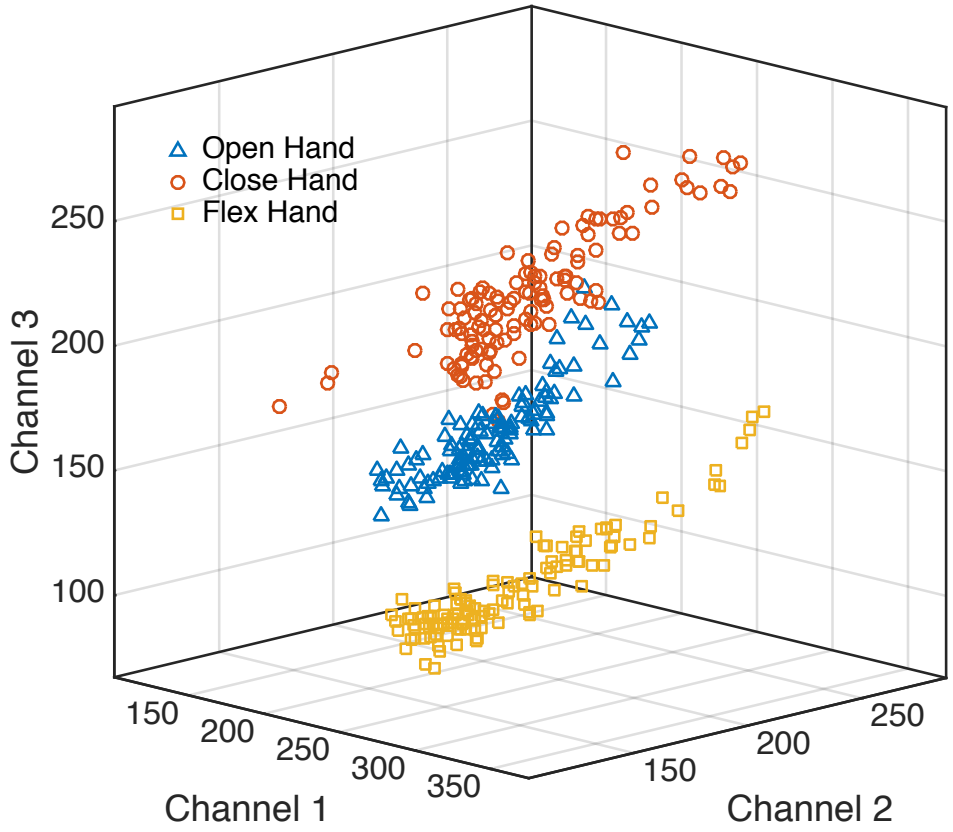
CHALMERS
UNIVERSITY OF TECHNOLOGY



Electromyography Analysis by Classification Complexity Estimation

A Study of the Complexity of Myoelectric Pattern Recognition

Master's Thesis in the Master's programme in Biomedical Engineering

NICLAS NILSSON

Department of Signals and Systems
Division of Signal processing and Biomedical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2016
Master's Thesis EX078/2016

MASTER'S THESIS EX078/2016

Electromyography Analysis by Classification Complexity Estimation

A Study of the Complexity of Myoelectric Pattern Recognition

Master's Thesis in the Master's programme in Biomedical Engineering

NICLAS NILSSON

Department of Signals and Systems
Division of Signal processing and Biomedical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2016

Electromyography Analysis by Classification Complexity Estimation
A Study of the Complexity of Myoelectric Pattern Recognition
Master's Thesis in the Master's programme in Biomedical Engineering
NICLAS NILSSON

Master's Thesis EX078/2016
Department of Signals and Systems
Division of Signal processing and Biomedical Engineering
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: + 46 (0)733805237

Cover:
Clusters of features representing three movements plotted in a three dimensional feature space. The descriptive feature cardinality was calculated for time windows extracted from a recording of three EMG channels. EMG feature space is further explained in section 2.3.

Electromyography Analysis by Classification Complexity Estimation
A Study of the Complexity of Myoelectric Pattern Recognition
Master's Thesis in the Master's programme in Biomedical Engineering

NICLAS NILSSON
Department of Signals and Systems
Division of Signal processing and Biomedical Engineering
Chalmers University of Technology

ABSTRACT

Intuitive control based on myoelectric pattern recognition (MPR) can be used in clinical applications such as prosthetic limbs and Phantom Limb Pain treatment. Electromyography (EMG) patterns representing limb movements are learned by a pattern recognition algorithm to enable classification of future EMG observations. These EMG patterns are commonly constituted by descriptive features extracted from raw EMG. The complexity of the classification task is highly influenced by both the selection of such features and the differentiation between movements in the raw EMG. A reliable estimation of classification complexity would facilitate selection of features and elimination of detrimental EMG patterns. Two such algorithms, Separability Index and Nearest Neighbor Separability, were found to be highly correlated with classification accuracy and enable efficient feature selection for three classifiers commonly used for MPR (Linear Discriminant Analysis, Multi-Layer Perception and Support Vector Machine).

The algorithms were implemented in the data analysis and feature selection modules of BioPatRec, an open source tool developed at Chalmers University of Technology for development and benchmarking of algorithms in MPR. The implementation included dedicated graphical user interfaces to ease visualization. This thesis deepens the understanding of the complexity of MPR and provides tools for prediction of classification performance and analysis of MPR applications.

Key words: Classification Complexity, Myoelectric Pattern Recognition, Electromyography, Feature Selection, Feature Space

Contents

Acronyms

ADC	Analog to Digital Converter
CCE	Classification Complexity Estimate
CCEA	Classification Complexity Estimating Algorithm
EMG	Electromyography
GUI	Graphical User Interface
LDA	Linear Discriminant Analysis
MLP	Multi-Layer Perceptron
MPR	Myoelectric Pattern Recognition
PLP	Phantom Limb Pain
PRA	Pattern Recognition Algorithm
SVM	Support Vector Machine

1 Introduction

Afflictions caused by amputation, such as function loss and Phantom Limb Pain (PLP), instantly decrease the quality of life, and many everyday tasks that use to come naturally become demanding challenges. Function loss is most commonly treated with prosthetic limbs, while PLP treatment is based on medication and rehabilitation [1]. Prostheses has historically been mechanical and passive, but in recent decades electrically controlled and powered prostheses are available [2]. Complex control algorithms using machine learning for motor volition decoding had been extensively studied [3] and shown to be useful for both prosthetic control and in PLP treatment [4]. This thesis aims to study the complexity of motor volition decoding.

1.1 Context and Motivation

Machine learning for motor volition decoding uses Pattern Recognition Algorithms (PRAs), or classifiers, that are fed with data constituted by input signals and corresponding expected outputs. The algorithms learn to recognize patterns in the data and response to the inputs according to the expected outputs.

A classifier can, after a successful learning (or training) process, be used for reliable classification of inputs earlier unseen by the algorithm. Figure 1 shows a simple illustration of the learning process of a PRA.

