using EEG signals collected from 11 stroke patients over a period of six weeks. To simulate the real world scenario of long-term BCI use for stroke rehabilitation, the data will be evaluated chronologically, only using previously collected sessions for transfer learning. As such when evaluating the first target session only the screening session will be available for inter-session transfer learning and when evaluating session two both the screening session and session one will be used for transfer learning. This chapter will compare the effects of the weighting and alignment separately as well as the combined effect on classification performance. Furthermore, the performance of the proposed r-KLwDSA algorithm will also be compared with the performance of the session-specific BCI model trained with only the available trials from the current target session and the naive transfer learning model trained with only the previous source sessions without any alignment.

5.2 Methodology

In this chapter, we assume that EEG trials of J sessions, previously recorded from our current user, are available. These sessions are called source sessions. The j^{th} source session, $\hat{\mathrm{S}}_{\mathrm{j}}$, can be represented as $\hat{\mathrm{S}}_{\mathrm{j}} = (\hat{\mathbf{X}}_j^i, \hat{y}_j^i)_{i=1}^{\hat{m}_j}$, where $\hat{\mathbf{X}}_j^i \in \hat{\mathbf{X}}_j \subset R^{ch \times t}$ represents the i^{th} EEG trial from $\hat{m}_j$ total

EEG trials available in $\hat{S}_j$, and $\hat{y}_j^i \in \hat{\mathbf{y}}_j \subset \mathbb{R}$ represents the corresponding class label. Moreover, *ch* and *t* respectively denote the number of channels and the number of time samples recorded in each EEG trial.

Similarly, in this chapter, we have access to a small number of EEG trials from a new session, collected in a short calibration session from the same user. This session, referred to as the target session, is presented as $S = (\mathbf{X}^i, y^i)_{i=1}^m$, where $\mathbf{X}^i \in \mathbf{X} \subset R^{ch \times t}$ is the i^{th} recorded trial and $y^i \in \mathbf{y} \subset \mathbb{R}$ represents its corresponding class label. Moreover, m refers to the total number of trials in the target session.

As can be seen in Fig. 5.1, the proposed r-KLwDSA algorithm consists of three steps, each attempting to address one of the challenges of transfer learning in BCI. Step 1 reduces the non-stationarity between the EEG data from the source sessions and those from the target session. For this purpose, a linear transform is performed on the EEG data of each source session to reduce their distribution difference from the target data. Subsequently, step 2 defines the similarity between the EEG distributions of each linearly aligned source session and the target session using a proposed weighting mechanism. Finally, step 3 fuses the weighted aligned trials from the source sessions with the few available trials of the target session using a proposed regularisation method. In fact, the regularisation controls a trade-off between the target model from the new session and the weighted aligned source model from the past sessions. These three steps are explained in detail in the following subsections.

5.2.1 Linear alignment to reduce non-stationarities

When performing transfer learning, one of the key issues is the presence of non-stationarities which can cause large differences in the properties of EEG data from session to session. These differences in the data space can have a very detrimental effect on the performance of transfer learning in BCI. To address this issue, similar to Chapter 3, we propose the use of a linear transformation, $\mathbf{L}_j$, to reduce the mismatch between the distribution of each source session, $P(\hat{\mathbf{X}}_j, \hat{\mathbf{y}}_j)$, and the distribution of the target session, $P(\mathbf{X}, \mathbf{y})$. For this purpose, $\mathbf{L}_j \subset R^{ch \times ch}$ needs to be calculated such that the distribution dissimilarity between $P(\mathbf{Z}_j, \hat{\mathbf{y}}_j)$ and $P(\mathbf{X}, \mathbf{Y})$ is minimised where $\mathbf{Z_j} = \mathbf{L}_j \hat{\mathbf{X}}_j$.

Assuming that EEG signals have Gaussian distributions [69], we used Kullback Leibler

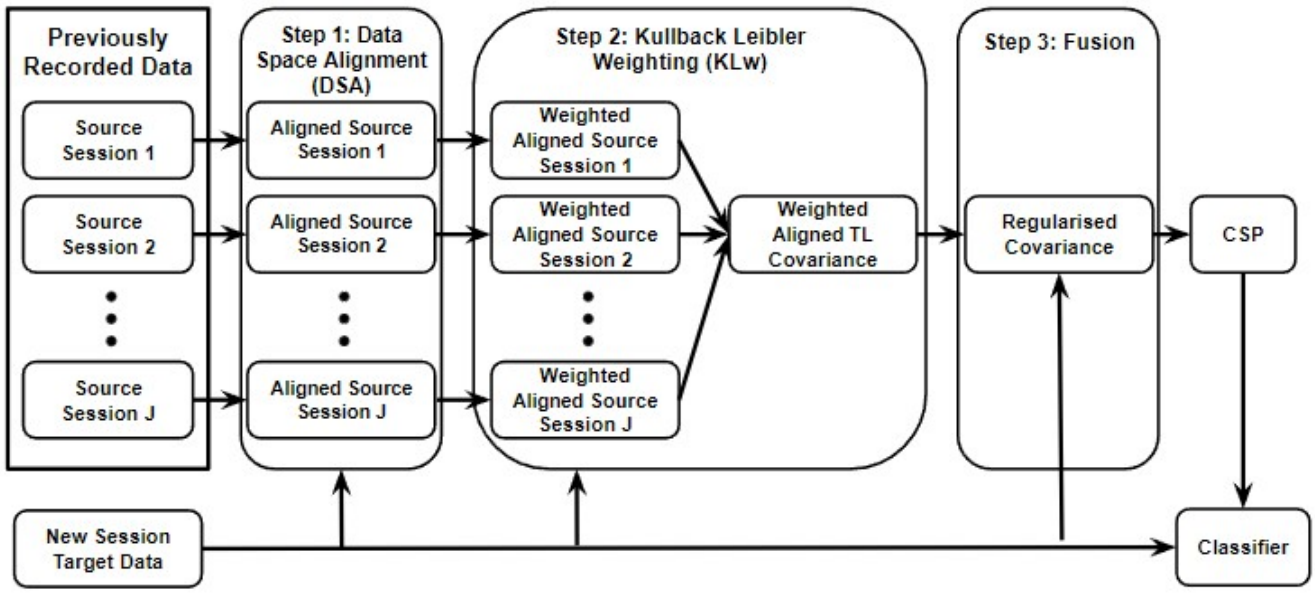

Figure 5.1: The proposed r-KLDSA algorithm is comprised of three steps, 1) the EEG data from the source sessions are aligned to the EEG data from the available target data using data space alignment, 2) weighting the aligned data of the source sessions based on their similarities with the data of the target session, 3) fusing the weighted aligned source data with the target data using a regularisation method.

(KL) divergence for Gaussian distributions to measure the distribution dissimilarity between the target session and each source session. Given two Gaussian distributions, $N_1(\mu_1, \mathbf{\Sigma_1})$ and $N_2(\mu_2, \mathbf{\Sigma_2})$ with μ_1 and μ_2 as the means and $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma_2}$ as the covariance matrices, the KL divergence between N_1 and N_2 is measured as

$$KL[N_1 \parallel N_2] = \frac{1}{2}(\mathrm{tr}(\mathbf{\Sigma_2}^{\dagger}\mathbf{\Sigma_1}) + (\mu_2 - \mu_1)^{\mathrm{T}}\mathbf{\Sigma_2}^{\dagger}(\mu_2 - \mu_1) - \ln(\frac{\det \mathbf{\Sigma_1}}{\det \mathbf{\Sigma_2}}) - \mathrm{k}), \tag{5.1}$$

where tr, det and ln denote the trace function, the determinant function and the natural logarithm function respectively. † and T denote the pseudo-inverse and the transpose functions, respectively. Finally, k refers to the dimension of the data.

As the EEG data of the source and target sessions are band-pass filtered, they have zero means. The co-variance matrices, representing the distributions of the target session, S, and the j^{th} source sessions, $\hat{\mathrm{S}}_{\mathrm{j}}$, are calculated using (5.2) and (5.3) respectively,

$$\overline{\mathbf{\Sigma}}^{c} = \frac{1}{m^c}\sum_{i=1}^{m^c}\frac{\mathbf{X}^{c,i}(\mathbf{X}^{c,i})^{\mathrm{T}}}{\mathrm{tr}(\mathbf{X}^{\mathrm{c,i}}(\mathbf{X}^{\mathrm{c,iT}}))}, \tag{5.2}$$

$$\hat{\mathbf{\Sigma}}_j^{c} = \frac{1}{\hat{m}_j^{c}}\sum_{i=1}^{\hat{m}_j^{c}}\frac{\hat{\mathbf{X}}_j^{c,i}(\hat{\mathbf{X}}_j^{c,i})^{\mathrm{T}}}{\mathrm{tr}(\hat{\mathbf{X}}_{\mathrm{j}}^{\mathrm{c,i}}(\hat{\mathbf{X}}_{\mathrm{j}}^{\mathrm{c,i}})^{\mathrm{T}})}, \tag{5.3}$$

where, c denotes the class, and m^c is the total number of trials for the class c. Subsequently, the linearly transformed $\hat{\mathrm{S}}_{\mathrm{j}}$, presented as $\mathbf{L_j}\hat{\mathrm{S}}_{\mathrm{j}}$, has a zero mean and the covariance matrix calculated as $\mathbf{L_j}\hat{\mathbf{\Sigma}}_j^{c}\mathbf{L_j}^{\mathrm{T}}$. Given (5.1), the distribution dissimilarity between the linearly transformed j^{th} source session and the target session can be calculated as:

$$KL[\mathbf{L_j}\hat{\mathrm{S}}_{\mathrm{j}} \parallel \mathrm{S}] = \frac{1}{2}\sum_{\mathrm{c}=1}^{2}[\mathrm{tr}(\overline{\mathbf{\Sigma}}^{\mathrm{c}\dagger}\mathbf{L_j}\hat{\mathbf{\Sigma}}_{\mathrm{j}}^{\mathrm{c}}\mathbf{L_j}^{\mathrm{T}}) - \ln(\frac{\det(\mathbf{L_j}\hat{\mathbf{\Sigma}}_{\mathrm{j}}^{\mathrm{c}}\mathbf{L_j}^{\mathrm{T}})}{\det(\overline{\mathbf{\Sigma}}^{\mathrm{c}})}) - \mathrm{ch}]. \tag{5.4}$$

The linear transform $\mathbf{L}_j$ aims to minimise the distribution dissimilarity between $\hat{\mathrm{S}}_{\mathrm{j}}$ and S. To calculate $\mathbf{L}_j$, the first order derivation of the loss function (5.5) with respect to $\mathbf{L}_j$ is computed and set to zero, as shown in (5.6) and (5.7). For more details on how optimum $\mathbf{L}_j$ has been calculated, please see the appendix.

$$A(\mathbf{L}_j) = \min_{\mathbf{L}_j}\sum_{c=1}^{2}\frac{1}{2}[\mathrm{tr}(\overline{\mathbf{\Sigma}}^{\mathrm{c}\dagger}\mathbf{L_j}\hat{\mathbf{\Sigma}}_{\mathrm{j}}^{\mathrm{c}}\mathbf{L_j}^{\mathrm{T}}) - \ln(\frac{\det(\mathbf{L_j}\hat{\mathbf{\Sigma}}_{\mathbf{j}}^{\mathbf{c}}\mathbf{L_j}^{\mathrm{T}})}{\det(\overline{\mathbf{\Sigma}}^{c})}) - ch]. \tag{5.5}$$

$$\frac{dA}{d\mathbf{L}_j} = \sum_{c=1}^{2}\frac{1}{2}[\frac{d}{d\mathbf{L}_j}\mathrm{tr}(\overline{\mathbf{\Sigma}}^{\mathrm{c}\dagger}\mathbf{L_j}\hat{\mathbf{\Sigma}}_{\mathrm{j}}^{\mathrm{c}}\mathbf{L_j}^{\mathrm{T}}) - \frac{d}{d\mathbf{L}_j}\ln(\det(\mathbf{L_j}\hat{\mathbf{\Sigma}}_{\mathbf{j}}^{\mathbf{c}}\mathbf{L_j}^{\mathrm{T}}))] = 0. \tag{5.6}$$

$$\mathbf{L}_j = \sqrt{2}\sum_{c=1}^{2}(\hat{\mathbf{\Sigma}}_j^c \overline{\mathbf{\Sigma}}^{c\dagger})^{\dagger 0.5}. \tag{5.7}$$

By applying the linear transform $\mathbf{L}_j$ to each of the source sessions, the KL divergence between the target and the aligned source session is minimised. This reduces the effect of the non-stationarities from session to session.

