NOVEL CLUSTER-BASED SVM TO REDUCE CLASSIFICATION ERROR IN NOISY EEG DATA: TOWARDS REAL-TIME BRAIN-ROBOT INTERFACES

Martin Johansson
mjn13021@student.mdh.se

Examiner: Ning Xiong
Mälardalen University, Västerås, Sweden

Supervisor: Elaine Åstrand
Mälardalen University, Västerås, Sweden

Supervisor: Jonatan Tidare
Mälardalen University, Västerås, Sweden

June 7, 2018
Abstract

To be able to control a robotic platform using signals from the human brain is something that has been considered science fiction for a long time. With the technology available today however, it has more or less become a reality. Electroencephalography or EEG for short is a widely used method for extracting signals from the brain. The signals in this thesis contain motor imagery (MI) commands that are supposed to be sent through a brain computer interface (BCI) to control a mobile robot. This thesis investigates deeper into classification of these signals, specifically optimising classification accuracy of noisy EEG data, that previously has been unsatisfactorily classified with support vector machine (SVM). It is paramount that the classification accuracy is as high as possible when used in a BCI since the robot can cause damage if the commands are faulty. A new cluster-based SVM is developed that discards uncertain trials and minimises the false positive rate in an attempt to increase the accuracy. The algorithm increases the classification accuracy compared to SVM alone by 8 percentage points. Alongside this new algorithm, eye movement artefacts and separability of the MI commands are analysed to further investigate classification accuracy influences.
# Table of Contents

1 Introduction .................................................. 4

2 Problem Formulation ........................................... 5

3 Hypothesis ...................................................... 6

4 Background ...................................................... 7
   4.1 Electroencephalogram ........................................ 7
      4.1.1 EEG setup ........................................ 7
      4.1.2 EEG signals ........................................ 7
      4.1.3 Pre-processing ........................................ 7
   4.2 Brain Computer Interface ..................................... 8
      4.2.1 Classification ........................................ 8
      4.2.2 Robot control ........................................ 9
      4.2.3 Visual feedback ....................................... 10

5 Expected Outcome .............................................. 11

6 Method .......................................................... 12
   6.1 Pre-processing stage ......................................... 12
      6.1.1 Signal acquisition .................................... 12
      6.1.2 Data ................................................ 12
      6.1.3 Feature extraction .................................... 12
   6.2 Data analysis ................................................ 13
      6.2.1 Eye artefact removal .................................. 13
      6.2.2 Cluster analysis ...................................... 13
   6.3 Classification ............................................... 13
      6.3.1 Support Vector Machine ................................ 13
      6.3.2 Cluster-based SVM .................................... 13

7 Limitations ..................................................... 14

8 Design .......................................................... 15
   8.1 Eye-blink component removal ................................ 15
   8.2 Feature extraction ........................................... 15
   8.3 Cluster separability analysis ................................ 15
   8.4 Classification ............................................... 18
      8.4.1 Cluster-based SVM .................................... 18

9 Results .......................................................... 19
   9.1 Eye artefact removal ........................................ 19
   9.2 Cluster analysis ............................................. 19
   9.3 Classification ............................................... 22
      9.3.1 Support Vector Machine ................................ 22
      9.3.2 Cluster-based SVM .................................... 23

10 Discussion ..................................................... 24
   10.1 Eye artefact removal ....................................... 24
   10.2 Cluster separability analysis ................................ 25
   10.3 Classification accuracy and performance ..................... 25

11 Ethics .......................................................... 27

12 Conclusion ..................................................... 28
   12.1 Future work ................................................ 28
13 Acknowledgements

References
1 Introduction

The ability to control a vehicle using only your brain and without moving any muscle contributes a promising technique for our society. Not least for people with a movement hindering disability. Electroencephalogram (EEG) is a non-invasive, portable and relatively inexpensive recording technique that measures the ongoing brain activity with high temporal resolution. Thanks to these advantages, it is often used as a tool in Brain-Computer Interface (BCI) based robot controllers for different vehicles. For example, a BCI-manoeuvred wheelchair is only one out of many promising applications [1].

EEG has relatively low signal-to-noise ratio as well as low spatial resolution which means that the signals needs to undergo several steps of signal processing before it can be utilised in a BCI robot controller. This can be accomplished with bandpass filters and artefact removal algorithms together with constantly evolving classification and feature extraction methods.

Even with well pre-possessed EEG data, the BCI has limitations when the brain is used by itself to control the robot. Mental commands generated from the EEG by today’s standards can be unreliable due to a variety off different sources such as mental fatigue or muscle signal noise. Bäckström and Tidare (2016) attempted to overcome these issues by developing a reliable BCI based robot controller that only partly uses the command generated by the EEG data [2]. Specifically, the mental command extracted from the brain activity was merged with environmental information from a laser scan plot generated by the vehicle that is supposed to be controlled. The direction and velocity of the vehicle was determined by the relative weight of each input in the objective of reliably and intuitively manoeuvring a vehicle in a simulated environment. The mental command was extracted from recorded EEG activity simultaneously as the subject was imagining a pre-defined body movement. The intelligent system, consisting of intelligent interpretation of the laser scan data, would help with the navigation if the EEG signals were too noisy or unreadable and to make decisions when the mental commands in question are limited.

With learning algorithms, the mental commands produced by the EEG signals can be classified and then recognised for later use. This is done by selecting specific features for these mental commands and then using a classification algorithm to be trained to recognise these features. The classification in Bäckström and Tidare’s work generated an over-representation of wrongly classified classes which supposedly obscured the navigation of the robot. For a robot controller it is paramount that the classified signals is of the correct task, otherwise the robot might turn the completely wrong way which can have dire consequences. This specific reason makes it more valuable to toss uncertain trials to increase the quality at the expense of generating less mental commands. There are various classification methods that utilises threshold filtering which could possibly increase the classification accuracy by generating true and false positives. The algorithm can then be targeting the false positives to try and minimise its presence. Krauss et al. proposed an interesting cluster analysis method which could be expanded in to a threshold filtering classification algorithm to filter out false positives [3]. This is one of the research areas that has been investigated in this thesis.