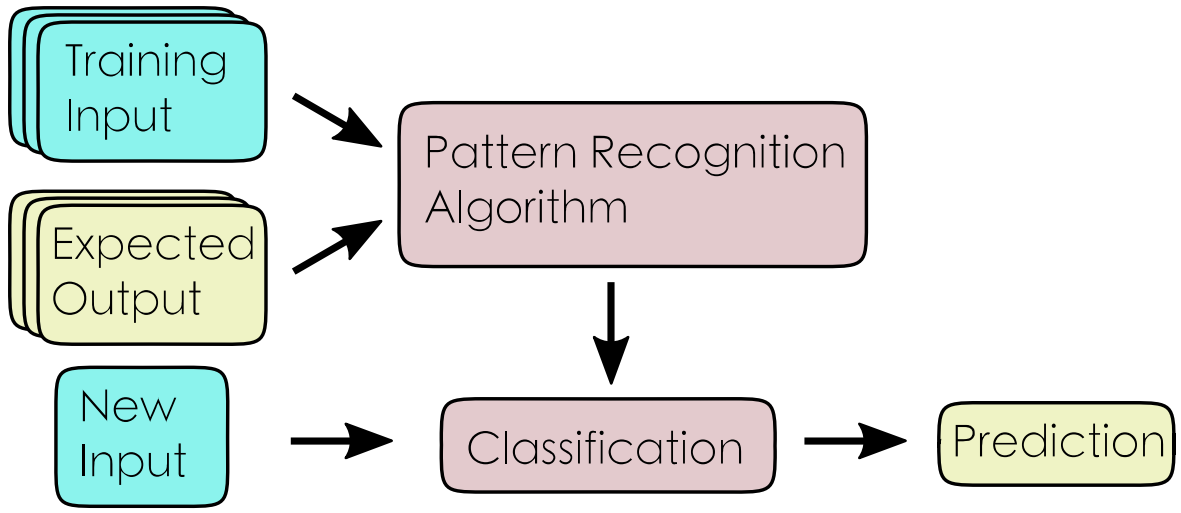


Figure 1: Simple illustration of the learning process of a Pattern Recognition Algorithm

Training data for Myoelectric Pattern Recognition (MPR) are electromyography (EMG) recordings representing limb movements [5]–[7]. The adequacy for classifier learning of such training data is influenced by factors that are irreversible, such as lost muscle tissue, but also factors possible to eliminate, *e.g.* dry skin. Both kinds influence the complexity of the pattern recognition task.

PRAs are rarely designed for raw EMG. Instead they are fed with descriptive features extracted from raw EMG with the aim to reduce redundant information, while maintaining useful information. Such features are differently efficient, and the selection of features highly influence the complexity of the pattern recognition.

MPR has been extensively studied, but few studies has been done on its complexity, *i.e.* the level of influence on the classification performance caused by limiting factors in the application that makes the patterns less separable. A reliable estimation of the classification complexity could contribute to both feature selection and elimination of detrimental influence of EMG.

1.2 Scope and Contribution

This study aimed to find Classification Complexity Estimating Algorithms (CCEAs) that provide information on MPR complexity enabling reliable estimations of classification performance. The CCEAs found to be reliable and computationally sufficient were implemented as part of data analysis and feature selection in BioPatRec, an open source tool developed at Chalmers University of Technology for development and benchmarking of algorithms in MPR.

This thesis work resulted in two papers, one conference paper that was accepted for publication and one journal paper still being processed.

2 Background

This chapter describes machine learning as a control strategy for treatment of amputees. The first section is a short literature review on the history of prosthesis control. The acquisition of EMG and its attributes are described in section 2.2. Section 2.3 explains the extraction of features and the resulting feature space. Finally, in section 2.4 the concept of machine learning and a number of classifiers commonly used for MPR are described in more detail.

2.1 Prosthesis control

The history of prosthetic devices goes back to 2100 B.C. [8] and the technology has gradually developed ever since. However, the use of myoelectric prosthesis control, which is now extensively studied [9], was first documented in the beginning of the 1940's [10]. A simple and commonly used implementation of myoelectric control is the so called on/off control [9]. An EMG amplitude threshold distinguishes between muscle contraction and surrounding noise. An excess of the threshold indicated the start of a contraction and the exceeding EMG amplitude is used for dynamic control. A common setup has two EMG channels recording opposing muscle groups (*e.g.* flexor and extendor) for control of one degree of freedom (*e.g.* open and close hand) as illustrated in Figure 2 inset A. There are controls that additionally use contraction sequences and co-contraction to change between state of the control to enable more degrees of freedom, *e.g.* going from open and close hand to thumb and index finger pinch control. The robustness of the on/off strategy has made it the state of the art in myoelectric prosthesis control today. Even though the on/off control can be considered a simple form of MPR, it is limited compared to more sophisticated MPR algorithms that classifies the input based on patterns built up by all channels combined.

Such algorithms enables seamless transition between the aforementioned states and even supports mixed prediction of simultaneous movement [11]. A simple illustration of myoelectric classification applied to forearm movements is shown in Figure 2 inset B.

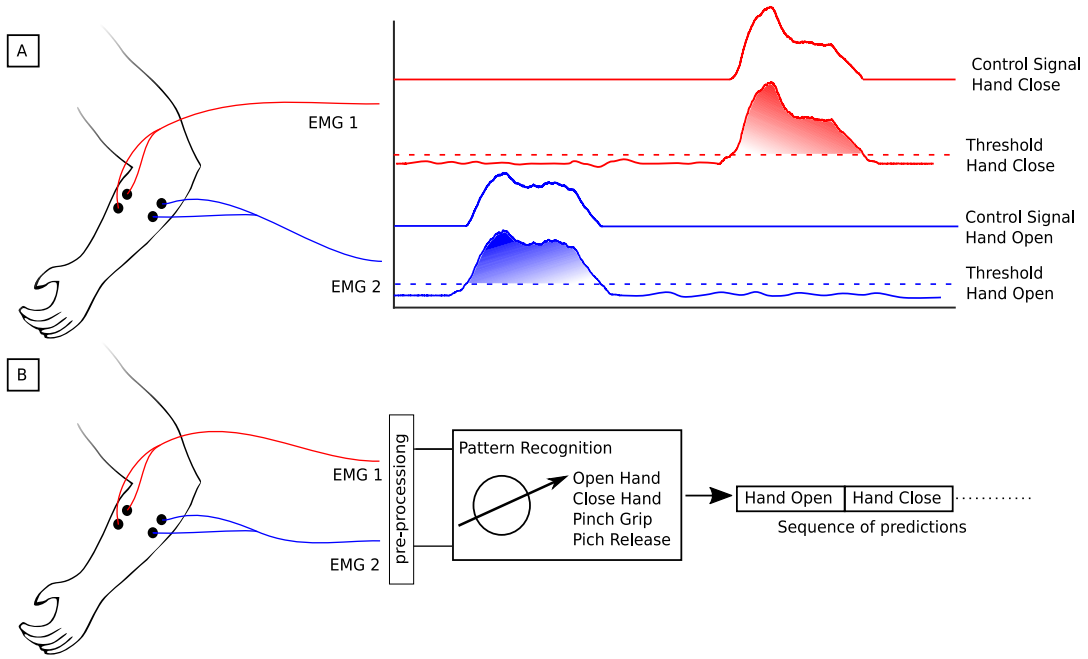


Figure 2: Inset A illustrates a on/off control strategy for prostheses. The doted line is the threshold for start of control and the marked area shows the level of contraction used for dynamic control. Inset B illustrates the structure of a myoelectric classification application.

MPR can be divided into three parts; EMG acquisition, Feature Extraction and Pattern Recognition. The three parts are described in more detail in section 2.2-4. In order to achieve natural control there is need of an additional part deciding how to act on the predicted motor volition. This part is not trivial, but outside the scope of this thesis.

2.2 Electromyography acquisition

EMG is considered to be a key component in natural control of prosthetic devices [3], but is also used within management and rehabilitation of motor disability among other things [12]. EMG measures the biopotential that activates muscle contraction. Figure 3 shows four EMG channels recorded on the forearm using surface electrodes, while the subject is closing and relaxing the hand repeatedly.