5.2.2 Weighting according to EEG distribution similarity

Although reducing the non-stationarity can help improve transfer learning, some source sessions can still be detrimental to the BCI. The second step of the proposed r-KLwDSA algorithm, shown in figure 5.1, weights the aligned source data of each previous session to reduce the impact of adverse data while placing more weight on data that is similar to the target session. In (5.4), we proposed using the KL divergence between Gaussian distributions to measure dissimilarity between the aligned j^{th} source session and the target session. Subsequently, the assigned weight for the aligned j^{th} source session, ω_j, presenting its distribution similarity to the target session, is calculated through equation (5.8),

$$\omega_j = \frac{(KL[\mathbf{L_j}\hat{\mathrm{S_j}} \parallel \mathrm{S}])^{-1}}{\sum_{i=1}^{J}(KL[\mathbf{L_i}\hat{\mathrm{S_i}} \parallel \mathrm{S}])^{-1}}, \tag{5.8}$$

where, $KL[\mathbf{L_j}\hat{\mathrm{S_i}} \parallel \mathrm{S}]$ is calculated using (5.4). According to (5.8), source sessions with similar data to the data of the target session are assigned larger weights, whereas aligned source sessions with less similarity to the target session are given small weights. Consequently, the weighted aligned source data are used to calculate the co-variance matrix of past data, called the transfer learning co-variance matrix, $\hat{\mathbf{\Sigma}}_{TL}^c$, as

$$\hat{\mathbf{\Sigma}}_{TL}^c = \sum_{j=1}^{J} \omega_j \mathbf{L}_j \hat{\mathbf{\Sigma}}_j^c \mathbf{L}_j^T. \tag{5.9}$$

5.2.3 Regularised Transfer Learning between past and present data

Transfer learning can be very effective for some of the target sessions, while for some other target sessions, the source data might be detrimental, even after weighting and alignment. These target sessions usually tend to be able to achieve high classification accuracy even when only

a few target trials are available for training. As such the third step of the proposed r-KLwDSA algorithm uses a regularisation method to find a trade-off between data from the previous sessions and the new data from the new target session in the final BCI model. Thus, the final regularized co-variance matrices are calculated using (5.10) with a regularisation parameter, $r \in \{0, 0.1, ..., 1\}$. The individualised regularisation parameter is calculated for each target session and selected through leave-one-out cross validation method on the available target trials. The parameter achieving the highest average leave-one-out classification accuracy was then used to produce the final co-variance matrix, $\mathbf{\Sigma}_F^c$ for class c. The final co-variance matrices are then used for training the Common Spatial Patterns (CSP) features [91], as further elaborated in Section 3.2.

$$\mathbf{\Sigma}_F^c = r\overline{\mathbf{\Sigma}}^c + (1-r)\hat{\mathbf{\Sigma}}_{TL}^c \tag{5.10}$$

5.3 Experiment

5.3.1 Dataset

The dataset used to evaluate the proposed algorithm is known as the nBetter dataset [7]. This dataset was collected by the Institute for Infocomm Research, A*Star, Singapore to evaluate the efficacy of the Neurostyle Brain Exercise Therapy Towards Enhanced Recovery (nBETTER) system in post-stroke upper limb rehabilitation. The clinical trial obtained ethical approval from the Institution's Domain Specific Review Board (IRB), National Healthcare Group, Singapore and is registered in ClinicalTrials.gov under NCT02765334. The use of this dataset to evaluate our proposed algorithm was approved through IRB Reference: 2020-103.

All participants in the study had their first-ever stroke 3 to 24 months before participating the clinical trial, affecting their upper limb movements. They all provided informed consent before enrollment in the study. Potential participants attended a 40 minute BCI screening session, and only those who achieved BCI accuracy above 57.5% 10-fold cross validation accuracy were invited to attend the nBetter intervention. The EEG data was collected from 24 EEG channels, placed in the international 10-20 system positioning: F3, F4, FC3, FC4, C3, C4, CP3, CP4, P3, P4, FT7, FT8, T3, T4, TP7, TP8, Fz, Oz, FCz, Cz, CPz, Pz, A1 and A2, and digitally sampled

at 256 Hz for a voltage range of $\pm 300mV$.

In total the nBetter dataset contains the EEG data from 11 stroke patients completing one screening session, 18 supervised sessions and 18 therapy sessions supervised by an occupational therapist. The screening session, collected at the start of the study, contains four runs, each consisting of 20 idle trials and 20 motor imagery trials. For the idle class the participant was instructed to relax, whereas for the motor imagery class the participant was instructed to imagine movement of their affected hand. As shown in Figure 5.2, each supervised session followed by one therapy session on the same day, conducted thrice weekly over a six week period. In each of the supervised sessions 40 labelled trials were collected, half motor imagery and the other half idle trials. The therapy sessions contain four runs each consisting of 40 motor imagery trials. Each of the trials in these sessions lasted 13 seconds, as illustrated in Figure 5.3 with the instruction to perform motor imagery being presented for four seconds after giving the participant two seconds to prepare.

To evaluate the proposed r-KLwDSA algorithm, only the screening and supervised sessions were used. These sessions contained clearly labelled trials with equal numbers of each of the two classes. When considering each supervised session as the target session the first 10 trials of each class were used as training data, while the rest were kept for evaluation. When the supervised session was used as a source session all trials were used for transfer learning. To simulate a real world scenario the supervised sessions were evaluated chronologically. As a result, when the supervised session one was evaluated as the target session, only the screening session was used as source data. Similarly, when the supervised session 18 was used as the target session, the supervised sessions 1 to 17 and the screening session were used as the source sessions.

5.3.2 Data Processing

Any of the trials missing time samples were removed, other than this no artifact rejection algorithms used. A zero phase elliptic band pass filter from 8 to 35 Hz was used to filter the EEG data as this range contains the key range of frequencies that are linked to motor imagery. The band-passed EEG signals from 2.5 to 5 seconds after the presentation the cue were used for feature extraction. This time interval considers sufficient time for the participant to react

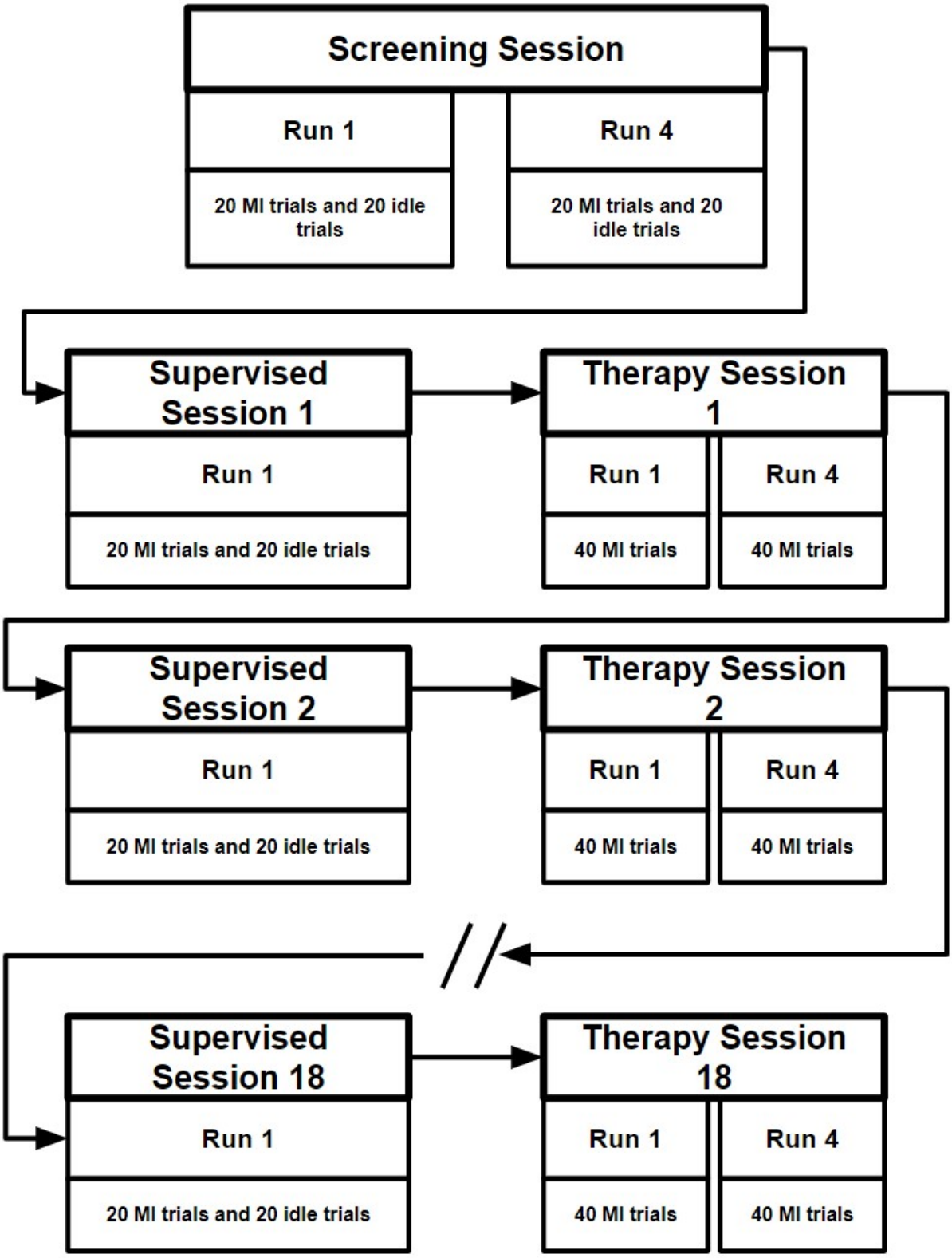

Figure 5.2: Illustration of the data collection of the nBetter dataset. First a screening session is collected at the start of the six weeks. Following the screening session, a supervised session followed by a therapy session is collected three times a week.

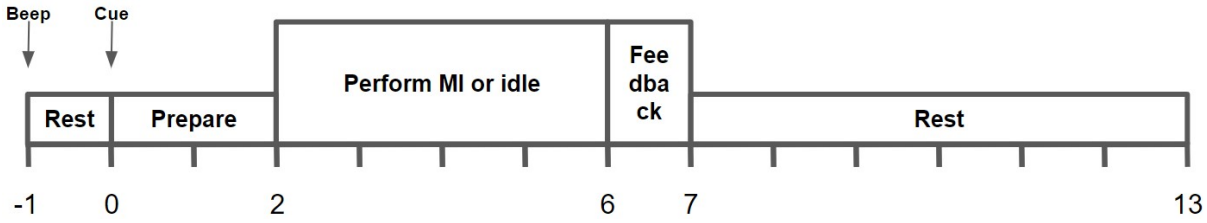

Figure 5.3: A trial from the nBetter supervised session. Each supervised session consists of 20 motor imagery (MI) trials and 20 idle trials.

to the motor imagery instruction. Six different feature extraction algorithms, including the proposed r-KLwDSA algorithm, are used in this chapter. These six algorithms utilise CSP filters to calculate the features. The CSP diagonalises the covariance calculated for each class to find the subspace that maximises the variance of one class while minimising the variance of the second classes. The first and last two rows of the CSP were selected as the most discriminative spatial filters for feature extraction. The normalised variances of the spatially filtered EEG signals from the training part of the target session were used as the features to train a Linear Discrete Analysis (LDA) classifier.