Bäckström and Tidare did their work in an offline scenario where the mental imagery data was recorded beforehand and was applied to the robot controller with the use of a reference joystick. To get visual input from the robot while doing the mental commands could supposedly improve the accuracy of the robot [4, 5]. This however presents new challenges like filtering, artefact removal and classifying without too much of a delay. This thesis takes this in consideration for eventual future real-time usage.

In this thesis the above-described issues and challenges are investigated in order to implement a classification algorithm that shall increase the classification accuracy of the EEG data Bäckström and Tidare collected during their thesis. Specifically a threshold filtering method to decrease the amount of false positives.
2 Problem Formulation

How can we improve the classification that Bäckström and Tidare did on the EEG-derived imaginary body movements?

- How does a blink and eye movement reduction algorithm affect the EEG data and classification accuracy?
- Can we find another algorithm that increases the classification accuracy?
- Is there an algorithm or method that increases the classification performance?

Question one aims to find out if a blink reduction algorithm is a necessity for the BCI to work properly. As of today, a blink reduction algorithm for online usage cannot be performed without delay. Question two aims to investigate classification algorithms or improve the one Bäckström and Tidare used in order to increase the classification accuracy. Question three aims to investigate if the performance of the new classification algorithm will increase as well. Performance includes loss of data, execution time etc.
3 Hypothesis

This thesis performs analyses and investigates new cluster based classification methods for EEG data with this hypothesis: The classification accuracy for noisy EEG data will increase if the classes/clusters are separable and uncertain trials are removed.
4 Background

EEG based BCI robot controllers are a popular research field due to its many useful applications and the non-harmful nature of the brain activity acquisition method. Feature extraction and classification are particularly challenging since the EEG data is typically very noisy and the features tend to have high dimensionality, which gives the classification a low accuracy [2, 6]. Having an over-representation of wrongly classified classes or false positives, if a discard state is available, is especially troublesome since it lowers the reliability of the BCI significantly. Special means to eliminate false positives are sometimes preferred even if the true positive rate is lowered as well [7, 8, 9].

4.1 Electroencephalogram

EEG is a noninvasive method that measures electrical activity produced in the brain. The method can be used in a variety of medical and scientific fields ranging from controlling robotic vehicles, detecting epileptic seizures and to analyse driver drowsiness [1, 10, 11].

4.1.1 EEG setup

To measure EEG signals, noninvasive electrodes are typically placed along the scalp of the subject’s head, however invasive electrodes do exist and are used in methods such as electrocorticography (ECoG). For noninvasive electrodes, a variety of gels and sanitation methods need to be applied to get a good signal acquisition. If the applying speed is more important than the signal clearness, specific caps can be used that has the electrodes already correctly placed. The amount of electrodes applied varies depending on the area of use and the spatial resolution needed for the procedure. With a higher amount of electrodes the spatial resolution increases. For example, beamforming analysis (a brain source reconstruction spatial filter) requires high spatial resolution, thus performs better with a higher amount of electrodes, up to 256 [12]. For an experiment similar to this thesis’, 62 noninvasive electrodes are sufficient.

4.1.2 EEG signals

The data collected by the EEG needs to be correctly interpreted by the BCI. Suitable EEG signals for a BCI are event-related desynchronisation/synchronisation (ERD/ERS), steady state visual evoked potentials (SSVEP) and P300 component of event related potentials [13]. SSVEP and P300 both need a specific monitor for the subject to look at. Depending on the setup, the monitors have different command sections on the screen that the subject needs to focus on in order to generate a specific stimuli for the brain to create the signal. ERD/ERS on the other hand is event-related, which means that the signal correlates with an event in the brain, for example motor imagery (MI). ERD/ERS measures rhythmic electrical activity to find and classify its changes. The synchronisation and desynchronisation originates from the idea that thousands of neurons in the brain synchronise or desynchronise their activation, thus creating electrical rhythmic activities that can be measured by the EEG sensors [2]. MI commands are particularly useful when controlling a robot via BCI since no specific stimuli is required to create the signal. This gives freedom in how the commands are generated to control the robot and the subject can even see the robot in action when performing the MI commands since there is no monitor required, which gives visual feedback.

4.1.3 Pre-processing

The signal acquired from the EEG is very noisy due to low signal-to-noise ratio and low spatial resolution as well as various outer sources such as artefacts and interfering frequencies. The low signal-to-noise ratio comes from the fact that the EEG electrodes are applied on the surface of the scalp. Comparing EEG with ECoG, where the electrodes are placed directly on the brain, the signal-to-noise ratio is much higher. This noise needs to be processed in order to extract the important information from the signal. The frequency range usually sought after when performing EEG generated MI are between 8-30 Hz when ERD/ERS is used [14, 15, 16]. Unwanted signals
Figure 1: A simplified flowchart showing the different steps performed in a BCI. In an online scenario the subject is getting feedback in real-time. In an offline scenario the feedback can be given after the commands are performed.

with higher frequency are typically generated by electromyography (EMG) from the muscles in the subject’s body [17]. Standard power lines can create interference as well in the frequency range of 50-60 Hz. Interference in frequencies below 8 Hz are typically created by electrooculography (EOG) activity but also working memory activation generated from the brain itself [17]. A bandpass filter can be utilised to focus the attention on the desired frequencies.

Artefacts can occur from various sources and need to be reduced for a cleaner signal. EMG and EOG artefacts gets mostly eliminated by the bandpass filter but other sources like eye-blinking or similar facial movements are more prominent even with the bandpass filter active. These has to be taken care of by other means. A widely used method is Independent Component Analysis (ICA) which divides the signal into its statistically independent components [13]. The artefacts is then visually selected and removed.

The ICA algorithm for artefact removal used by Bäckström and Tidare will not work in real time and must be performed in a different fashion [2]. One promising approach is used by Matsusaki et al. which expands the ICA-based algorithm for online usage [18].

4.2 Brain Computer Interface

A BCI is an interface between a brain and a computer that translates brainwaves to actual actions that can be performed by other software or hardware. It can also present feedback to the subject in forms of visual and physical stimulus. Bäckström and Tidare used MI as mental commands with the ERD/ERS signals acquired from the EEG. These MI commands represented the different directions of a moving vehicle [2]. To distinguish between these commands the subject was imagining the movement of the right arm, left arm, the feet and the tongue. A simplified flowchart of a BCI can be seen in Figure 1.