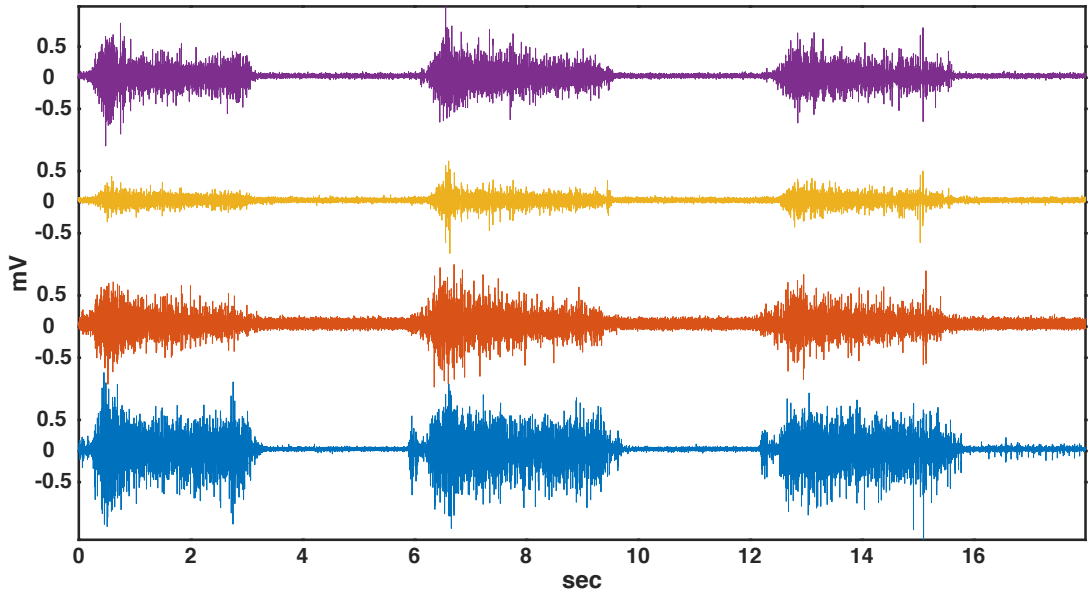


Figure 3: Four EMG channels recorded with surfaces electrodes on a for arm while the subject was closing and relaxing the hand repeatedly.

An acquired EMG signal is always contaminated. Ambient noise from surrounding power lines, motion artifacts from the interface at the skin, and EMG from unrelated muscle tissue close to the target muscle area are examples of unwanted components of the acquired signal. A more complete summary of contaminating factors was given by Raez *et al.* [12].

EMG is acquired through an electrode interface transforming the biopotential from the muscle contraction to electrical signal [13]. Surface electrodes constitute the most commonly used electrodes for MPR because of its non-invasiveness. It creates the interface at the skin and generates EMG with a frequency range between 0-500 Hz and an amplitude of 10 mV (-5 to +5) [14]. Surface electrodes are usually used in a bipolar configuration with 10-20 mm apart and placed along the target muscle fibers to detect its biopotential gradient [15]. An alternative to surface electrodes are implanted electrodes that creates the same interface but on the surface of the muscle tissue or inside the muscle fiber [16]. Implanted electrodes was shown by Ortiz-Catalan et al. to improve MPR compared to surface electrode [17].

A bipolar electrode configuration is naturally used in combination with a differential amplifier. Differential amplifiers are consequently a common amplifier configuration for EMG acquisition [15]. A high-quality amplifier for EMG have adjustable amplification between 100 and 10,000 to handle the variance in EMG amplitude.

Before used in any specific application, EMG is typically filtered with high- and low-pass filters with cut-off frequencies of 10-20 Hz and 400-450 Hz respectively [15]. High-pass filters to avoid motion artifacts and low-pass filters to remove high frequency noise not being components of the EMG.

The signal is finally digitized with an Analog to Digital Converter (ADC). A sampling frequency of 2 kHz is considered a sufficient sampling rate for most related applications. The complete acquisition setup is illustrated in Figure 4.

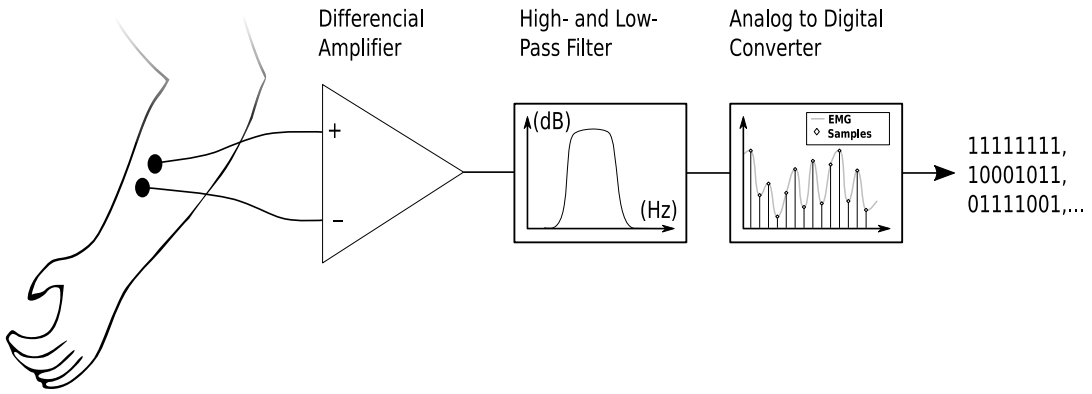


Figure 4: Electromyography aquacition including the electrode interface, amplification, filtering and analog to digital conversion.

2.3 Feature Extraction, Feature Space and Feature Selection

The use of raw EMG for MPR applications is rare and can be considered an exception, even though it has been done [18]. As briefly explained in the introduction, the general method is to extract features from time windows of EMG to reduce redundant information, while maintaining the useful information in the signal. The features can be statistical descriptors such as variance or other signal attributes *e.g.* the number of zero crossings or the waveform length [19]. The length of the time windows changes the performance of the MPR implementation. Long windows leads to more stable feature extraction, *i.e.* lower feature variance, but will increase the response delay of the MPR, which will ultimately affect the control performance [3]. Time windows can be extracted with or without overlap as illustrated in Figure 5.

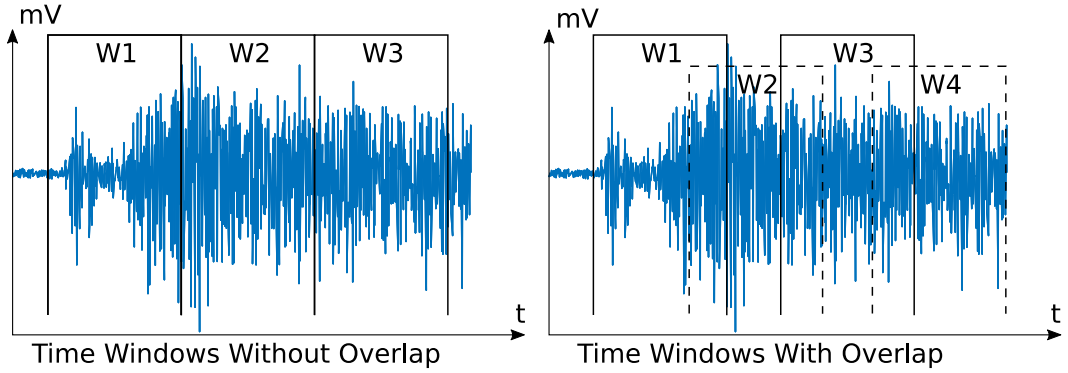


Figure 5: Time windows extracted from electromyography with and without overlap.