Despite all using CSP for calculating the features, the covariance matrix of each class, used to calculate CSP, was obtained differently in the six applied methods. For the proposed r-KLwDSA algorithm, the covariance of each class was calculated as described above in (5.10). The proposed KLwDSA algorithm is a special case of the proposed r-KLwDSA algorithm with $r = 0$. In other words, the proposed KLwDSA algorithm uses the covariance matrices calculated in (5.9) to obtain the CSP filters. The standard session-specific (SS) algorithm is also a special case of the proposed r-KLwDSA algorithm with $r = 1$. As when $r = 1$ no transfer learning occurs making it the same as a standard CSP-LDA BCI. Thus, the SS uses only the training data available in the target session to calculate the CSP filters. The naive transfer learning algorithm, nTL, concatenates all the source sessions with equal weights and without alignments to calculate the covariance of each class for the CSP algorithm. The DSA and KLw algorithms are extensions to nTL. The DSA algorithm applies the DSA linear transform to each of the source sessions before calculating the CSP covariance matrices by concatenating the aligned source trials. The KLw algorithm weights each of the source sessions using the weighting method proposed in step 2 of the proposed algorithm without any alignment.

Then the weighted covariance matrices of the source sessions are used for calculating the CSP filters. All these algorithms are compared in terms of the classification results to understand their merits and disadvantages.

5.4 Results and Discussion

5.4.1 Comparison of classification accuracy results

Figure 5.4 shows the average classification accuracy of the six above-mentioned algorithms across all the subjects and sessions when a different number of target trials were available for BCI calibration. As shown in Figure 5.4, the proposed r-KLwDSA algorithm outperformed all the other algorithms across different numbers of available target trials. Given the number of available target trials between 2 to 10 per class, r-KLwDSA consistently outperformed SS by an average more than 4%. The sensitivity and specificity were also calculated for the proposed r-KLwDSA algorithm and have been included in a table in the supplementary materials.

A 6 (algorithms = SS, nTL, KL, DSA, KLwDSA and r-KLwDSA) $\times$ 5 (target trials per class = 2, 3, 4, 5 and 10) $\times$ 18 (available source sessions= 1, 2,..., 18) repeated measures ANOVA test was performed on the classification results using the SPSS software. The statistical results showed that only the number of trials satisfied Mauchly Sphericity, so the Greenhouse Geisser was used to evaluate the effects of the algorithms, the number of target trials and sessions on the classification results. The results showed that the number of target trials and the algorithms had statistically significant effects on the classification accuracy with P-values of less than 0.001 and 0.048, respectively. The post hoc analysis showed using 3, 4, 5, and 10 target trials per class led to significantly better classification results compared to when we used only 2 target trials per class ($p < 0.001$). Similarly using 10 target trials per class significantly outperformed the results of using 3 trials per class ($p = 0.008$). When comparing the algorithms separately, the post hoc analysis showed that the proposed r-KLwDSA algorithm significantly outperformed all the other algorithms. P-values of less than 0.001 were obtained when comparing the proposed r-KLwDSA with the SS, nTL, DSA, KLw and KLwDSA algorithms. The post hoc analysis did not show any significant difference between the SS, nTL, KLw and DSA algorithms. Interestingly, by combining KLw and DSA, the proposed KLwDSA algorithm sig-

nificantly outperformed the SS, nTL, DSA and KLw algorithms, with the P-values of 0.006, 0.013, 0.012 and 0.032 respectively. We corrected the p-values for the multiple comparisons using the Bonferroni correction method.

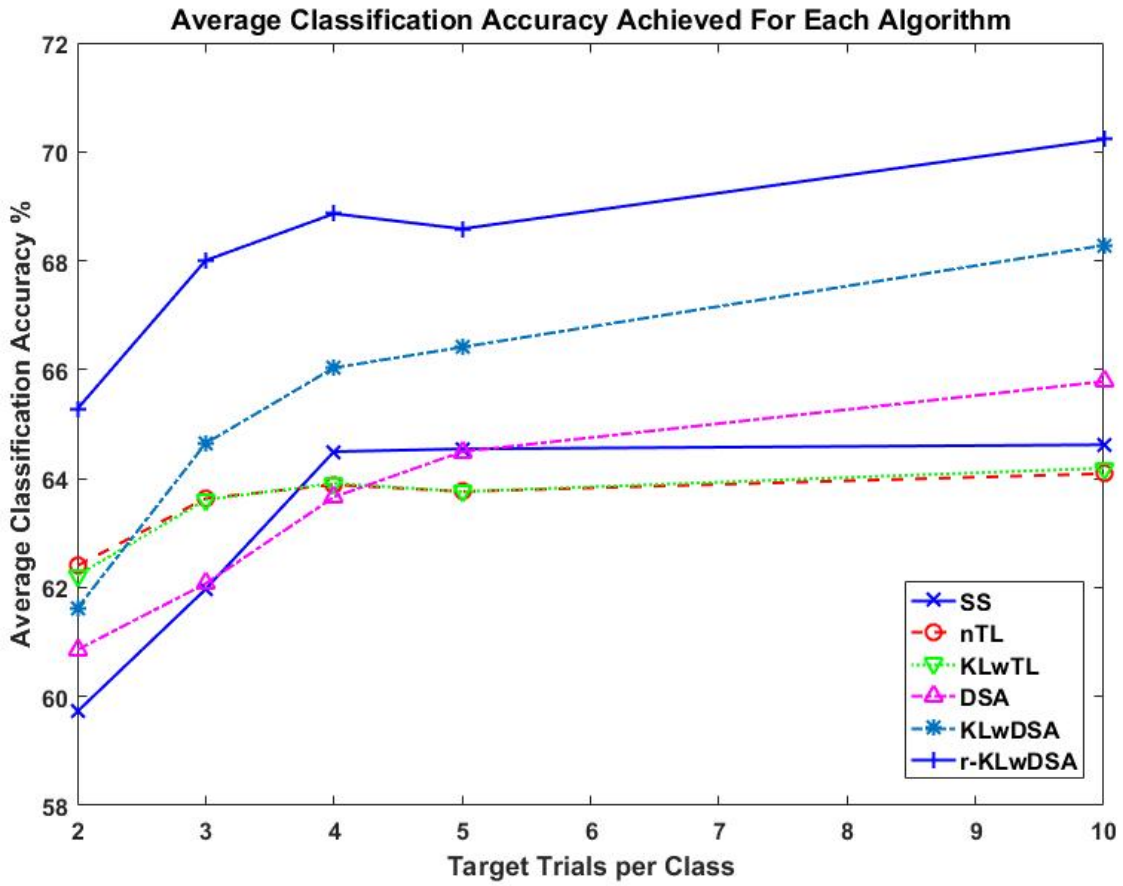

Figure 5.4: Average classification accuracy of six different algorithms across all subjects and sessions, when different number of target trials were available for calibration. SS denotes the target session-specific algorithm; nTL, naive transfer learning; proposed KLw, Kullback Liebler weighted transfer learning; proposed DSA, data space alignment transfer learning; proposed KLwDSA, aligned and weighted transfer learning; and proposed r-KLwDSA, the regularised, aligned and weighted transfer learning algorithm.

To better understand the merits of the proposed r-KLwDSA over the standard SS algorithm, the best two features of these two algorithms, obtained using the target data from subject 6, session 16, were compared in Figure 5.5. Figure 5.5 highlights the benefit of implementing the proposed algorithm when there are only a limited number of target trials available. As shown

the SS algorithm suffers greatly from overfitting due to the lack of target training trials. The SS algorithm extracts session-specific CSP features which perform very well with the available training data with only two features on the incorrect side of the hyperplane. However when transferring to the target test data the features from both classes overlap. The r-KLwDSA algorithm is not affected by this overfitting due to the integration of the source sessions data. While the trained target features overlap slightly more than the SS algorithm the test target features are much more distinctive.

5.4.2 Average Sensitivity and Specificity of each algorithm

Table 5.1 compares the average Sensitivity and specificity of the proposed r-KLwDSA algorithm with the Sensitivity and specificity of SS, nTL, KL, DSA, and KLwDSA.

	2 target trials per class			**3 target trials per class**			**4 target trials per class**			**5 target trials per class**			**10 target trials per class**		
	Mean Accu.	Mean Spec.	Mean Sens.	Mean Accu.	Mean Spec.	Mean Sens.	Mean Accu.	Mean Spec.	Mean Sens.	Mean Accu.	Mean Spec.	Mean Sens.	Mean Accu.	Mean Spec.	Mean Sens.
SS	59.72	60.20	59.24	61.97	61.62	62.32	64.49	62.83	66.16	64.55	63.79	65.30	64.62	65.51	63.74
nTL	62.40	59.14	65.66	63.64	62.02	65.25	63.89	62.42	65.35	63.76	61.67	65.86	64.09	59.09	69.09
KL	62.22	58.89	65.56	63.61	61.97	65.25	63.91	62.42	65.40	63.76	61.67	65.86	64.19	59.65	68.74
DSA	60.86	59.65	62.07	62.07	62.63	61.52	63.66	63.08	64.24	64.49	64.34	64.65	65.78	61.31	70.25
KLwDSA	61.62	58.79	64.44	64.65	62.32	66.97	66.04	63.43	68.64	66.41	64.14	68.69	68.28	64.80	71.77
r-KLwDSA	65.28	62.63	67.93	68.01	65.51	70.51	68.86	64.55	73.18	68.59	64.75	72.42	70.23	67.27	73.18

Table 5.1: The average classification accuracy, specificity and sensitivity are shown for each algorithm as the number of target trials increases. These averages are calculated across all the users and sessions. Accu., Spec. and Sens. denote accuracy, specificity and sensitivity respectively.

5.4.3 Effects of number of target trials and source sessions on the performance of r-KLwDSA

Further statistical analyses were carried out to investigate the effects of the number of target trials and source sessions on the performance of the proposed r-KLwDSA algorithm. A 5 (target trials per class = 2, 3, 4, 5 and 10) $\times$ 18 (available source sessions= 1, 2,..., 18) repeated measures ANOVA test was performed on the r-KLwDSA classification results. Mauchly Sphericity was satisfied for the number of trials per class, so the sphericity assumed results were used. The