4.2.1 Classification

In the work of Bäckström and Tidare, the four-class Support Vector Machine (SVM) classification algorithm for extracting the mental commands from the EEG data generated an over-representation of wrong classifications. Even when they used a rest state for a five-class SVM, the false positive rate was too high and obstructed the movement of the vehicle and was suggested to be the main cause that prevented the vehicle to move along the simulated path.

Support vector machine is a supervised learning algorithm that, with a set of labelled training data, builds a model that can separate two or more categories or classes. This separation can be done in different ways. A hyper-plane is one of the more simple ones that linearly separates low dimensional classes . For higher dimensions or non-linear classification problems a kernel method can be used instead of a hyper-plane [19]. This maps the inputs in to a high-dimensional feature space. The kernel can vary from linear to Gaussian. There exists SVM methods for unsupervised learning as well, for example support vector clustering proposed by Ben-Hur et al. [20].
According to Lotte et al., SVM should be a sufficient classification algorithm for EEG based BCI [21]. Many of the papers referred to in their analysis performed well with SVM which implies that it should be possible to increase the classification accuracy on Bäckström and Tidare data.

According to many sources, a Gaussian classifier could be more suitable for a high dimensional EEG based robot controller [2, 9, 22]. Millán et al. uses Gaussian classifiers for their EEG-based robot controller for classifying EEG-data and is a promising technique to get reliable classification accuracy for a BCI. This was specifically suggested in Bäckström and Tidare report as well. Additional to the Gaussian classification algorithm Millán et al. utilises a threshold method that functions as a filter for unrecognisable MI tasks/trials. This was done to minimise false positives in the classification.

Yavuz and Aydemir used k-Nearest Neighbour (k-NN) and Linear Discriminant Analysis (LDA) as classification algorithms for their EEG based BCI and presented promising results with up to 82% classification accuracy [23]. Mean Derivative (MD) and Hilbert Transformation (HT) were used to extract the features to represent the signals. These are other classification algorithms that might be interesting to utilise on the data by Bäckström and Tidare.

Sreeja et al. compared the LDA and SVM classification methods with a Gaussian Naïve Bayes (GNB) classifier for their MI-based BCI [22]. They proposed that the GNB method would improve the classification accuracy over the other two. This further implies that a Gaussian method would be a suitable classifier for Bäckström and Tidares EEG data. They also stated that the method can be further implemented for a real-time MI-based BCI system.

An new cluster separability analysis method proposed by Krauss et al. suggests that a cluster based classification method could work on EEG data if the clusters are separable enough [3]. The cluster analysis uses the intra and inter cluster distances to calculate a discrimination value that represents the separability of two clusters. The method has not yet been extended as a classification algorithm and therefore has been applied on the EEG data for this thesis and with good results a new cluster based supervised SVM learning algorithm has been implemented.

4.2.2 Robot control

The classification algorithm plays a big part in how reliable the BCI-based robot control is. If the classification error is too high the robot will be unreliable and might cause devastating damage. The signals generated from the brain are individual to the subject performing the MI tasks. Therefore, several signals from different subjects cannot be collectively assembled to get quality out of quantity. When performing the mental commands in the training stage of the classification, the signals must be subject specific and it can take several attempts for the subject before the results are satisfactory. Millán et al. had two subjects participating for their Gaussian classification algorithm [7]. In the beginning, before the subjects were familiarised with the mental commands in question, the classification accuracy was low, i.e. false positive rates were high. But after several days of familiarising themselves with the commands the accuracy went up, although in a different pace and fashion. The first subject, who had more experience with MI commands, had a linear descent of false positives. The second subject had the false positive rate vary form day to day but with an overall descending trace. This shows that the performance of the robot controller are dependent on the subject and can vary from day to day.

When there exists a rest state or a discard state where uncertain trials are disposed, false positive rate determine the performance of the BCI, as stated by Liu et al. [24]. False positives are trials that should have been in the rest state but are instead classified as active classes, which is undesirable. Liu et al. developed a SVM-based binary classifier that uses false positive rate control schemes to force the false positive rate in to a desirable amount. The effect is subject independent which can increase the robot controller performance for a subject with minimal MI command experience.

The data provided by Bäckström and Tidare did not initially have a rest state or discard state [2]. This forces every trial to be classified to active classes which can decrease the classification accuracy if the MI commands are difficult to distinguish, hence making the BCI unreliable. Bäckström and Tidare did implement a rest state in pursuit of increasing the classification accuracy by classifying some of the trials in to that state. The difference in results with and without the rest state was minimal however and the rest state will not be used in this thesis.
4.2.3 Visual feedback

Classification of EEG data can be done both offline and online. Offline means that the training data and prediction data is prerecorded. For an MI driven BCI, this means that the subject performing the tasks gets no feedback of how the classification of the MI commands worked until after the whole session. In an online scenario, the training data is typically performed beforehand which allows the classification to learn how to recognise the MI commands. However, the prediction data is recorded and predicted at the same time which creates opportunities for the subject to adapt depending on the feedback of the BCI. Visual feedback while operating a BCI has been shown to promote higher performance resulting from neural learning and adaptation [4, 5]. When doing MI commands, a visual representation of the commands is often visualised by a moving cursor or a simplified robot simulation. Bäckström and Tidare used Gazebo, which is a Linux-based simulation tool that is specialised in simulating different models of robots and environments [2]. Gazebo is a toolbox that comes with ROS (Robot Operating System) which is an operating system dedicated for controlling and simulating robots. Bäckström and Tidare used this tool for visual validation in an offline scenario. To fully justify a BCI based robot controller it must be implemented in an online scenario which provides visual input of the moving vehicle whilst performing the mental tasks.
5 Expected Outcome

The main expected outcome of this thesis is a classification algorithm that aims to find uncertain trials and discard them in order to increase the quality of the data. The aim is to increase the classification accuracy of the data recorded by Bäckström and Tidare. Alongside, this thesis aims to obtain further knowledge related to 1) the separability of the classes with the use of cluster analysis methods 2) the effect of artefact removal on the data and classifier, specifically blink and eye movement artefacts.
6 Method

This thesis uses an exploratory, iterative and quantitative engineering methodology. The classification experiments was done in an offline scenario but with online capabilities in mind. The offline data is prerecorded by Bäckström and Tidare and permission for usage was granted. All participants for the recording of the data are anonymous and will remain anonymous. The thesis is divided in to three stages: pre-processing, data analysis, and classification.