The aforementioned features are usually used for all EMG channels, making the number of features for one time window the number of channels times the number of features. See Figure 6

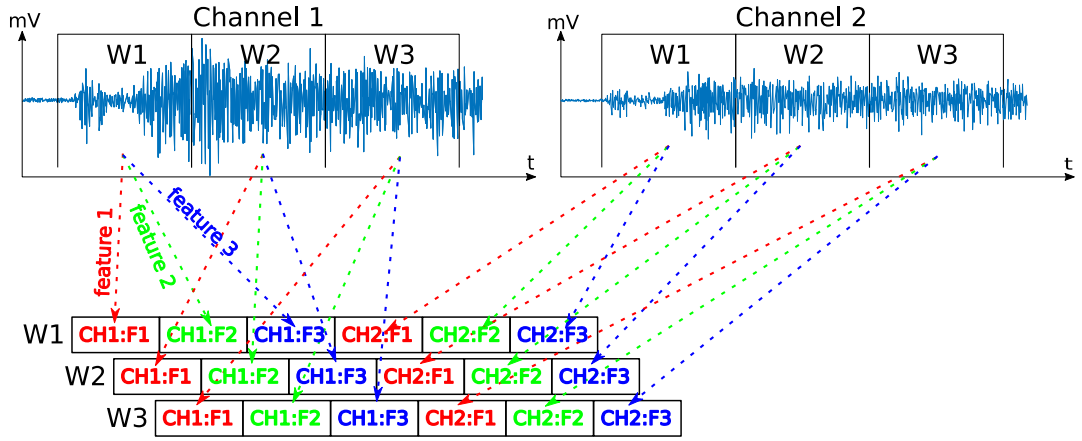


Figure 6: Feature vectors constructed by features extracted from three time windows (W1-3) in a two channel electromyography recording.

Features extracted from all the time window constitutes the distribution of the recording in feature space. One point in feature space is a representation of a time window with coordinates according to the features vector. Figure 7 inset A shows one point in a two dimensional feature space, while inset B shows how training data forms clusters representing different recorded movements. These clusters are the patterns which the classifier learns so that future observations can be classified and labeled correctly.

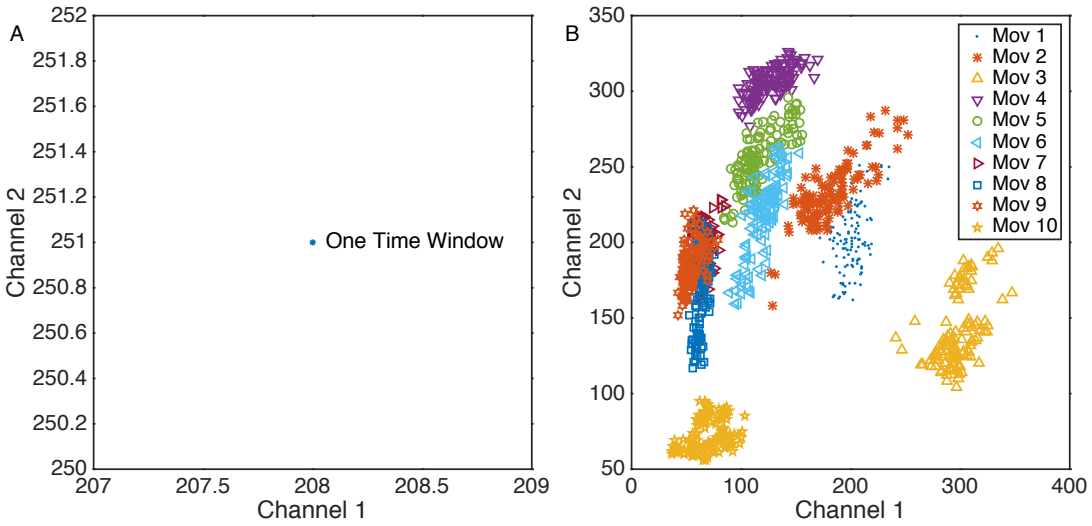


Figure 7: Feature vectors extracted from time windows of a two channel electromyography recording using the feature cardinality [20]. The feature vectors are illustrated as points in the feature space corresponding to the feature vector structure. Inset A show one such feature vector, while Inset B shows clusters of feature vectors representing 10 movements differentiated by color and marker.

Figure 7 inset B shows how clusters representing 10 movements are distributed when using the feature "cardinality" [20]. Comparing Figure 7 inset B with Figure 8, where "slope sign change" was used instead of "cardinality", emphasizes the importance of carefully considered feature selection.

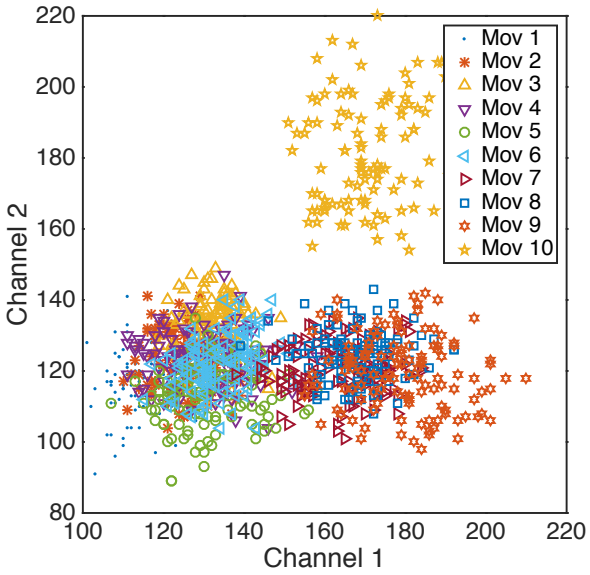


Figure 8: Clusters of feature vectors representing 10 movements differentiated by color and marker. The feature vectors were extracted from time windows of a two channel electromyography recording using the feature "slope sign change". The feature vectors are illustrated as points in the feature space corresponding to the feature vector structure.

A commonly used feature set is the so called *Hudgins set*, first introduced by Hudgins *et. al.* based on there high performance in extensive experiments [21]. Similar studies of features general performance have suggested other sets of features [19], [20], [22], [23], while other studies focus on algorithms automatically selecting features based on the data acquired for classifier training [24]–[27]. Such feature selecting algorithms evaluates the data on its separation and redundancy.

The separabillity of data is determined by the conflicts between the classes in feature space, while redundancy points out the amount of repeated information in the data. *E.g.* two instances in a set of features that are highly correlated. These attributes of the data are highly influencing the complexity of the classification, making the algorithms used for feature selection highly relevant to this thesis work. Such algorithms are described in more detail and are used for classification complexity estimation in the appended papers.

2.4 Classification and Pattern Recognition Algorithms

MPR is a part of the supervised machine learning family. Supervised machine learning can be further divided into classification and regression. Classification aims to correctly label a given input, *e.g.* tell if an image contains a dog or a horse, while regression responds with a continuous output, *e.g.* the resulting torque output from an engine model [28]. Even though some classifiers used for MPR gives a continuous output corresponding to the probability that the input is representing a specific class, the aim is to provide the class label, which makes MPR a classification task.