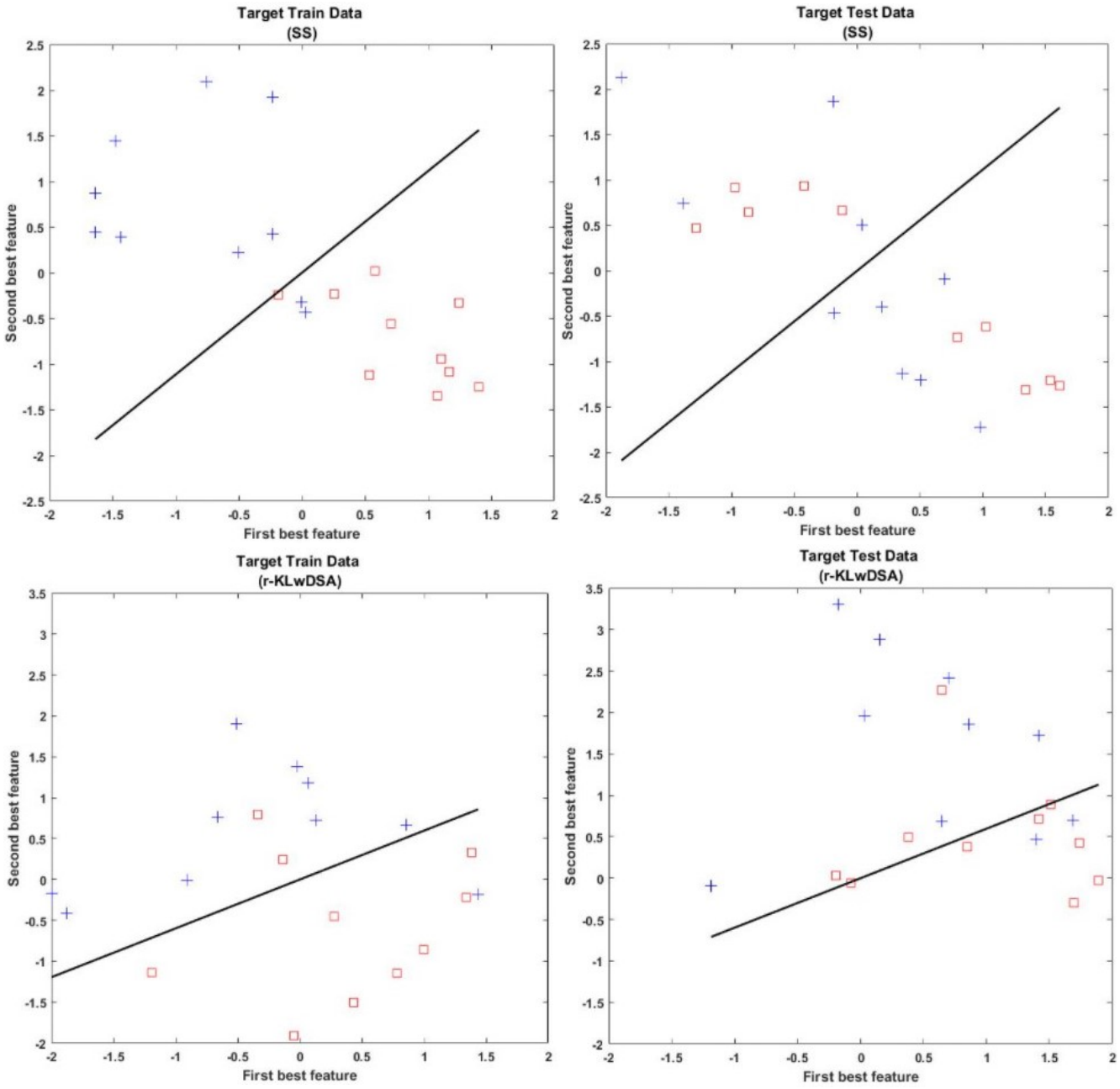

Figure 5.5: An example of the distribution of the two best features obtained by the session-specific CSP and the proposed r-KLwDSA. These features were collected from subject 6, session 16. The blue crosses and red squares denote the normalized features of the hand motor imagery and the rest class, respectively. The black line represents the LDA hyperplane obtained by the target train data.

ANOVA results confirmed that the number of available target trials had a main affect on the classification accuracy with a P-value of less than 0.001. This is aligned with previous literature, as increasing the number of target trials improves the estimation of the average target

trials for each class. The improved average target trial results in a better DSA alignment and a more accurate KL weighting, consequently improving the r-KLwDSA accuracy.

While increasing the available target trials significantly improved the classification accuracy, the number of available source sessions did not have a main effect on the classification results of the proposed r-KLwDSA algorithm (P-value $= 0.472$). A potential factor contributing to the lack of a significant effect of the number of source sessions on the r-KLwDSA results could be the non-stationarity of the EEG signals. The users' EEG signals vary from session to session, and these variations can be significant over extended periods. Thus, increasing the number of the source data could not necessarily improve the BCI accuracy. Please note that to mimic practical scenarios, we considered the data chronologically and used all the available source sessions for training r-KLwDSA. Thus, our results did not make a direct comparison between the different number of source sessions as by increasing the number of the source sessions the target sessions were changed. To better analyse the impact of number of source sessions on the r-KLwDSA performance, we fixed the target session to session 18 and used different numbers of the nearest source sessions. However still we did not observe a statistically significant effect of number of source sessions on rKLwDSA results.

5.4.4 Effects of Increasing Source Session Availability on Accuracy

One of the interesting findings of this chapter is how increasing the available source session does not always lead to an improvement in the classification accuracy. Figure 5.6 shows the changes in the classification accuracy for target session 18 for all 11 subjects. First only the previous session, session 17, is used to train the BCI then session 17 and 16 are used followed by sessions 17, 16 and 15 and so on until all the previous data has been used.

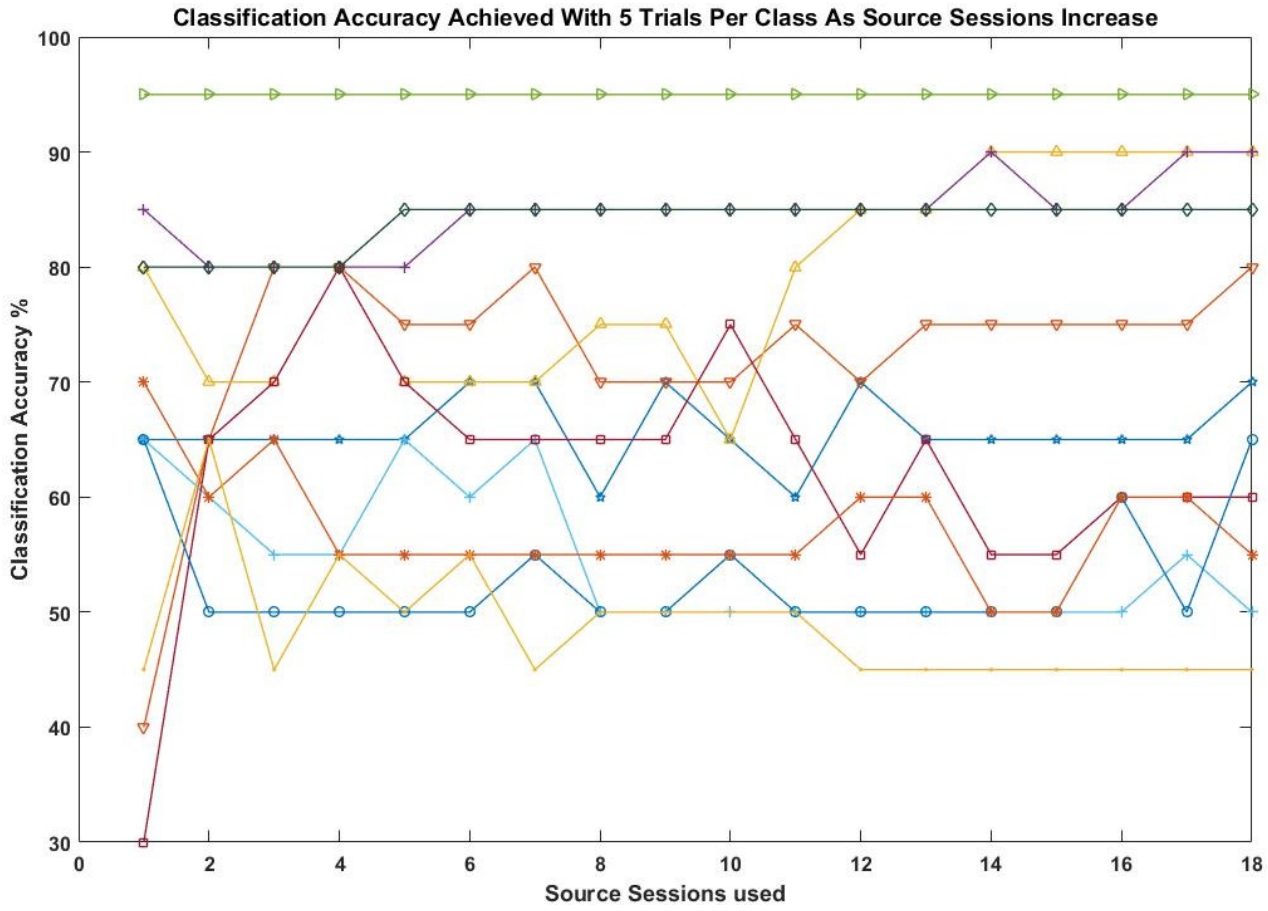

Figure 5.6: Classification results of the proposed r-KLwDSA algorithm using 5 trials per class from the session 18 as the target session and different number of source sessions. Each curve presents one participant.

Figure 5.7 shows the classification accuracy across all 11 subjects as a box plot. The highest mean classification accuracy across all the users is when there are 4 source sessions available. However the actual best number of source sessions varies significantly from user to user.

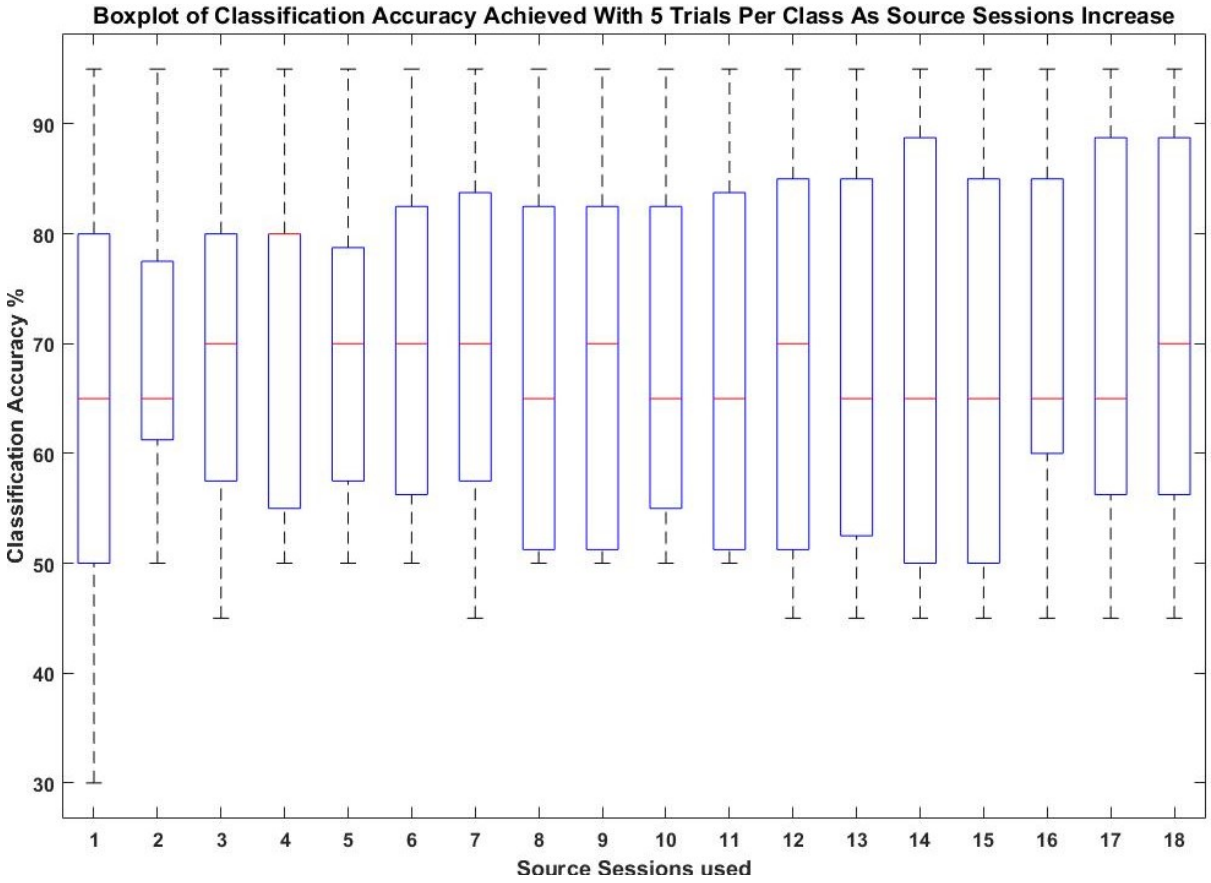

Figure 5.7: Box-plots of the classification results for the proposed r-KLwDSA algorithm using 5 trials per class from the session 18 as the target session and different number of source sessions.