6.1 Pre-processing stage

To be able to perform classification algorithms on the EEG data, it had to undergo pre-processing stages such as filtering and artefact removal. This was mainly done by Bäckström and Tidare in their previous thesis and documentation from their work was handed over to this thesis.

6.1.1 Signal acquisition

In this thesis, data collected by Bäckström and Tidare were used and the same signal processing technique applied on their data was utilised. This provides a good opportunity for comparison when evaluating different classification techniques. The data provided by Bäckström and Tidare was recorded using an EEG system with 64 and 65 Ag/AgCl active electrodes, however only 62 of the channels are of importance for the EEG data. One high-pass FIR filter of order 3000 with a cutoff frequency of 2 Hz was applied on their data together with a low pass FIR filter of order 100 with a cutoff frequency of 40 Hz. The same filtering was done in this thesis. The widely used ICA algorithm was used as the eye movement artefact reduction method in the same fashion as Bäckström and Tidare. The classification algorithms were tested both with and without the ICA algorithm in order to investigate its impact on the accuracy.

6.1.2 Data

Two occasions of data recordings are available from the work of Bäckström and Tidare. Each occasion contains five recording sessions of data. In a recording session, the subject performed the four MI commands successively in a random fashion with short breaks in between. The subject sat still in a dimly lit room with no other disturbances. The first recording occasion contains three sessions from their first subject and two from their second. The second recording contains five sessions from the same subject and these recordings were mostly used in this thesis. A BCI is individual to the user’s own brain and when the data goes through the training phase of the classification it is of most importance that the data is from the same subject that will use the BCI. The EEG data from the 62 channels are divided into epochs for each session. Each epoch contains data temporally aligned to the onset of the MI task in a time-window of -1000 to 4000 ms. The MI tasks in the epochs were performed semi-randomly, that is that the order was fairly random, however the distribution of MI tasks were evenly distributed throughout the epochs.

6.1.3 Feature extraction

Power Spectral Density (PSD) was used as features when the machine learning algorithm was applied. PSD is the distribution of power a signal has when measured by frequency. When averaging the PSD, the patterns of the signal can be visualised as ERD/ERS and the tasks can be recognised [2]. The PSD was calculated with Morlet wavelets. Just as in Bäckström and Tidares work 20 complex Morlet wavelet was calculated for all the 62 channels of the EEG. The wavelets are logarithmically spaced from 4 Hz to 30 Hz covering 20 different frequencies. This yields 1240 power features for each epoch of all the sessions. As already tested in Bäckström and Tidares report, a feature reduction algorithm will make little difference for the classification on the data and therefore was excluded in this thesis.
6.2 Data analysis

There were two main analyses that has been performed on the EEG data before the major classification algorithms were implemented. The analyses helped determine the design of the classification algorithms.

6.2.1 Eye artefact removal

The ICA method was applied on the EEG data and an analysis on how major the artefact reduction algorithm changed the data was performed. The amplitude of the blink artefacts was noted before and after the ICA algorithm was applied and a statistical analysis determined the impact it had on the data.

6.2.2 Cluster analysis

The cluster analysis proposed by Krauss et al. was done on the data to investigate the separability of the four MI tasks, or in terms with more pertinence, the four clusters [3]. The method is described for two clusters by Krauss et al., therefore a more developed version of the algorithm was used and is described more thoroughly in section 8.3. The basic concept is that the intra and inter-cluster distance are calculated and used as variables to form a discrimination value that determines the separability of the clusters. Since there are more than two clusters this method is done in two ways, one where each cluster are binary paired with all other clusters, which gives several discrimination values. The other method has the cluster distances calculated between the clusters all at once, which gives one discrimination value.

6.3 Classification

The classification of the EEG data that Bäckström and Tidare recorded was expanded to minimise classification error or false positives. Depending if there exists a rest state or a discard state for trials that are uncertain, false positives can be targeted to create a higher accuracy. A standard SVM was implemented at first in a similar fashion as Bäckström and Tidare did. Later a cluster-based version was implemented and compared with the results from the standard SVM. All the tests were done with five sessions of data, each contains around 103 trials recorded by Bäckström and Tidare if nothing else is specified. The results are presented by confusion matrices which have the predicted classes on the x-axis and the true classes on the y-axis. The values in the matrix represents how much percentage of one class that was classified as a specific class, i.e. the accuracy of that class.

6.3.1 Support Vector Machine

An SVM with a linear kernel was implemented for classification by using fitcsvm in Matlab on the data provided by Bäckström and Tidare. PSD was used as features and the training were performed on all the 103 epochs for each of the five sessions from the second recording occasion. 5-fold cross-validation was performed to minimise the risk of over-fitting the training exercise. In the report of Bäckström and Tidare it is described that they tried different window sizes with different starting points for the PSD to represent the MI task most accurately. Centre of window and window size is not clear from the report however. After close investigation of their report, a window size of 800 ms centred on 1000 ms after the onset of MI was used in this master thesis.

6.3.2 Cluster-based SVM

The results from the cluster analysis method proposed by Krauss et al. determined a new cluster based SVM algorithm. The theory is that the MI tasks can be separable as clusters, thus creating an opportunity to divide the classification between these clusters. That is, implementing SVM on each cluster to learn how to recognise the specific clusters and determine if a trial/point belongs to it or not. These SVM results can then be compared and if two points were classified as more than one cluster it was discarded. This will hopefully minimise false positives by discarding uncertain trials/points. A more detailed explanation of the algorithm is presented in section 8.4.1.
7 Limitations

The offline EEG data handed over from the previous master thesis can be both a limitation and an advantage. The quick delivery of already recorded EEG data saves a lot of time but limits the experiments to data recorded by someone else. The classification learning and testing were done in an offline scenario due to already recorded data and limited time. This compromises the actual usability of the algorithm when an eventual online application is implemented. Although, when the algorithms chosen for this thesis were implemented and designed, online capabilities were in mind.
8 Design

The first goal when developing the BCI controller was to reconfirm the SVM classification results of Bäckström and Tidare. That is because the second goal was to implement a complementary threshold-based classification that utilizes both SVM and the cluster analysis proposed by Krauss et al. [3]. This algorithm has potentially decrease the classification error that the SVM generated.