Classifiers are evaluated on their ability to correctly label observations as the class that they originate from, but also on attributes more related to control performance, such as response time. Offline tests, where the classifiers are used on pre-recorded data, clearly reveals the former performance attributes, while real-time tests, *e.g.* the *Motion Test* introduced by Kuiken *et al.* where classification is done as data is acquired [29], reveals a more over-all performance of the classifier. Classification accuracy is a commonly used measurement of classifier performance and is the ratio between correctly classified observations and the total number of observation.

A wide range of classifiers have been used and evaluated for MPR [30]. The following sections describes and motivates the selection of the PRAs used in this thesis.

2.4.1 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) was selected to be used in this thesis because of its simplicity, robustness [31] and because it has been extensively used in studies of MPR [3], [11], [19], [20], [31]. LDA is special case of Fisher Discriminant Analysis, which is a measurement of separation in class labeled data defined as the ratio between the class variance, *i.e.* variance of the class means, and the within class variance [32]. LDA however, is simplified by the assumption of normality and common within class variance, *i.e.* $\sum_1 = \sum_2 = \ldots = \sum_c = \sum$ where $\sum_c$ is the covariance of class c and c is the number of classes in the classification task.

When LDA is used for classification the resulting classifier is built up by statistical models of the classes derived under the aforementioned assumption of LDA, *i.e.* normal distribution models according to $X_c \sim N(\boldsymbol{\mu_c}, \sum)$ for $c = 1, 2, \ldots, C$ were C is the number of movements, $\boldsymbol{\mu_c}$ is a vector of mean values for all dimensions of the feature space. [33]. The models are illustrated together with the corresponding clusters in Figure 9.

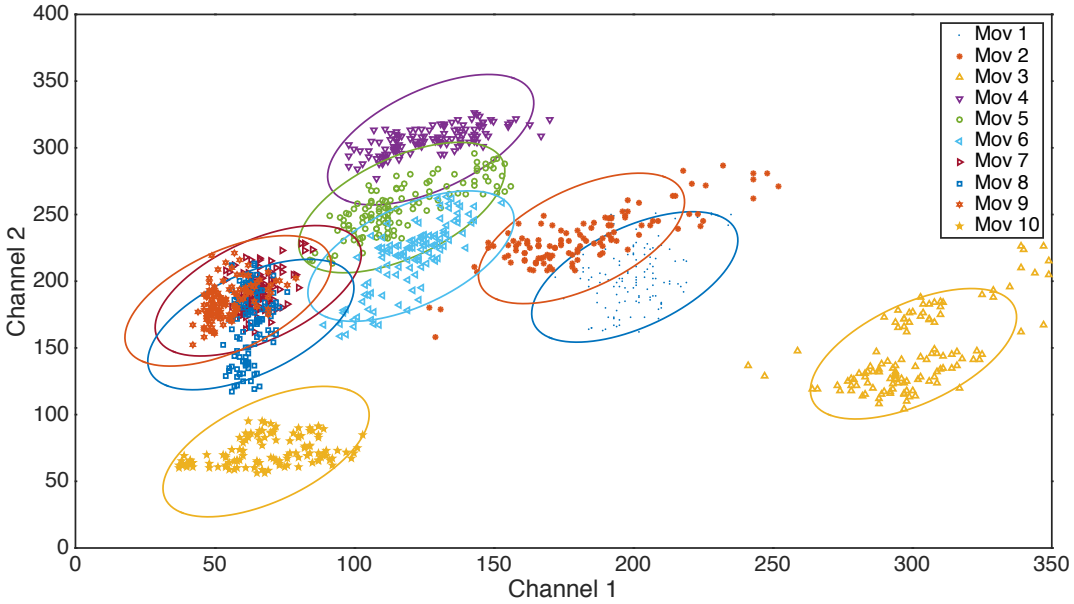


Figure 9: The clusters of Figure 7 representing 10 movements plotted together with the models used by LDA to classify new observations. The models are presented as ellipsoids blotted around the center of the clusters (mean values for the two dimensions) and constructed according to the assumed common covariance.

A new observation is classified based on the models, but the exact use differs between implementations. A common way is by so called discriminant functions, which describes decision boundaries derived from the models [34], Figure 10 illustrates a classification problem with movement 1, 3 and 6 from Figure 9 including the decision boundaries and the resulting decision regions. A new observation is classified according to the decision region in which is it occurs.

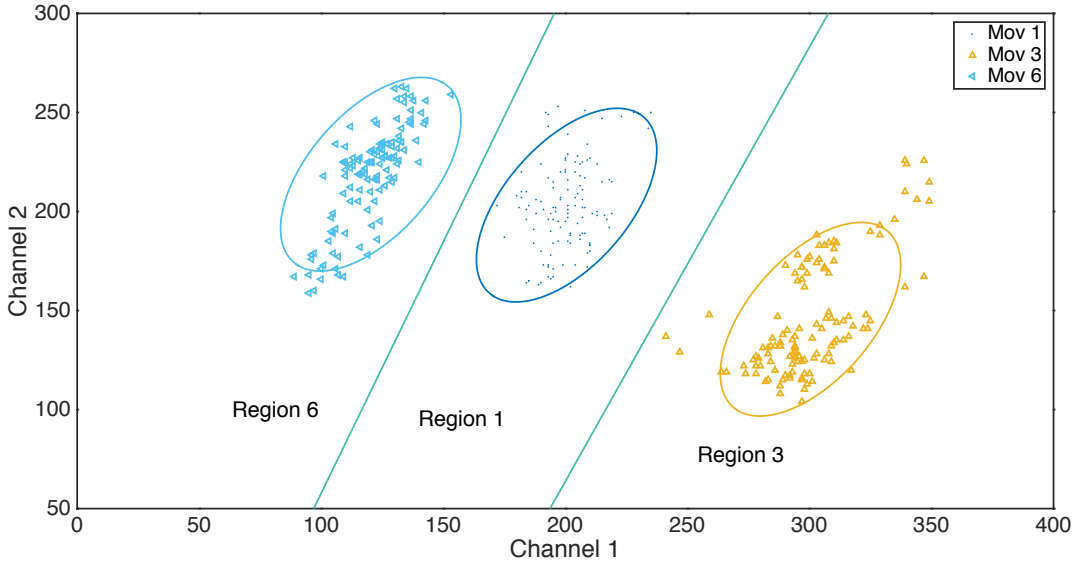


Figure 10: A classification problem with movement 1, 3 and 6 from Figure 9. The movements are represented by the original data cluster, the model of the data and the decision region. A new observation is classified according to the region in which it occurs.

2.4.2 Support vector machine

Support vector machine (SVM) was selected for this thesis because of the general interest in the algorithm within and outside of the MPR field. SVM is based on decision boundaries similar to those derived using LDA described in the previous section. SVM however, uses a different approach.

First of SVM is based on so called *kernel functions*. The basic idea behind *kernel functions* is to map an original feature space, where clusters of training data are not linearly separable, into one were they are. Figure 11 illustrates the transformation of a two dimensional feature space using a mapping function $\Phi(\cdot)$ [34].