5.4.5 Change in classification accuracy for those encountering BCI deficiency

Figure 5.8 presents scatter plots showing all the classification results obtained using SS against those obtained using the proposed r-KLwDSA algorithm, when 2, 3, 5 and 10 target trails were available for BCI calibration. As can be seen, compared to the SS algorithm, the increased classification accuracy from using the proposed r-KLwDSA algorithm was pronounced for stroke users encountering BCI deficiency (i.e. SS accuracy less than 60%). As expected, increasing the number of available target trials led to a larger improvement in the classification accuracy of the users who were identified as BCI deficient using the SS algorithm.

To better investigate the benefit of using r-KLwDSA, Table 5.2 splits the 198 available target sessions based on the classification accuracy achieved by the SS algorithm, when ten target trials per class were available for BCI calibration. Impressively, for the total 73 sessions where the SS encountered BCI deficiency (i.e. accuracy below 60%), the use of r-KLwDSA yielded

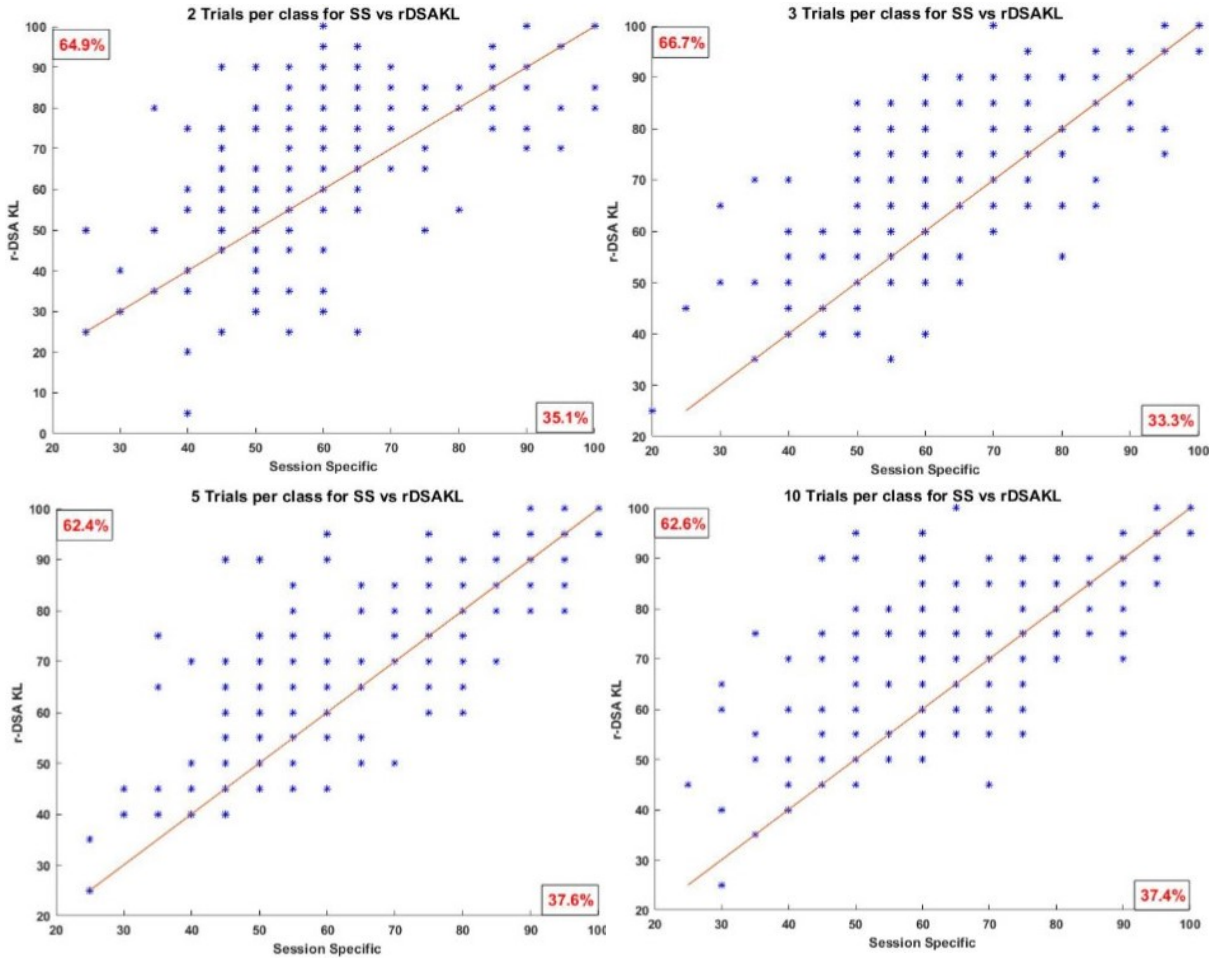

Figure 5.8: Four scatter plots showing the SS classification accuracy against the classification accuracy of the proposed r-KLwDSA algorithm. Each star represents one test session of a patient.

a significant increase in the classification accuracy with an average improvement of 13.22% and p-value of 0. Moreover, the proposed r-KLwDSA significantly improved the classification accuracy of the total 87 sessions achieving between 60% and 85% accuracy using the SS algorithm. However, the observed average improvement was smaller, with the average accuracy increasing by 2.99%. On the contrary, sessions with a SS classification accuracy more than 85% observed an average decrease in the accuracy when r-KLwDSA was applied. This shows when the session-specific model performs very well, adding source sessions to the model could be detrimental and the proposed regularisation method could not deal with it as expected. This could be because the regularisation values were chosen using cross validation on only very few target trials, which increases the risk of over fitting. Thus, there is a need to investigate

novel ways to find the optimum regularisation values, particularly for those with very good initial SS performance.

	Below 60%	60% to 85%	85% to 100%
SS Mean Acc	45.82%	68.56%	91.71%
r-KLwDSA Mean Acc	59.04%	71.55%	88.68%
SS Count	73	87	38
P-Value	0	0.016	0.003

Table 5.2: The sessions are separated into those achieving below 60%, between 60% and 85% and above 85% classification accuracy using the session specific (SS) BCI model, when there were 10 target trials per class available for calibration. The average classification accuracy achieved by these sessions using the proposed r-KLwDSA and the session specific BCI are presented with the p-value calculated from the t-test between them.

Considering the r-KLwDSA results and regardless of the number of target trials available, we observed consistent improvements in the classification accuracy of the sessions with BCI deficient SS models. For those sessions, the proposed r-KLwDSA algorithm improved the classification accuracy by an average of 9.29%, 9.54%, 8.93%, 9% and 13.22% for 2, 3, 4, 5 and 10 trials per class, respectively. Importantly, the observed improvements in the classification accuracy were significant for all these different number of available target trials with P-values of less than 0.001.

In summary, Figure 5.8 and Table 5.2 show that the proposed r-KLwDSA could potentially reduce the number of sessions encountering BCI deficiency while limiting the calibration time to less than 4 minutes. Thus, r-KLwDSA could help more stroke patients have a meaningful and potentially effective BCI-based rehabilitation.

5.4.6 Impact of number of source sessions on the regularisation value

Figure 5.9 illustrates the effects of the number of available target trials and source sessions on the regularisation value in the proposed r-KLwDSA algorithm. The regularisation value, r, defines a trade-off between the weighted aligned source trials and the target trials in the

final r-KLwDSA model. Figure 5.9 shows larger weights were given to the target trials when more target trials per class were available for training the r-KLwDSA model. For example, given one source session available, when ten target trials per class were used for training, the target trials on average got weighted as $r = 0.52$, whereas the average obtained r was 0.17 when there were two target trials per class available for calibration. These results suggest that when more target trials are getting available for calibration, r-KLwDSA gets more similar to the target session-specific model rather than the transfer learning model extracted from the source sessions. However, when there are only a couple of target trials available for training, although the proposed r-KLwDSA algorithm still finds them useful, the focus has to be on the source data available as a clear representation of the target session cannot be calculated from the limited data available.

When there were less than 5 source sessions available, the effects of the number of source sessions on r was opposite to the effect of the number of target trials. Figure 5.9 shows when there was only one previous source session available, the average r-value was high; however, as the number of sessions available increased, the average weight given to the target trials decreased. This drop in the r-value is presumably due to the increasing amount of source data available for transfer learning. As the amount of transfer learning data increased, the proposed r-KLwDSA could find more source sessions which were similar to the target session and could produce more robust features without relying too much on the available target trials. However, this trend is not consistent as the further increase in the number of available sessions did not lead to a further decrease in the r-value. This end to the trend could be due to the user adapting to using the BCI over time. As the user continues to use the BCI for rehabilitation, they would start learning how to produce more consistent and separable target EEG signals. Due to this the proposed r-KLwDSA algorithm would adapt to this change and increase the r-value to rely more on the target data when calibrating the BCI model.

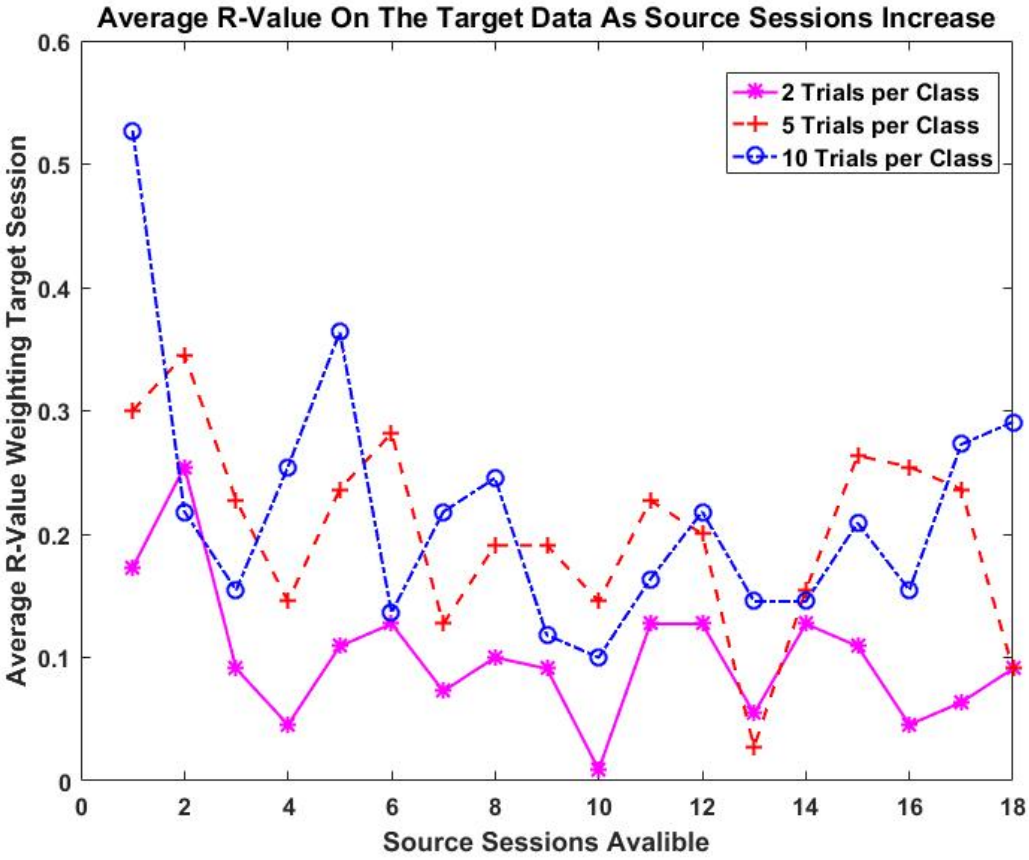

Figure 5.9: The average r-value used for regularising the proposed r-KLwDSA algorithm for each of the target trials per class is plotted against the number of source sessions available. In the proposed r-KLwDSA, the r-value, r, is used to weight the available target trials while $1 - r$ is used to weight the aligned weighted source trials.