8.1 Eye-blink component removal

The well tested algorithm ICA was used on the raw EEG data to remove artefacts, specifically eye-blink components. The independent components were visualised separately and discarded manually [25]. 62 different components were generated per session and only one or two for every session matched with the eye-blink component shown in Figure 2. They are named from IC1 to IC62 as can be seen in Figure 2 and 3 which are the components IC4 and IC14. The component that corresponded to the blink was removed manually for every sessions of data.

A typical blink artefact component that is visually represented when running the ICA algorithm can be seen in Figure 2 in relation to another non-blink component in Figure 3. In the head-shaped colour graph (topoplot) generated by IC4 it can be seen that the activity is heavily centred on the frontal part of the head with gradually decreasing activity when going from frontal to posterior brain regions which could imply eye movement. Also, in the Continuous data colour graph for the same component, heavy sporadic and transient spikes occur throughout the time span which also implies eye movement, specifically blinking. When looking at the Continuous data colour graph for component IC14, the power is randomly spread out in the time span which most probably represents other brain activity.

8.2 Feature extraction

The features used in Bäckström and Tidare’s work was Power Spectral Density (PSD), since it correlates with the ERD/ERS. The dimensionality of the data increases rapidly with a high amount of samples which can limit the usefulness of various classification algorithms. Although, according to Bäckström and Tidare a feature reduction method does little difference for an SVM classification and therefore was not used in this thesis either [2].

8.3 Cluster separability analysis

The cluster separability analysis done in this thesis is based on the multidimensional cluster statistics (MCS) method proposed by Krauss et al. [3]. In their report they present the MCS by explaining the steps required to investigate the separability between two clusters, although they mention that it can be applied on multi-class problems as well. The basic concept is to find a discrimination value that represents the separability between two clusters or classes. The lower the discrimination value is the more separable the clusters are. In short, for a two class problem, the algorithm calculates the euclidean distances between all trials for both classes. That is equivalent to the intra- and inter-cluster distances between the two classes. This can be represented as a distance matrix which can be seen in Figure 4. For two classes the distance matrix is divided in to four quadrants or boxes where the top left and the bottom right box represents the intra-cluster distances and the other two the inter-cluster distances. These boxes are used to calculate the discrimination value for the two clusters. The discrimination value is calculated by adding the mean of the intra-cluster distances and subtract the mean of the inter-cluster distances. The formula can be represented as follows:

\[ d(A, B) = \text{mean}(A, A) + \text{mean}(B, B) - \text{mean}(A, B) - \text{mean}(B, A) \] (1)

where \( d(A,B) \) is the discrimination value between the clusters A and B and \( \text{mean}(\cdot) \) is the mean value for a specific box from Figure 4. For example \( \text{mean}(A,A) \) is the mean of the top left box.

The discrimination value calculated from the two classes does not convey that much information by itself. A statistical comparison to evaluate if the discrimination value represents two clusters that are statistically separated must be performed. Krauss et al. took the original clusters and
relabeled them randomly, thus creating a new set of clusters and calculated the discrimination value for the new set. These clusters will, if the original clusters contain statistical separability, have higher discrimination values than the originals because the new clusters are mixed together at random. This procedure of randomly mixing labels and calculating discrimination value was repeated 1000 times to generate a non-biased discrimination value distribution. If the original discrimination value is within the top five percentages of all the randomly generated cluster pair discrimination values the original cluster pair can be viewed as separable according to the statistical confidence interval concept (one-tailed non-parametric random permutation, $p < 0.05$).

In this thesis there are four classes, or clusters if viewed as that. The above described cluster analysis was faced in two different ways that are applicable for this thesis. The first one is a one-vs-one method where each cluster is paired with one other cluster to calculate the discrimination value. This was done for every possible combination pair, that is all clusters were paired with all clusters. Since there are always only two clusters paired with each other, formula (1) could be used. The other way this analysis was faced was a one-vs-all method. When calculating the inter-cluster distances in this method the clusters was paired with all the other clusters at once. Since all the clusters are used for one calculation the distance matrix becomes a four times four matrix with 16 boxes as seen in Figure 5. This gave another discrimination value formula based on the distance matrix:

$$d(A, B, C, D) = \text{mean}(A, A) + \text{mean}(B, B) + \text{mean}(C, C) + \text{mean}(D, D) - \text{mean}(A, (B : C : D))$$
$$- \text{mean}(B, (A : C : D)) - \text{mean}(C, (A : B : D)) - \text{mean}(D, (A : B : C))$$  \hspace{1cm} (2)$$

Where $d(A,B,C,D)$ is the discrimination value for all the clusters combined and $\text{mean}(A,(B:C:D))$ is the mean of the distance boxes on the top row in Figure 5 except for $(A,A)$.

When the discrimination values for both the one-vs-one and one-vs-all methods was calculated, 1000 randomly labelled clusters with corresponding discrimination values for both the methods were generated. The results of the cluster analysis will be further examined in 9.2.
Figure 4: A and B denotes the boxes that are used to calculate the discrimination values. The x-axis and y-axis are the amount of features. This particular distance matrix is generated from the distance between right arm and left arm.

Figure 5: A, B, C, and D denotes the boxes that are used to calculate the discrimination values. The x-axis and y-axis are the amount of features. This distance matrix is generated from all four classes, right arm, left arm, tongue, and feet.
8.4 Classification

Bäckström and Tidare got their best result when they used SVM and according to Lotte et al. it is one of the best classification algorithms to use on EEG data specifically [2, 21]. Never-the-less the results of Bäckström and Tidare were unsatisfactory and another method could increase the accuracy as well as reducing the false positive rate. By using the cluster analysis, presented in the previous section, a new method that combines SVM and cluster separability is proposed. The goal is to investigate if the classification error can be reduced by modifying the classification technique.

When the algorithms were tested, exactly the same features of the data was used. The same window size was also used for all the algorithms. This ensures that it is only the classification algorithm that influences the results. Some of the algorithms might perform better with other features or window sizes, but this is outside of the scope of this master thesis.