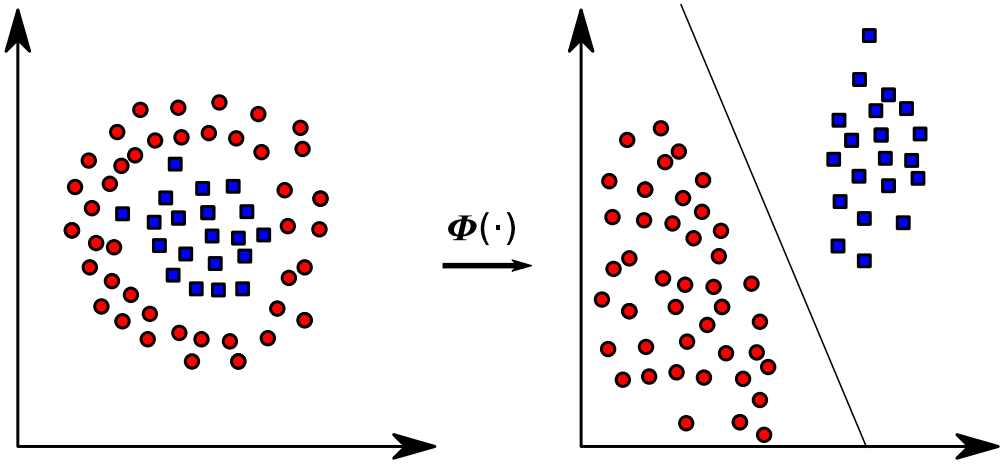


Figure 11: Illustration of the idea behind kernel functions. The original space to the left is transformed by the function $\Phi(\cdot)$ to the kernel space to the right with the aim to enable linear separation.

A simple mapping function $\Phi(\boldsymbol{x}) = \boldsymbol{x}$, where $\boldsymbol{x}$ is a feature vector from the training data, is introduced for explanatory reasons. The corresponding kernel function, called the linear kernel, is defined as

$$k_l(\boldsymbol{x}, \boldsymbol{x}') = \Phi(\boldsymbol{x})^T \Phi(\boldsymbol{x}') = \boldsymbol{x}^T \boldsymbol{x}'$$

where $\boldsymbol{x}'$ is an interacting vector with the size of $\boldsymbol{x}$. Note that the linear kernel does not transform the original feature space [34].

For algorithms where the input vector $\boldsymbol{x}$ is only used in a scalar product, as for the linear kernel function, that scalar product can be replaced with some other choice of kernel function with the aim to increase the class separability. SVM is such an algorithm. Other examples of kernel functions are

- Polynomial kernel, $k_p(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}^T \boldsymbol{x}')^d$
- Radial basis function kernel, $k_g(\boldsymbol{x}, \boldsymbol{x}') = exp(-\|\boldsymbol{x} - \boldsymbol{x}'\|)^2/2\sigma^2$

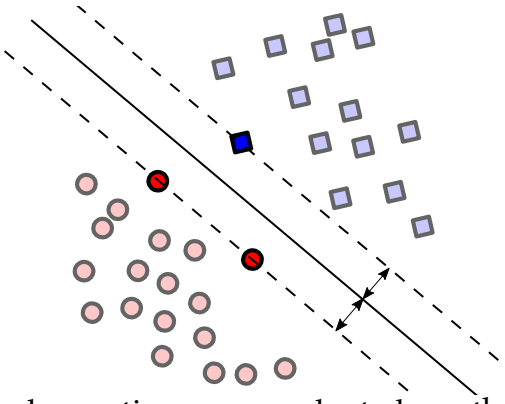


Figure 12: A decision boundary that maximizes the margin to the closest data point of both classes.

SVM is, in its original form, a two-way classifier, *i.e.* classifies inputs as one out of two classes. Regardless the selection of kernel function, SVM classifies new observations using so called *support vectors*. A decision boundary is chosen to maximize the margin to the closest data points of both classes, see Figure 12. These points are saved to be used as *support vectors*. The decision boundary is not needed for classification as new observations are evaluated on there relation to the support vectors in the transformed feature space [34].

When SVM is used for classification of multiple classes, the decision is commonly based on a majority voting of multiple classifiers. Either one for ever pair of classes or one for every class, where a target class is compared with a class representing all other classes.

2.4.3 Multi-Layer Perceptron

Multi-Layer Perceptron (MLP) was shown to yield high classification accuracy in a number of studies of MPR [4], [11], [20], [31], [35], and is therefor a natural PRA for this thesis.

MLP is a common implementation of a so called *artificial neural network*, that originates from an attempt at describing the data processing of biological system. The concept is a network built up by units representing the neurons in a biological neural network. The units are connected to receives and provide information. A MLP network consists of layers of units were all units of one layer is connected to receive input from all units in the previous layer. This results in a directional flow of information, making MLP a feed-forward neural network [34]. There are three subgroups of layers; the input layer, providing the network with initial input; hidden layers, where the inputs are processed; and output layer, were the interpretation of the input is provided. Figure 13 inset A illustrates a network with one hidden layer. The functionality of one unit is illustrated in Figure 13 inset B, where the inputs are received from either the input layer or a previous hidden layer. The sum of the weighted inputs is fed to an activation function, which vary depending on the desired attribute of the network.

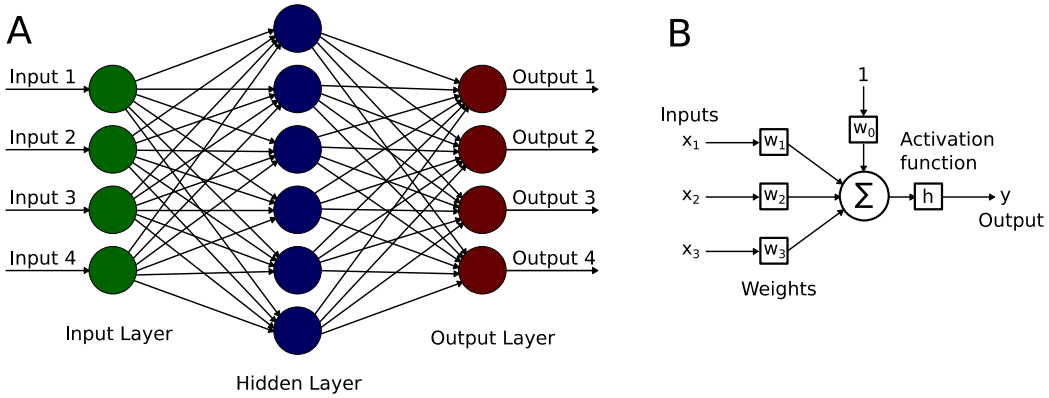


Figure 13: Inset A illustrates a feed-forward neural network with four inputs, one hidden layer with six units and 4 outputs. Inset B explains the structure of one unit of a feed-forward neural network.

The following equation describes the computation of the output of unit k.

$$y_k = h\left(\sum_{i=0}^{K} x_i w_i\right)$$

where x_i is the output of the i:th unit of the previous layer, w_i is the weight for the corresponding input, K is the number of units in the previous layer and $h(\cdot)$ is the activation function. Note that i starts at 0. Adding an input $x_0 = 1$ enables the activation function to operate around a fixed value, usually 0 and can consequently be equal for all units.

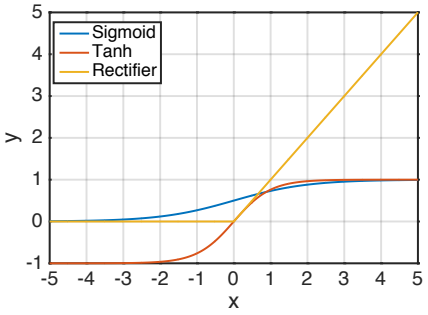


Figure 14: Plot of three activation function for Multi-Layer Perception networks, namely logistic sigmoid function, Tanh [34] *and rectifier function* [38].

Examples of activation functions are illustrated in Figure 14.