5.4.7 Limitations and Future Work

Although the results collected show that the proposed r-KLwDSA performed best for the majority of the stroke patients in a few cases other method performed better. In particular for some users the SS performed much better for a couple of sessions.

Ideally if the correct regularisation parameter r was calculated for each session the proposed r-KLwDSA should always outperform the standard SS algorithm. Utilising regularisation improved the classification accuracy however applying leave-one-out cross validation to select the r value is rather lacking. This method is prone to overfitting due to the limited number of target trials available. Finding a better alternative method to calculate the r value

that works well with limited trials would be very beneficial.

The current proposed r-KLwDSA assumes that there are no non-stationarities within each session which is not an always correct assumption. Further work could be done to reduce the effects of these non-stationarities. Different variants of CSP have been produced to reduce the effects of these non-stationarities such as the KL-CSP and DTW-CSP [92] [93]. Alternatively, online adaptation algorithms have also been developed to reduce these non-stationarities and could further improve the classification accuracy [46].

5.5 Conclusion

This chapter proposed a novel algorithm for transfer learning combining linear alignment, weighting and regularisation to reduce the calibration time for long-term BCI users. The linear alignment aimed to reduce the non-stationarity between the source and target sessions, whereas the weighting mechanism adjusted the impact of each source session on the BCI model based on its similarity to the target data. Finally, the regularisation step combined the weighted aligned source data and the few available target data to build the final BCI model. The proposed algorithm significantly outperformed the session-specific model and a number of other state-of-the -art transfer learning algorithms when the number of available target trials was very few and the number of available source sessions was between 1 to 18. Importantly, the proposed algorithm remarkably reduced the number of BCI sessions with deficient session-specific accuracy (i.e. less than 60%) with an average accuracy improvement of around 10%.

Chapter 6

Conclusion and future work

This thesis has focused on developing novel transfer learning frameworks to reduce the calibration time required for long-term BCI users while maintaining an effective classification accuracy. In order to do this, two key challenges needed to be addressed effectively. Firstly reducing the effects of non-stationarities between the source and target data. Secondly, accounting for the differences between the available source sessions so that the effects of detrimental data are limited while the useful data is prioritised.

6.1 Conclusion

Throughout this thesis, several different transfer learning algorithms and frameworks have been proposed and explored, and their ability to address these challenges has been evaluated. Overall an alignment algorithm, a measure of similarity and two frameworks have been developed and proposed to fulfil the objectives set.

In chapter two, we initially explored and evaluated the current work that has been completed by different studies to resolve these challenges. Through the exploration of the current methods that are being applied, it appeared that transfer learning has the potential to fulfil the objectives set out with alignment and selective transfer learning, particularly areas which show promise in addressing challenges we had identified.

Initially a novel method of alignment, called MDSA, was proposed to align the distributions of EEG data from multiple classes. In chapter three we explored this proposed trans-

fer learning alignment and its ability to reduce effects of non-stationarity between the source and target sessions. By collecting a small number of labelled target session trials, the proposed MDSA algorithm was able to create a linear alignment which minimised the Kullback Liebler divergence between the target and source sessions. The proposed alignment reduced the effects of non-stationarity between raw EEG data before feature extraction or classification. Therefore it could be applied to any feature extraction and classification techniques later if required. The results showed the proposed MDSA improved the mean classification accuracy and even outperformed one state-of-the-art form of continuous adaptation when more than ten target trials per class were available.

Although the proposed MDSA algorithm showed some success, it also highlighted some issues and limitations that need to be addressed. The first is the limited nature of the proposed MDSA. By aligning the target session to a source session, the transfer learning technique is limited to that single source session, as the target session would not be able to directly align with multiple source sessions. To counter this, the proposed MDSA was later altered to align the source session to the target session. This allowed for the inclusion of multiple source sessions while retaining the MDSA's benefit. Another limitation of the proposed MDSA which we sadly did not get time to address, was the assumption that the EEG non-stationarity did not have an effect withing each source/target session. An interesting final point discovered while developing this alignment was the variations between source sessions. It was found that for some target sessions, when transfer learning was applied, even in a naive form, it significantly improved the classification accuracy, while for others, the effect was the opposite. In fact for some target sessions a BCI trained using the few labelled trials from the target session outperformed any transfer learning method explored. From this, the importance of selective transfer learning became clear.

In chapter four we proposed a method to identify the target sessions that would benefit from transfer learning as well as a new measurement of similarity between sessions. Subsequently, we created a framework which first identified the target sessions that would benefit from transfer learning and then found the source session with the highest similarity to train the model. A significant correlation between sessions that found transfer learning detrimen-

tal and sessions that obtained high levels of classification accuracy with session-specific BCIs were found. As such, to identify the target session that performed best without transfer learning, leave-one-out validation was applied using the few available labelled trials from the target session. Target sessions obtaining above 70% average leave-one-out classification accuracy did not use transfer learning, while the other sessions utilised transfer learning. For these sessions, the Jensen Shannon Ratio (JSR) was proposed. A measurement which worked out the ratio between the similarity of the same classes in the target and source session and the similarity between the opposite classes in the target and source sessions. This proposed measurement selected one source session to train the BCI for the target data. This similarity measure had a statistically significant improvement of 8% against a standard BCI for the sessions that encountered BCI deficiency. Importantly, the proposed leave-one-out method for selecting whether or not to implement transfer learning identified the correct approach around two-thirds of the times. To summarise, the proposed framework on average improved the classification accuracy even when only four target trials per class where available; however, both approaches contained some limitations.

Finally, in chapter five, we proposed a new framework, called r-KLDSA, to build on the previous framework exploring inter-session instead of inter-subject transfer learning, using a similarity measurement to weight source sessions instead of selecting a single source session and regularising the target and source sessions instead of just selecting between transfer learning and session-specific. Along with this, the proposed r-KLwDSA framework also incorporated the previously proposed MDSA alignment algorithm, improved as previously explained so that it can work on aligning multiple source sessions to the target session. This framework initially utilised alignment to address the challenge of the non-stationary nature of the EEG. Then regularisation and weighted transfer learning to address the variations between all the source and target sessions. This proposed framework was evaluated using the nBetter data set, which contained data collected from 11 subjects with the arm weakness due to stroke. Each subject provided 18 BCI sessions over six weeks. As one key application of BCI currently being developed is for stroke rehabilitation, this data set contained very interesting data. In order to try and replicate a real-world scenario, the data set was evaluated chronologically, with each

session being added to the pool of available source sessions only after it had been a target session. As such, only one source session was available for session two, while session eighteen had seventeen source sessions available.

The proposed r-KLDSA provided encouraging results improving classification accuracy for all the subjects with as few as two labelled target trials per class. Sessions that encountered BCI deficiency when using session-specific BCI, in particular, had a statistically significant mean improvement of over 13% classification accuracy when ten labelled target session trials per class were used. Along with these results, a lot of interesting points became apparent, such as how some target sessions perform better with fewer source sessions being available; potentially, the changes between each session accumulate until the older source sessions only provide detrimental data.

Overall the final proposed framework addressed the challenges laid out at the start and completed the objective of reducing the calibration time required by the BCI while improving the accuracy. This reduction in the calibration time will lead to a more efficient BCI technology, which can be set up in a shorter time frame, making them much more practical for real-world applications.

6.2 Future Works

The proposed frameworks in this thesis could potentially be improved with more work to address the limitations identified during the thesis and other general challenges in BCI.

- The proposed MDSA algorithm designed in chapter three aligns the data spaces linearly. It is interesting to investigate if non-linear alignments could be useful. Moreover, the MDSA works regardless of the feature extraction method explored or the classifier. Due to their reliability and common use, the common spatial patterns (CSP) filter was used for feature selection, while linear discrete analysis was used for classification. Other methods have been explored briefly for comparison; however, more work could be explored using different feature extraction and classification approaches within the framework.

- In chapter four, a new method of similarity was proposed called the Jensen Shannon

ratio. This measure performed well when evaluated using the BCI Competition IV data set 2a; however, once it was applied to the nBetter stroke data set, it was found that relying on Kullback Liebler divergence performed better. Although this was the case for the nBetter data set, evaluating the proposed JSR on more data sets would be interesting.

- In chapters four and five, two frameworks are proposed to reduce the need for a long calibration session. Although effective, these frameworks are only evaluated on their ability to classify two classes of EEG data. Increasing the number of classes the algorithm can classify would significantly impact the BCIs' real-world applications.

- In chapter five, the best regularisation parameter was selected among a set of predefined values using cross-validation on the available target trials. Although the proposed method proved effective, it is still far from optimal, especially when the labelled target trials are limited. Exploring other regularisation methods to find the optimal regularisation parameter with the limited number of trials would be interesting.

- In chapter five it also became clear the it is not necessarily the best approach to provide a large number of previously recorded sessions as the source sessions for the same user. An interesting area which could be explored is identifying the best number of source sessions for inter-session transfer learning. This probably wouldn't be a set number but it may be possible to develop an algorithm possibly based on similarity.

- Throughout the thesis, several algorithms are proposed and explored; however, in all cases, it is assumed there is no non-stationarity between trials. Although the proposed work reduces the effects of non-stationarity between sessions, they could be improved by extending the framework to accommodate continuous adaptations.

- One data set containing data collected from stroke patients over an extended period was used to explore the proposed r-KLwDSA framework along with one publicly available dataset containing two sessions for each subject. Evaluating this framework with more data sets containing EEG collected over an extended time could be interesting to explore the idea of intra-subject transfer learning further.

- The proposed algorithms in this thesis were all evaluated using offline data collected previously. The results collected appear encouraging, but an online experiment could be useful. As much as we can try to simulate a real-world experiment by evaluating the data chronologically, there are limits to this. Testing these algorithms in real-world experiments would reveal new flaws that need to be addressed.

- An interesting area which still requires much research is understanding the co-adaptation process between the BCI and the user. The proposed frameworks utilise machine learning to reduce the effects of the non-stationary nature of the EEG signals; however, it does not account for the learning process. The user can be also trained to reduce the effects of non-stationarity in EEG signals by generating more discriminative and stable brain signals. Importantly, when BCIs are used for rehabilitation, it is important to obtain high levels of classification accuracy but also to train the brain.