8.4.1 Cluster-based SVM

To utilise the separability of clusters for an increase in classification accuracy, a new method of classification will here be presented. It still uses the same SVM as before but input data is driven by cluster separability in an attempt to discard uncertain data (i.e. data that are far from its cluster). The four classes: right arm, left arm, tongue, and feet are divided into 4 clusters represented in a multidimensional feature space. The same data as the previous SVM experiment are used to more accurately compare the two results. The first step is to calculate the distance of all points in one cluster to its own centre. The point that are furthest apart from the cluster centre is considered as a threshold for that cluster. The second step is to calculate the cluster distances from the other three clusters to the first cluster’s centre. All points from the other clusters that are outside of the threshold, i.e further apart from the cluster centre than the cluster centres own points, are tossed aside and not used as training data for that specific cluster or class. This is done for all four classes which creates four different sets of training data.

Each new training data are labelled so that one trial either belongs to the current cluster (labelled 'in') or one of the other clusters (labelled 'out'). In other words the four new training sets are binary. Now the SVM is implemented on each new training set to train the classifiers to distinguish its own class from the others. This creates four classifiers that can only find if a new trial belongs to its class or not. After the training is complete, each classifier has judged if the trial belongs to any of the clusters. If two or more classifiers say that the same trial belongs to its class then it is labelled as unknown and is tossed aside. If only one of the four classifiers recognise a new trial as its own class it is labelled as classified. This will in theory further minimise wrongly classified trials.
9 Results

In this section, the results of the described methods from the previous section is presented. They are presented in order of implementation with the analysis first and the classification algorithms in the end.

9.1 Eye artefact removal

The ICA algorithm was implemented on the data provided by Bäckström and Tidare in a similar fashion as they did in their work. A more detailed description on the method can be read in section 8.1. In Figure 6 it can be seen, for one example epoch, that the amplitude of the EEG data in the moment of a blink artefact has been neutralised. The upper subplot shows the EEG signal of one channel located near the eyes and the blue graph in the subplot represents the signal with the artefact reduction algorithm applied. The red graph in the same subplot represents the same signal without the artefact reduction. As for the lower subplot in Figure 6, it represents the eye movements in the horizontal axis (red) and the vertical (blue) axis of the same example epoch. A blink artefact generates a clear amplitude change in the vertical axis. The bar-graph in Figure 7 shows the average amplitude reduction across blink artefacts in all sessions. A significant reduction of amplitude change created by blink can be observed ($p < 0.001$, Wilcoxon signed rank test). As can be seen from the p-value representation in the graph (***) the reduction is a significant change to the data.

9.2 Cluster analysis

To investigate how separable the four classes are as clusters in a multidimensional space, the MCS method described in 8.3 was implemented. Significantly separated clusters in the multidimensional feature-space suggest that classification of the MI tasks should be possible. In short, a discrimination value between clusters are calculated and the lower the value is the more separable the clusters are. To make the analysis as thorough as possible, two experiments were performed where the results yielded was calculated both with one-vs-one and one-vs-all. That is the binary cluster distance and the distance calculated for all four clusters. The discrimination values for the one-vs-one session are presented in Figure 8 as a table where all the pairs of clusters possible are shown. The discrimination value between right arm and left arm are the lowest, around -9.7. This suggests that those two clusters are the most separable. Right arm and tongue has the highest value, which suggests that those two clusters are least separable.

These values with no relation to anything else does not say that much. Therefore 1000 randomly labelled clusters were constructed and the corresponding discrimination values were calculated. The original discrimination values were then compared with the values from the randomly generated cluster distances to see how separable they are in relation to randomly labelled clusters. Figure 9 presents the p-values that represent the probability that the clusters are non-separate. The p-value is gathered by comparing the original discrimination value with discrimination values from the 1000 permutations. When the permutations are sorted, like in Figure 10, the area where the original discrimination would be placed are calculated and the percentage for that area compared to the whole scale becomes the p-value. If the p-value is below 0.05 it is considered to represent separate clusters (i.e. the probability that the clusters originates from the same cluster is below 5%). These p-values correlates directly to the discrimination values in Figure 8. As can be seen, right arm and left arm have the lowest p-value suggesting that those MI tasks are most separable. Some MI tasks does not have a p-value below 0.05 like right arm and tongue. This suggests that those clusters are not separable enough as clusters.

To see if the four clusters are overall separable, the one-vs-all method was implemented. This was done by calculating the distances between one cluster and all the others. This was done four times, one for every cluster and gave the discrimination value -11.62 in accordance with Formula 2. By performing the same random permutations as described above, the discrimination value is observed as highly significant suggesting that the clusters are statistically separate in the multi-dimensional feature-space ($p = 0.002$, 1-tailed non-parametric random permutation test). In Figure 10, the 1000 random permutation discrimination values can be seen in a sorted graph.
Figure 6: The upper graph represents one epoch of one session of the EEG data, the red is without the ICA algorithm applied and the blue is with the ICA algorithm. The lower graph represents the EOG, eye movement, for the same epoch. The red is vertical eye movement and the blue is horizontal eye movement.

Figure 7: This bar-graph shows the amplitude reduction the ICA algorithm does on the EEG data. All the blink components for all five sessions are used in this graph to represent the difference in amplitude when using, and not using, ICA on the data.
Figure 8: This matrix represents the discrimination values calculated between all clusters separately. As can be seen, the discrimination value is zero when doing the calculation between the same cluster.

Figure 9: This matrix represents the p-values obtained when comparing the original discrimination values with the ones from the 1000 randomly labelled new clusters. For example the box denoted by left arm and right arm has the p-value 0.002 which indicates that the original discrimination value is within the top 0.2% of the randomly labelled clusters.
Figure 10: The discrimination values from all the 1000 randomly labelled clusters. They are sorted from the lowest value to the highest. As can be seen, the original value of -11.62 would be at the lowest most part of the graph, hence giving it a low p-value.

9.3 Classification

The classification of the MI tasks from the EEG data was the main objective in this thesis. The goal was to find a way to increase classification accuracy or more specifically reduce the amount of false positives or wrongly classified trials. At first an SVM algorithm was used similarly as Bäckström and Tidare. Later the new Cluster-based SVM was used on the same data to compare with the results from the SVM.