An MLP network is trained by applying something called *Error Backpropagation*. An error function, providing the error in the actual output with respect to the expected output of the network, is used to derive the gradient of the error as a function of the weights. The weights are adapted in the direction of the gradient to ultimately decrease the error. The process is repeated for all the training data and until the error reaches a desired level.

3 Implemention in BioPatRec

BioPatRec is a modular software tool for benchmarking and development of algorithms within the MPR field that was made freely available online [7], [36]. It includes all steps needed from EMG acquisition for PRA training to control of a virtual arm and is structured as shown in Figure 15.

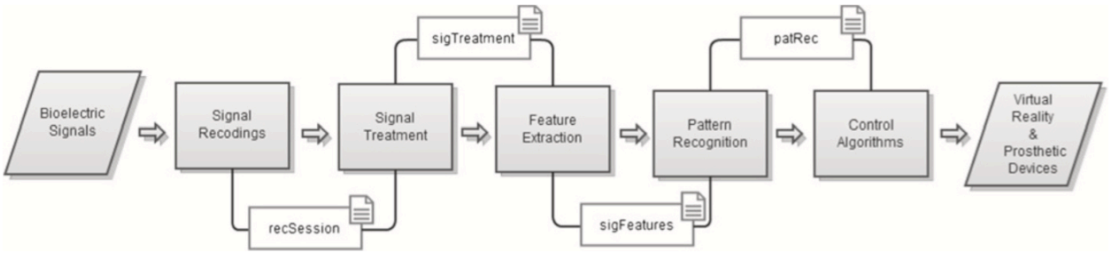


Figure 15: The modular structure of BioPatRec . The modules are connected through data arraws, which can be saved and loaded between the different modules.

Three CCEAs that were found relevant for MPR through the experiments of Paper II were implemented in BioPatRec as part of introduced data analysis and feature selection modules. The three CCEAs were the following; *Separability Index* based on *Modified Mahalanobis* and *Bhattacharyyas Distance* as well as *Nearest Neighbor Separability*. These three algorithms will be referred to as the *usefull algorithms* in the continuation of this section. [37]

3.1 Data Analysis

The aim for the data analysis module is to visualize the complexity yielded by an EMG recording acquired for PRA training and predict the resulting offline classification accuracy. It receives a recSession, which includes the recorded EMG, and can be used to edit the recSession for better separation, hence it is a part of the Signal Recordings module, see Figure 15 [7]. A more detailed description with references to the screen shot of the graphical used interface (GUI) in Figure 16 follows bellow.

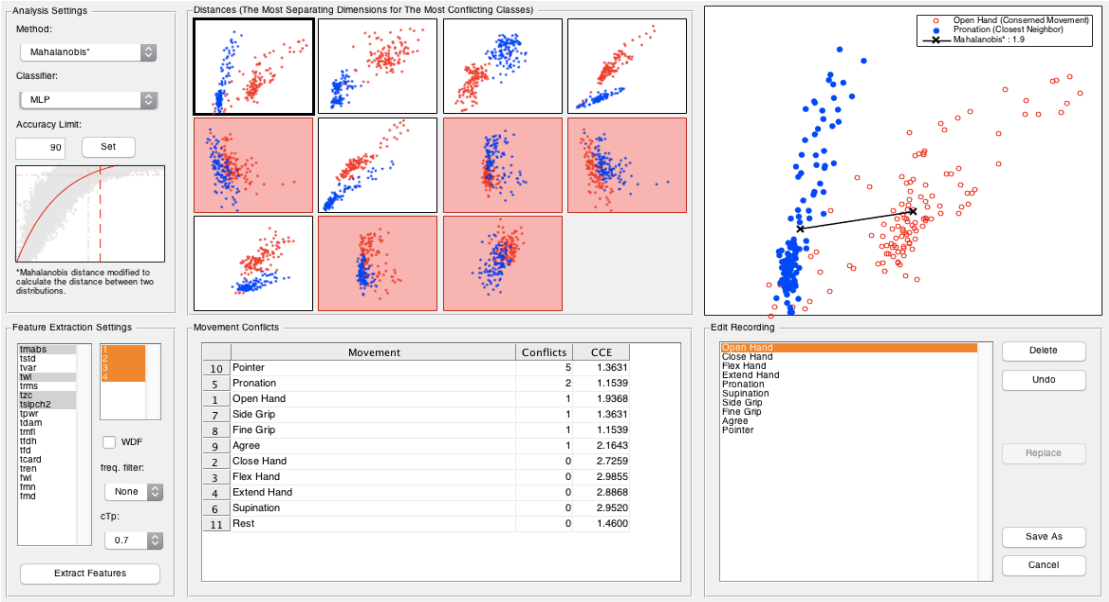


Figure 16: A screenshot of the data analysis tool of BioPatRec. Showing a recording of a hand performing 10 movement plus rest.

3.1.1 Analysis Setting

This part allows the user to select the algorithm for classification complexity estimation in the pop-up menu under “Methods”. The algorithms available are the three *usefull algorithms*. A prediction of offline classification accuracy is presented in the small predictions plot in the lower part of the section. The predicted result is for the classifier selected in the pop-up menu under “Classifier”. LDA used in single classifier and “One-Vs-One” topologies are together with MLP used in a “All-Movements-as-Individual” configuration available for selection. For more details on the topologies and configurations, se appended Paper II. The red line represents the predicted accuracy as a function of the Classification Complexity Estimate (CCE) of the selected *useful algorithm*. The function was derived from data collected during the experiment of the appended papers. The collected data is plotted as grey dots in the prediction plot. A desired offline accuracy limit is selected by the user, under “Accuracy limit”, and plotted together with the corresponding CCE limit in the prediction plot as “— · —“ lines. The CCE for the movement selected in the “Distances” section is plotted as a clear red “— —“ line. The content of the prediction plot is copied to a separate window with explanatory legends by mouse control. 1

3.1.2 Distances

The “Distances” section illustrates the classification complexity for each movement by plotting the training data for the movement considered and for its most conflicting neighbor in the different instances of the plot matrix. The two dimensions are selected as the most separating of the two movement. The separation evaluation and the selection of most conflicting neighbor are both done by the selected *usefull algorithm.* A selection of an instance in the plot matrix is done by mouse control and copies the content of the selected plot to the larger window next of the matrix, allowing closer analysis with explanatory legends. The CCE of the movement corresponding to the selected instance is shown in the legend as well as in the prediction plot of the ”Analysis Setting” section. Instances resulting in a offline accuracy prediction lower than the limit selected in the “Analysis Setting” section are highlighted with red color.

3.1.3 Feature Extraction Settings

This section controls the extraction of features from the recorded EMG. The features and channels to be used are selected in the two listboxes. The pop-up menu under “freq. filter” allows selection of frequency based filter to be used and the “WDF” checkbox activates a wavelet denoising filter. The filters are adapted from the Feature Extraction module of BioPatRec [7]. Contraction time percentage [7] is selected in the pop-up menu under “cTp”. The button “Extract Features” activates a feature extraction according to the settings followed by an analysis according to the settings in “Analysis Settings”.

3.1.4 Movement Conflict

The table in the “Movement Conflict” section shows the number of times the movements are selected as the most conflicting neighbor to emphasize how detrimental the movements are to the classification task. It also contains all CCEs.