Chapter 7

Appendix

7.1 Mathematical Proof of Equation (5.7)

To obtain the optimal $\mathbf{L}_j$, presented in (5.7), we need to calculate the first-order derivative of (5.4), with respect to $\mathbf{L}_j$, and set it to zero. To do this we use the properties presented below [94]:

$$\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CAB}) = \mathrm{tr}(\mathbf{CBA}) \tag{S1}$$

$$\frac{d}{d\mathbf{L}_j}\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\frac{d\mathbf{A}}{d\mathbf{L}_\mathrm{j}}) \tag{S2}$$

$$\frac{d}{d\mathbf{L}_j}\det(\mathbf{A}) = \det(\mathbf{A})\mathrm{tr}(\mathbf{A}^{-1}\frac{d\mathbf{A}}{d\mathbf{L}_\mathrm{j}}) \tag{S3}$$

where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are real matrices. As the EEG covariance matrices are positive and symmetric we can conclude that:

$$\frac{d}{d\mathbf{L}}\mathrm{tr}(\overline{\mathbf{\Sigma}}^{\mathrm{c}\dagger}\mathbf{L_j}\hat{\mathbf{\Sigma}}_\mathrm{j}^\mathrm{c}\mathbf{L_j}^\mathrm{T}) = \frac{d}{d\mathbf{L}}\mathrm{tr}(\hat{\mathbf{\Sigma}}_\mathrm{j}^\mathrm{c}\mathbf{L_j}\overline{\mathbf{\Sigma}}^{\mathrm{c}\dagger}\mathbf{L_j}^\mathrm{T}) \tag{S4}$$

$$\frac{d}{d\mathbf{L}}\mathrm{tr}(\overline{\mathbf{\Sigma}}^{\mathrm{c}\dagger}\mathbf{L_j}\hat{\mathbf{\Sigma}}_\mathrm{j}^\mathrm{c}\mathbf{L_j}^\mathrm{T}) = 2\mathrm{tr}(\overline{\mathbf{\Sigma}}^{\mathrm{c}\dagger}\mathbf{L_j}\hat{\mathbf{\Sigma}}_\mathrm{j}^\mathrm{c}) = 2\mathrm{tr}(\hat{\mathbf{\Sigma}}_\mathrm{j}^\mathrm{c}\overline{\mathbf{\Sigma}}^{\mathrm{c}\dagger}\mathbf{L_j}) \tag{S5}$$

$$\frac{d}{d\mathbf{L}_j}\ln(\det(\mathbf{L_j}\hat{\mathbf{\Sigma}}_\mathbf{j}^\mathbf{c}\mathbf{L_j}^\mathrm{T})) = \frac{\det(\mathbf{L_j}\hat{\mathbf{\Sigma}}_\mathbf{j}^\mathbf{c}\mathbf{L_j}^\mathrm{T})}{\det(\mathbf{L_j}\hat{\mathbf{\Sigma}}_\mathbf{j}^\mathbf{c}\mathbf{L_j}^\mathrm{T})}2\mathrm{tr}(\hat{\mathbf{\Sigma}}_\mathbf{j}^\mathbf{c}\mathbf{L_j}^\mathrm{T}(\mathbf{L_j}\hat{\mathbf{\Sigma}}_\mathbf{j}^\mathbf{c}\mathbf{L_j}^\mathrm{T})^{-1}) \tag{S6}$$

$$\frac{d}{d\mathbf{L}_j}\ln(\det(\mathbf{L_j}\hat{\mathbf{\Sigma}}_\mathbf{j}^\mathbf{c}\mathbf{L_j}^{\mathrm{T}})) = \frac{d}{d\mathbf{L}_j}\ln(\det(\mathbf{L_j}\hat{\mathbf{\Sigma}}_\mathbf{j}^\mathbf{c}\mathbf{L_j}^{\mathrm{T}})) = 2\mathrm{tr}(\mathbf{L_j}^{-1}) \tag{S7}$$

By substituting equation (S5) and (S7) into equation (5.6), we find

$$\frac{dA}{d\mathbf{L}_j} = \sum_{c=1}^{2}\mathrm{tr}(\hat{\mathbf{\Sigma}}_\mathbf{j}^\mathbf{c}\overline{\mathbf{\Sigma}}^{\mathbf{c}\dagger}\mathbf{L_j} - \mathbf{L_j}^{-1}) = 0 \tag{S8}$$

One of the solutions for equation (S8) is

$$\hat{\mathbf{\Sigma}}_j^1\overline{\Sigma}^{1\dagger}\mathbf{L_j} + \hat{\mathbf{\Sigma}}_j^2\overline{\mathbf{\Sigma}}^{2\dagger}\mathbf{L_j} - 2\mathbf{L_j}^{-1} = 0 \tag{S9}$$

This solution can then be simply re-arranged as equation (S9).

Bibliography

[1] A. J. Busti and D. Kellogg, "Homunculus: Somatosensory and Somatomotor Cortex," 2015.

[2] T. W. Berger, J. K. Chapin, G. A. Gerhardt, D. J. McFarland, J. C. Principe, W. V. Soussou, D. M. Taylor, and P. A. Tresco, *Brain-Computer Interfaces*. Dordrecht: Springer Netherlands, 2008.

[3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control.," *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–91, 2002.

[4] N. Sharma, V. M. Pomeroy, and J. C. Baron, "Motor imagery: A backdoor to the motor system after stroke?," *Stroke*, vol. 37, no. 7, pp. 1941–1952, 2006.

[5] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. Ramsey, I. Sturm, G. Curio, and K. R. Müller, "The Berlin brain-computer interface: Non-medical uses of BCI technology," *Frontiers in Neuroscience*, vol. 4, no. DEC, pp. 1–17, 2010.

[6] D. Zapała, P. Francuz, E. Zapała, N. Kopiś, P. Wierzgała, P. Augustynowicz, A. Majkowski, and M. Kołodziej, "The impact of different visual feedbacks in user training on motor imagery control in bci," *Applied psychophysiology and biofeedback*, vol. 43, no. 1, pp. 23–35, 2018.

[7] R. Foong, K. K. Ang, C. Quek, C. Guan, K. S. Phua, C. W. K. Kuah, V. A. Deshmukh, L. H. L. Yam, D. K. Rajeswaran, N. Tang, E. Chew, and K. S. G. Chua, "Assessment of the

Efficacy of EEG-Based MI-BCI With Visual Feedback and EEG Correlates of Mental Fatigue for Upper-Limb Stroke Rehabilitation," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 3, pp. 786–795, 2020.

[8] M. Krauledat, M. Schröder, B. Blankertz, and K.-R. Müller, "Reducing calibration time for brain-computer interfaces: A clustering approach," *Adv. Neural Information Processing Syst.*, vol. 19, p. 753, 2007.

[9] A. M. Azab, J. Toth, L. S. Mihaylova, and M. Dr Arvaneh, "A review on transfer learning approaches in brain–computer interface," in *Signal Processing and Machine Learning for Brain-Computer Interfaces*, ch. 5, pp. 96 – 116, 2018.

[10] T. Dikanev, D. Smirnov, R. Wennberg, J. L. P. Velazquez, and B. Bezruchko, "EEG nonstationarity during intracranially recorded seizures: statistical and dynamical analysis," *Clinical Neurophysiology*, vol. 116, no. 8, pp. 1796–1807, 2005.

[11] Z. Peng, W. Zhang, N. Han, X. Fang, P. Kang, and L. Teng, "Active Transfer Learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 1022–1036, 2020.

[12] I. Hossain, A. Khosravi, and S. Nahavandhi, "Active transfer learning and selective instance transfer with active learning for motor imagery based BCI," in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 4048–4055, 2016.

[13] S. Perdikis and J. d. R. Millan, "Brain-Machine Interfaces: A Tale of Two Learners," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 6, pp. 12–19, jul 2020.

[14] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of EEG-based brain–computer interface paradigms," *Journal of Neural Engineering*, vol. 16, p. 11001, jan 2019.

[15] J. Giles, K. Ang, L. Mihaylova, and M. Arvaneh, "Data Space Adaptation for Multiclass Motor Imagery-based BCI," in *EMBC 2018*, IEEE, apr 2018.

[16] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kübler, "An meg-based brain–computer interface (bci)," *NeuroImage*, vol. 36, no. 3, pp. 581–593, 2007.

[17] L. Parkkonen, "Real-time magnetoencephalography for neurofeedback and closed-loop experiments," in *Clinical Systems Neuroscience*, pp. 315–330, Springer, 2015.

[18] S. Fazli, J. Mehnert, J. Steinbrink, and B. Blankertz, "Using nirs as a predictor for eeg-based bci performance," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4911–4914, IEEE, 2012.

[19] J. W. Christoph M. Michel, Thomas Koenig, Daniel Brandeis, Lorena R. R. Gianotti, *Electrical Neuroimaging*. Cambridge University Press, 2009.

[20] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of eeg-based brain–computer interface paradigms," *Journal of neural engineering*, vol. 16, no. 1, p. 011001, 2019.

[21] A. Searle and L. Kirkup, "A direct comparison of wet, dry and insulating bioelectric recording electrodes," *Physiological Measurement*, vol. 21, no. 2, pp. 271–283, 2000.

[22] C. Grozea, C. D. Voinescu, and S. Fazli, "Bristle-sensors—low-cost flexible passive dry eeg electrodes for neurofeedback and bci applications," *Journal of neural engineering*, vol. 8, no. 2, p. 025008, 2011.

[23] J. R. Daube and D. I. Rubin, *Clinical Neurophysiology*. Oxford university press, 3 ed., 2009.

[24] L. Koessler, L. Maillard, A. Benhadid, J. P. Vignal, J. Felblinger, H. Vespignani, and M. Braun, "Automated cortical projection of EEG sensors: Anatomical correlation via the international 10-10 system," *NeuroImage*, vol. 46, no. 1, pp. 64–72, 2009.

[25] Y. Chen, Y. Ke, G. Meng, J. Jiang, H. Qi, X. Jiao, M. Xu, P. Zhou, F. He, and D. Ming, "Enhancing performance of P300-Speller under mental workload by incorporating dual-task data during classifier training," *Computer Methods and Programs in Biomedicine*, vol. 152, pp. 35–43, 2017.

[26] J. Minguillon, M. A. Lopez-Gordo, and F. Pelayo, "Trends in EEG-BCI for daily-life: Requirements for artifact removal," *Biomedical Signal Processing and Control*, vol. 31, pp. 407–418, 2017.

[27] R. Ortner, J. Scharinger, A. Lechner, and C. Guger, "How many people can control a motor imagery based BCI using common spatial patterns ?," *7th Annual International IEEE EMBS Conference on Neural Engineering*, pp. 22–24, 2015.

[28] Z. Y. Chin, K. K. Ang, and C. Guan, "Multiclass voluntary facial expression classification based on Filter Bank Common Spatial Pattern," *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08*, pp. 1005 – 1008, 2008.

[29] S. Kumar, A. Sharma, and T. Tsunoda, "An improved discriminative filter bank selection approach for motor imagery eeg signal classification using mutual information," *BMC bioinformatics*, vol. 18, pp. 125–137, 2017.

[30] R. Aler, I. M. Galván, and J. M. Valls, "Evolving spatial and frequency selection filters for brain-computer interfaces," in *IEEE congress on evolutionary computation*, pp. 1–7, IEEE, 2010.

[31] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American statistical association*, vol. 84, no. 405, pp. 165–175, 1989.

[32] U. Chaudhary, N. Mrachacz-Kersting, and N. Birbaumer, "Neuropsychological and neurophysiological aspects of brain-computer-interface (bci) control in paralysis," *The Journal of physiology*, vol. 599, no. 9, pp. 2351–2359, 2021.

[33] R. Foong, K. K. Ang, C. Quek, C. Guan, K. S. Phua, C. W. K. Kuah, V. A. Deshmukh, L. H. L. Yam, D. K. Rajeswaran, N. Tang, *et al.*, "Assessment of the efficacy of eeg-based mi-bci with visual feedback and eeg correlates of mental fatigue for upper-limb stroke rehabilitation," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 3, pp. 786–795, 2019.

[34] C. D. Wolfe, "The impact of stroke," *British medical bulletin*, vol. 56, no. 2, pp. 275–286, 2000.

[35] S. Ojaghihaghighi, S. S. Vahdati, A. Mikaeilpour, and A. Ramouz, "Comparison of neurological clinical manifestation in patients with hemorrhagic and ischemic stroke," *World Journal of Emergency Medicine*, vol. 8, no. 1, pp. 34–38, 2017.

[36] A. A. Frolov, O. Mokienko, R. Lyukmanov, E. Biryukova, S. Kotov, L. Turbina, G. Nadareyshvily, and Y. Bushkova, "Post-stroke rehabilitation training with a motor-imagery-based brain-computer interface (BCI)-controlled hand exoskeleton: A randomized controlled multicenter trial," *Frontiers in Neuroscience*, vol. 11, no. JUL, 2017.

[37] D. T. Bundy, L. Souders, K. Baranyai, L. Leonard, G. Schalk, R. Coker, D. W. Moran, T. Huskey, and E. C. Leuthardt, "Contralesional Brain-Computer Interface Control of a Powered Exoskeleton for Motor Recovery in Chronic Stroke Survivors," *Stroke*, vol. 48, no. 7, pp. 1908–1915, 2017.