9.3.1 Support Vector Machine

When applying SVM on the five sessions of data provided by Bäckström and Tidare it yielded an overall classification accuracy of 40.4%. The results of the classification are represented in the confusion matrix that can be seen in Figure 11. The accuracy of every class ranges between 30.8% to 46.5% with right hand having the highest accuracy and feet having the lowest accuracy. Overall, the accuracy of the four different classes are relatively evenly matched. There is no class that is classified more often as the wrong class rather than the correct one. Even so, the accuracy of the classes are too low.

To investigate if the ICA algorithm for removing blink and eye movement affected the data in a major way, which was suggested by Figure 7, the SVM was also tested on data with no ICA performed. All of the other steps is exactly the same as before, with the same filters and feature extraction methods. The overall classification accuracy when no ICA was applied was 40.6%. Therefore since there was no major difference with or without ICA, there will be no further effort to find an artefact removal procedure that can be applied in an online scenario.

These results are quite similar to Bäckström and Tidares results when compared to their SVM classification of the data. In the results part of their report they presents an overall classification accuracy of 41%. Although they used a separate fifth class which was a relax state which makes a comparison questionable. This relax state was excluded in this master-thesis for simplification and since the accuracy did not change in a major way when the state was present.
Figure 11: The confusion matrix for the SVM classification of the four motor imagery. The percentage is the amount of trials classified as that class. Every square has two values. The upper one is accuracy for the class relative to its own class and the lower one is accuracy relative to all classes. The classifier yielded an overall accuracy of 40.4%. An over-representation of wrongly classified classes is generated which translates to the relatively low accuracy.

### 9.3.2 Cluster-based SVM

To reduce wrongly classified trials, a threshold based algorithm was implemented that utilises the cluster distances between the four MI classes to filter out uncertain classifications. The algorithms implemented are described in depth in 8.4.1. This algorithm was designed to reduce false positives by discarding trials that are uncertain and that would have been miss-classified without the threshold. In this analysis, the SVM was trained with 70% of the data and the last 30% was used in prediction. A confusion matrix of the result can be visualised in Figure 12. The overall classification accuracy is 48% and it is an increase of about 8 percentage points as compared to using the SVM alone. Around 45% of the data was labelled as uncertain and was not classified at all. Unlike the results form only using SVM the spread between the accuracy of the four classes are much bigger this time. The accuracy of right arm is 75% which is significantly better than with SVM alone. Although, the accuracy for feet are as low as 19% and it is more often classified as both right and and left arm which is significantly worse than SVM. For left arm and tongue the accuracy are similar to using SVM alone, although a bit higher with this new cluster-based algorithm.
Figure 12: The confusion matrix for the Cluster-based SVM classification of the four motor imagery. The percentage is the amount of trials classified as that class. Every square has two values. The upper one is accuracy for the class relative to its own class and the lower one is accuracy relative to all classes. The removed section shows how the removed trials were distributed percentage-wise amongst each class and the total is the amount of removed trials relative to all data. The classification yielded an overall classification accuracy of 48.2%.

10 Discussion

This thesis has overall achieved what was stated in the problem formulation. An analysis on how eye artefact reduction algorithms affect prerecorded EEG data has been done. An algorithm has been successfully implemented that reduces the false positives and thus increased the accuracy on the data provided by Bäckström and Tidare. To get this increase in accuracy, some data had to be discarded based on the multidimensional distance from its cluster-centre, which in turn reduced the classification performance of the algorithm.

10.1 Eye artefact removal

The results from the artefact reduction analysis is interesting when the comparison between classification accuracy with and without ICA are realised. According to the artefact reduction analysis, the ICA algorithm reduces a major part of the amplitude of the EEG signal and in turn changes the signal significantly. Yet, the classification accuracy without ICA is a small percentage higher than with ICA, 40.6% to 40.4%. This is most likely due to the fact that the blink artefact comes fairly rarely and relatively periodical. This suggests that the artefacts do not affect the classification for the majority of the signals span due to its rareness, and when the artefacts does an appearance in the signal it can be interpreted as a feature for the classification algorithm to recognise thanks to its periodicity. This can contribute to the small accuracy increase when no ICA was used. This agrees with the artefact reduction analysis performed as well, since it takes in to account only the amplitude and quantity of the artefacts, but not in relation to the length of the signal.

To get the most fair classification it is of course better to reduce the artefacts. Although, the specific algorithm used in this thesis, the Independent Component Analysis, does not work in an online scenario since the complete signal is required for the implementation. There are algorithms that take care of artefacts, like blink and eye movement, in an online scenario like the online ICA developed by Matsusaki et al. [18]. The algorithm still adds a short delay of around three seconds, which could be devastating depending on the use of the EEG signal. In a BCI where the user will control a robot using the MI tasks produced by the brain, it is problematic if a three second delay is added every time a new command is sent. This brings up the question: is it worth implementing
an eye artefact removal algorithm in an online scenario? In this case it seems like the artefact removal algorithm is unnecessary since it does not change the classification algorithm enough to consider sacrifice a real online experience with no delay. Millán et al. for example did not use eye artefact removal algorithms at all when performing their Gaussian classification algorithm in real-time, stating that it is not necessary [7, 9].

10.2 Cluster separability analysis

To investigate whether the brain activity patterns during mental imagery of different movements (moving the right arm, left arm, tongue and feet) are different statistically, the cluster separability analysis was implemented. The results of the analysis was satisfactory in the sense that most of the cluster pairs were separable. In Figure 9, this is represented by p-values and the cluster pairs that has a p-value under 0.05 are considered separable. Only the combinations of the clusters tongue-feet and tongue-right arm are above the defined threshold of 0.05, which could imply that they are not separate clusters. When performing the cluster separability test on all the clusters at once, the p-value was very low, around 0.002, which in theory says that the four clusters are separate. It would be interesting to try and test different body parts, or different MI tasks in general, to find the optimal combinations for separability. This could in turn lead to a better classification accuracy as well. The cluster analysis is also interesting because it allows to assess the multidimensional distance from each trial/data point to the cluster-centre thus providing a measure that could possibly be correlated with uncertainty. In an attempt to investigate this, an SVM was added that is fine-tuned on discriminating between data from one class/MI task against hypothesised uncertain data from other classes/MI tasks.