[38] S. Brandl, L. Frølich, J. Höhne, K.-R. Müller, and W. Samek, "Brain–computer interfacing under distraction: an evaluation study," *Journal of neural engineering*, vol. 13, no. 5, p. 056012, 2016.

[39] F. Lotte, F. Larrue, and C. Mühl, "Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design," *Frontiers in human neuroscience*, vol. 7, p. 568, 2013.

[40] A. Myrden and T. Chau, "Effects of user mental state on eeg-bci performance," *Frontiers in human neuroscience*, vol. 9, p. 308, 2015.

[41] H. Raza, D. Rathee, S.-M. Zhou, H. Cecotti, and G. Prasad, "Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related eeg-based brain-computer interface," *Neurocomputing*, vol. 343, pp. 154–166, 2019.

[42] P. Wang, J. Lu, B. Zhang, and Z. Tang, "A review on transfer learning for brain-computer interface classification," in *2015 5th International Conference on Information Science and Technology (ICIST)*, pp. 315–322, IEEE, 2015.

[43] A. Quattoni, M. Collins, and T. Darrell, "Transfer learning for image classification with sparse prototype representations," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, IEEE, 2008.

[44] S. J. Pan, V. W. Zheng, Q. Yang, and D. H. Hu, "Transfer learning for wifi-based indoor localization," in *Association for the advancement of artificial intelligence (AAAI) workshop*, vol. 6, The Association for the Advancement of Artificial Intelligence Palo Alto, 2008.

[45] L. Chen, "Deep transfer learning for static malware classification," *arXiv preprint arXiv:1812.07606*, 2018.

[46] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "EEG Data Space Adaptation to Reduce Intersession Nonstationarity in Brain-Computer Interface," *Neural Computation Massachusetts Institute of Technology*, vol. 25, pp. 2146–2171, 2013.

[47] F. Wang, J. Ping, Z. Xu, and J. Bi, "Classification of motor imagery using multisource joint transfer learning," *Review of Scientific Instruments*, vol. 92, no. 9, p. 094106, 2021.

[48] Y. Li, H. Kambara, Y. Koike, and M. Sugiyama, "Application of covariate shift adaptation techniques in brain-computer interfaces.," *IEEE transactions on bio-medical engineering*, vol. 57, pp. 1318–1324, jun 2010.

[49] H. Raza, H. Cecotti, Y. Li, and G. Prasad, "Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface," *Soft Computing*, vol. 20, no. 8, pp. 3085–3096, 2016.

[50] H. He and D. Wu, "Transfer learning for brain–computer interfaces: A euclidean space data alignment approach," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 399–410, 2019.

[51] P. Aggarwal, S. F. Ahamed, S. Shetty, and L. J. Freeman, "Selective targeted transfer learning for malware classification," in *2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*, pp. 114–120, IEEE, 2021.

[52] Z. Chen, C. Chen, Z. Cheng, B. Jiang, K. Fang, and X. Jin, "Selective transfer with reinforced transfer network for partial domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12706–12714, 2020.

[53] Y. Sun, L. Lan, X. Zhao, M. Fan, Q. Guo, and C. Li, "Selective multi-source transfer learning with wasserstein domain distance for financial fraud detection," in *BenchCouncil International Federated Intelligent Computing and Block Chain Conferences*, pp. 489–505, Springer, 2020.

[54] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted transfer learning for improving motor imagery-based brain–computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1352–1359, 2019.

[55] Y. Xu, X. Huang, and Q. Lan, "Selective cross-subject transfer learning based on riemannian tangent space for motor imagery brain-computer interface," *Frontiers in Neuroscience*, vol. 15, 2021.

[56] F. Lotte and C. Guan, "Learning from other subjects helps reducing brain-computer interface calibration time," in *2010 IEEE International conference on acoustics, speech and signal processing*, pp. 614–617, IEEE, 2010.

[57] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Omitting the intra-session calibration in EEG-based brain computer interface used for stroke rehabilitation," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4124–4127, aug 2012.

[58] J. Giles, K. K. Ang, L. Mihaylova, and M. Arvaneh, "Data Space Adaptation for Multiclass Motor Imagery-based BCI," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2018.

[59] H. Raza, D. Rathee, S.-M. Zhou, H. Cecotti, and G. Prasad, "Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related EEG-based brain-computer interface," *Neurocomputing*, vol. 343, pp. 154–166, 2019.

[60] D. Wu, "Online and offline domain adaptation for reducing bci calibration effort," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 4, pp. 550–563, 2017.

[61] H. Kang and S. Choi, "Bayesian common spatial patterns for multi-subject EEG classification," *Neural Networks*, vol. 57, pp. 39–50, 2014.

[62] C. Vidaurre, M. Kawanabe, P. Von Bünau, B. Blankertz, and K. R. Müller, "Toward unsupervised adaptation of LDA for brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3 PART 1, pp. 587–597, 2011.

[63] C. Vidaurre, C. Sannelli, K. R. Müller, and B. Blankertz, "Co-adaptive calibration to improve BCI efficiency," *Journal of Neural Engineering*, vol. 8, no. 2, 2011.

[64] L. F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Álvarez, and R. Hornero, "Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain-computer interfaces," *Neurocomputing*, vol. 159, no. 1, pp. 186–196, 2015.

[65] W. Wu, X. Gao, and S. Gao, "One-versus- the-rest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case," *Engineering In Medicine And Biology*, vol. 3, pp. 2387–2390, 2005.

[66] A. Llera, V. Gómez, and H. J. Kappen, "Adaptive Multiclass Classification for Brain Computer Interfaces," *Neural Computation Massachusetts Institute of Technology*, vol. 26, no. 6, pp. 1108–1127, 2014.

[67] P. Xu, P. Yang, X. Lei, and D. Yao, "An enhanced probabilistic LDA for multi-class brain computer interface," *PLoS ONE*, vol. 6, no. 1, 2011.

[68] C. Brunner, R. Leeb, G. R. Muller-Putz, A. Schlogl, and G. Pfurtscheller, "BCI Competition 2008 - Graz data set A," 2008.

[69] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.

[70] M. Tangermann, K. R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk,

A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI competition IV," 2012.

[71] J. Giles, K. K. Ang, L. S. Mihaylova, and M. Arvaneh, "A Subject-to-subject Transfer Learning Framework Based on Jensen-Shannon Divergence for Improving Brain-computer Interface," in *ICASSP 2019*, pp. 3087 – 3091, 2019.

[72] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, no. MAR, pp. 1–9, 2012.

[73] M. Arvaneh, I. Robertson, and T. E. Ward, "Subject-to-subject adaptation to reduce calibration time in motor imagery-based brain-computer interface," in *EMBC 2014*, pp. 6501–6504, 2014.

[74] V. Jayaram, M. Alamgir, Y. Altun, and M. Grosse-wentrup, "Transfer Learning in Brain-Computer Interfaces," *arXiv:1512.00296*, no. February, pp. 20–31, 2016.

[75] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Bio-Med. Eng.*, vol. 58, no. 2, pp. 355–362, 2011.

[76] H. Kang, Y. Nam, and S. Choi, "Composite Common Spatial Patterns for Subject-to-Subject Transfer," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, 2009.

[77] H. He and D. Wu, "Transfer Learning for Brain-Computer Interfaces : An Euclidean Space Data Alignment Approach," *arXiv:1808.05464*, pp. 1–10, 2018.

[78] F. Lotte and C. Guan, "Learning from other Subjects Helps Reducing Brain-Computer Interface Calibration Time," *International Conference on Audio Speech and Signal Processing (ICASSP), Mar 2010, Dallas, United States*, 2010.

[79] J. Giles, K. Ang, L. Mihaylova, and M. Arvaneh, "Weighted Transfer Learning of Dynamic Time Warped Data for Motor Imagery based Brain Computer Interfaces," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2020-July, 2020.

[80] J. Giles, K. Ang, K. S. Phua, and M. Arvaneh, "A Transfer Learning Algorithm to Reduce Brain-Computer Interface Calibration Time for Long-Term Users," *Frontiers in Neuroergonomics*, vol. 3, p. 837307, 2022.

[81] M. Sugiyama, S. Nakajima, H. Kashima, P. Buenau, and M. Kawanabe, "Direct Importance Estimation with Model Selection and Its Application to Covariate Shift Adaptation," in *Advances in Neural Information Processing Systems* (J. Platt, D. Koller, Y. Singer, and S. Roweis, eds.), vol. 20, pp. 1433–1440, Curran Associates, Inc., 2008.

[82] H. Raza, H. Cecotti, Y. Li, and G. Prasad, "Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface," *Soft Computing*, vol. 20, no. 8, pp. 3085–3096, 2016.

[83] S. Dalhoumi, G. Dray, J. Montmain, G. Derosière, and S. Perrey, "An adaptive accuracy-weighted ensemble for inter-subjects classification in brain-computer interfacing," in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 126–129, apr 2015.

[84] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted Transfer Learning for Improving Motor Imagery-Based Brain-Computer Interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1352–1359, 2019.

[85] S. Perdikis, L. Tonin, S. Saeedi, C. Schneider, and J. d. R. Millán, "The Cybathlon BCI race: Successful longitudinal mutual learning with two tetraplegic users," *PLOS Biology*, vol. 16, no. 5, pp. 1–28, 2018.

[86] D. Novak, R. Sigrist, N. J. Gerig, D. Wyss, R. Bauer, U. Götz, and R. Riener, "Benchmarking Brain-Computer Interfaces Outside the Laboratory: The Cybathlon 2016," *Frontiers in Neuroscience*, vol. 11, p. 756, 2018.

[87] L. Hehenberger, R. J. Kobler, C. Lopes-Dias, N. Srisrisawang, P. Tumfart, J. B. Uroko, P. R. Torke, and G. R. Müller-Putz, "Long-Term Mutual Training for the CYBATHLON BCI Race With a Tetraplegic Pilot: A Case Study on Inter-Session Transfer and Intra-Session Adaptation," *Frontiers in Human Neuroscience*, vol. 15, p. 70, 2021.

[88] C. Benaroch, K. Sadatnejad, A. Roc, A. Appriou, T. Monseigne, S. Pramij, J. Mladenovic, L. Pillette, C. Jeunet, and F. Lotte, "Long-Term BCI Training of a Tetraplegic User: Adaptive Riemannian Classifiers and User Training," *Frontiers in Human Neuroscience*, vol. 15, p. 118, 2021.

[89] C. S. Wei, Y. P. Lin, Y. T. Wang, T. P. Jung, N. Bigdely-Shamlo, and C. T. Lin, "Selective Transfer Learning for EEG-Based Drowsiness Detection," jan 2016.

[90] L. Cao, S. Chen, J. Jia, C. Fan, H. Wang, and Z. Xu, "An Inter- and Intra-Subject Transfer Calibration Scheme for Improving Feedback Performance of Sensorimotor Rhythm-Based BCI Rehabilitation," *Frontiers in Neuroscience*, vol. 14, p. 1444, 2021.

[91] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis," *IEEE SIGNAL PROCESSING MAGAZINE*, vol. XX, 2008.

[92] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing Spatial Filters by Minimizing Within-Class Dissimilarities in Electroencephalogram-Based Brain–Computer Interface," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 4, pp. 610–619, 2013.

[93] A. M. Azab, L. Mihaylova, H. Ahmadi, and M. Arvaneh, "Robust Common Spatial Patterns Estimation Using Dynamic Time Warping to Improve BCI Systems," in *ICASSP*, 2019.

[94] K. B. Petersen, M. S. Pedersen, *et al.*, "The matrix cookbook," *Technical University of Denmark*, vol. 7, no. 15, p. 510, 2008.