Krauss et al. mentioned in their report about the cluster separability analysis that it could be expanded in to a classification [3]. They specifically stated that their method was not a classification algorithm as it was in their report. This thesis has taken their advice and successively introduced a cluster based classification algorithm including SVM. It does also take inspiration from the algorithm presented by Millán et al. since it has a similar threshold based method to filter out uncertain trials [9]. This create the possibilities of false positives and false negatives since there is a discard state, thus giving the opportunity to target false positives and try to eliminate them. The algorithm presented in this thesis does have several variables that could be tuned in order to adapt it to the data used. In the case of this thesis, a significantly small amount of time was spent in tempering with these variables due to time constrains of the master thesis. For example, when selecting the thresholds for each cluster, the threshold was set so that all data points of the current cluster was included. Most clusters were more dense closer to the centre. This means that the threshold could be set closer to the centre, and therefore more points of that cluster relative to the other clusters are captured in the threshold, thus also discarding possible uncertain trials of the current MI task/cluster. This could in turn increase the classification accuracy of the SVM that is added later in the process. This is something that should be considered as a future work possibility.

10.3 Classification accuracy and performance

Stated in the research questions: does the accuracy and or overall performance increase with this new classification algorithm? As can be seen in the results, the classification accuracy does increase by approximately 8 percentage points to an overall accuracy of 48%. Comparing that to the results gathered from only using SVM, which yielded around 40%, and the results from Bäckström and Tidare the increase is percentage-wise major [2]. A big reason to this increase is most probably the threshold and the filtering of uncertain trials. The algorithm classifies cluster by cluster and discards any trials that are classified as two or more classes. The probability of a trial like that being classified as the wrong class is higher than a class that only was regarded as one class from the beginning, hence reducing false positives. Despite the enhancement, the maximum classification accuracy is still low and a BCI would benefit substantially if the classification performance was higher. Yavuz and Aydemir got up to 82% accuracy when they used k-NN and LDA [23]. Millán et al. got in their paper from 2000, with a few days of training, an over all accuracy around 90% [7]. Looking at the research question that asks if the overall performance is increased, the
results suggest that it does not. To determine the overall performance of the classification, not only the accuracy needs to be taken into account but also how the algorithm executes. In the case of the new cluster-based SVM introduced in this thesis, around half of the data is discarded because of uncertain recognition of the classes. This is in accordance with the results from Millán et al. as they got a discard rate of around 50-60% [7]. This can in theory make small delays when performing the algorithm in an online scenario if several input trials get discarded in a row, which decreases the classification performance. In practice though, the signals should stream fast enough for the user to not feel a major difference if some of the trials are tossed aside. However, due to visual feedback, studies have shown that feedback allows the user to adapt its brain activity in order to produce the correct behaviour of the robot [5].

When comparing the confusion matrices from both SVM and cluster-based SVM it can be seen that SVM has a more evenly distributed classification accuracy between the classes. The cluster-based SVM however has a really high classification accuracy for right hand, around 75%, but a really low for feet, around 19%. If the confusion matrix from the cluster-based SVM is compared with the p-values from Figure 9 with this new information the question arises if the right hand is most separable and feet are least separable. The right hand does have the two lowest p-values in the cluster separability analysis, although one of them is against feet. The p-value that are the highest are between right hand and tongue, suggesting that those are least separable form each other, which contradicts the classification results. Although, right hand does overall have low p-values and feet has relatively high p-values suggesting that parallels can be drawn. Possible solutions for this is to try and find MI tasks that are optimally separable from right hand since it has the highest classification accuracy. This could possibly increase the classification accuracy of the other classes.
11 Ethics

The work was conducted in accordance to the relevant guidelines for ethical research of Mälardalen University. Ethical issues related to data collection and personal integrity has been dealt with by keeping the subjects anonymous. In accordance to the local Ethical Review Board, no ethical approval was necessary for this thesis.
12 Conclusion

This master thesis has been a continuation of the master thesis Bäckström and Tidare did in 2016 [2]. They collected EEG data and attempted to classify MI tasks in order to control a mobile robotic platform. The focus for this specific thesis is to improve the classification accuracy, specifically reduce false positives, from the SVM classification algorithm that Bäckström and Tidare used. This thesis has also analysed the effect eye movement artefact reduction has on the EEG data as well as the separability of the four MI tasks using the cluster separability test method presented by Krauss et al. [3]. With the information collected by the analyses, a new clusters-based SVM classification algorithm has been implemented and tested on the data.

The result showed that an artefact reduction algorithm, specifically eye movement reduction, is unnecessary for classification accuracy reasons. In an online scenario it might be more profitable to not use it because it is more desirable to avoid time delays than having artefact free EEG data. The separability analysis showed that the MI tasks are separable but can be more distinguishable if other MI tasks are chosen. The cluster-based SVM generated better classification accuracy than only using SVM even if the clusters are not perfectly separable.

12.1 Future work

Dividing the classification into clusters showed promising results in this thesis and is something that could be extended in the future. The first thing that could increase the result significantly is to find MI tasks that are optimal for the cluster separability analysis. This will in turn make the cluster based SVM classification more efficient since the clusters are more recognisable. Another thing is the cluster-based algorithm itself which could be improved upon. As mentioned in section 10 the threshold that embraces the clusters when separating them for later classification could be altered in size for further classification improvement.

An obvious step forward for this project is to test the algorithm in an online scenario. The visual feedback from seeing the robot move when performing the MI tasks can help the subject readjust the brain activity depending on the result. This can in turn increase the accuracy and performance of the classification algorithm. The classification accuracy was improved from Bäckström and Tidare’s results and the robot controller they used could be implemented together with this new classification algorithm for an online application.

Using different features for the data could potentially increase the classification accuracy as well. There are various other methods to calculate power that can be used, like the Welch’s periodogram algorithm used by Millán et al. [9].
13 Acknowledgements

I would like to express my sincerest gratitude to my supervisors Elaine Åstrand and Jonatan Tidare for supporting me with knowledge and encouragement during this thesis. I would also like to thank Mattias Bäckström and Jonatan Tidare for lending me all of their material from their previous master thesis which helped a lot when starting this master thesis. Lastly, I would like to thank Mälardalen University for introducing this exciting master thesis.
References