3.1.5 Edit Recording

This section lists the movements found in the recSession and allows the user to edit it and save a new more separable version. Pressing the "Delete" or "Replace" button removes or initializes the replacement of the movements selected in the listbox. The replacement of movements is done by a repeated recording of the selected movements and is only allowed if the analysis tool is started from the recording session (Signal Recording in Figure 15). If movement have been deleted from the list, the classification task constituted by the new set of movement can be analyzed by simple clicking the "Extract Feature" button in the feature extraction section. The list is restored to the original set of movements by use of the "Undo" button. The "Save" button saves the edited version of the recSession in a new file.

3.2 Feature selection

The selection of features in BioPatRec is access by the Pattern Recognition module in Figure 15 [7]. The original methods are manual selection or selection of predefined sets of 2, 3 or 4 features commonly used for MPR. The method contributed by this thesis is an automatic selection made based on the training data and the selected *useful algorithm*. An additional GUI appears when the automatic feature selection is activated from the Pattern Recognition module, see Figure 17.

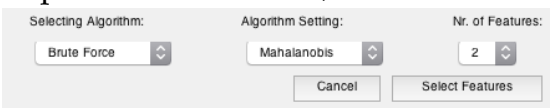


Figure 17: Graphical User Interface for automatic feature selection in the Patter Recognition module of BioPatRec.

A feature selecting algorithm is selected in the pop-up menus under "Selecting Algorithm". The available algorithms are the brute-force search used for the appended Paper II and a sequential selection adding the most contributing feature repeatedly till the desired number of features are selected. The evaluations of the feature sets are done by the *useful algorithm* selected under "Algorithm Setting". If f is the numbers of features to select from and n is the desired number features in the feature set, the sequential selecting algorithm will run the selected *useful algorithm* $f * n$ times. The brute-force search however, will evaluate every combination of n features, resulting in $\frac{f!}{n!(f-n)!}$ computations. That leads to 85 and 6188 computations for the sequential and brute-force selection algorithm respectively, given that $f = 17$ and $n = 5$. A comparison of sequential and the brute-force selection is shown in Figure 18. The sequential selection of features is, compared to brute-force selection, little detrimental for *Nearest Neighbor Separability* and *Separability index* used with *Modified Mahalanobis* actually benefit from it. Considering the significantly reduced computation complexity and the results presented in Figure 18, the sequential selecting algorithm should be the preferred selecting algorithm of the two.

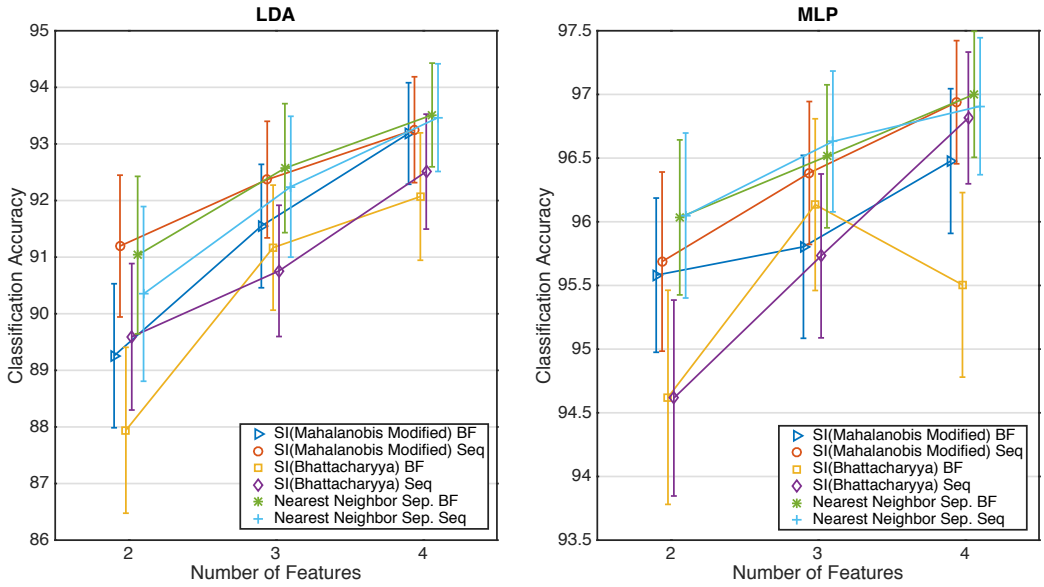


Figure 18: Yielded offline classification accuracy for brute-force search (BF) and sequential selection (Seq) of features sets. The markers are plotted at the means and the bars represent the standard error. The results are computed using the data referred to as IM data in the appended Paper II.

4 Summary of Papers

4.1 Paper I - Estimates of Classification Complexity for Myoelectric Pattern Recognition

N. Nilsson and M. Ortiz-Catalan, "Estimates of Classification Complexity for Myoelectric Pattern Recognition," *Pattern Recognition (ICPR), 2016 23nd International Conference on*, Cancun, 2016

In Paper I, different Classification Complexity Estimating Algorithms (CCEAs) were evaluated based on there correlation with classification accuracy using a number of commonly used classifiers for prosthesis control based on myoelectric pattern recognition (MPR). Two algorithms, namely Separability Index and Nearest Neighbor Separability, were found to yield high correlation and were further evaluated on the performance of feature sets selected based on the outcome of the algorithms.

The data set used in the experiments was limited to individual movements, *i.e.* the execution of one movement at the time, and the results presented were exclusively from offline tests.

Paper I contributes by deepening the understanding of the complexity of offline MPR of individual movements (classes).

4.2 Paper II - Electromyography Data Analysis for Myoelectric Pattern Recognition

N. Nilsson and M. Ortiz-Catalan, "Electromyography Data Analysis for Myoelectric Pattern Recognition," *In manuscript*

Paper II continues the evaluation of the algorithms found useful in paper I. An additional data set including simultaneous movement, *e.g* flexing and closing a hand simultaneously, was introduced together with two additional CCEAs. The additional algorithms, Purity and Repeatability Index, were found insufficient compare to the previously evaluated algorithms.

Paper II also deepened the analysis of the algorithms by further evaluating their different settings and attributes. The CCEAs dependency on correlation between features/channels and on the dimensionality of the classification task was in that way revealed.

In addition to offline tests done in Paper I, Paper II also evaluated the CCEAs relevance for real-time performance prediction.

Paper II further deepens the understanding of the complexity of MPR and provides tools to analyze MPR applications.

5 References

[1] M. A. Hanley, D. M. Ehde, K. M. Campbell, B. Osborn, and D. G. Smith, "Self-reported treatments used for lower-limb phantom pain: Descriptive findings," *Arch. Phys. Med. Rehabil.*, vol. 87, no. 2, pp. 270–277, 2006.

[2] A. O. Andrade, A. A. Pereira, S. Walter, R. Almeida, R. Loureiro, D. Compagna, and P. J. Kyberd, "Bridging the gap between robotic technology and health care," *Biomed. Signal Process. Control*, vol. 10, no. 1, pp. 65–78, 2014.

[3] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *J. Rehabil. Res. Dev.*, vol. 48, no. 6, pp. 643–660, 2011.

[4] M. Ortiz-Catalan, N. Sander, M. B. Kristoffersen, B. Håkansson, and R. Brånemark, "Treatment of phantom limb pain (PLP) based on augmented reality and gaming controlled by myoelectric pattern recognition: A case study of a chronic PLP patient," *Front. Neurosci.*, vol. 8, no. 8 FEB, pp. 1–7, 2014.

[5] J. Liu, X. Li, G. Li, and P. Zhou, "EMG feature assessment for myoelectric pattern recognition and channel selection: A study with incomplete spinal cord injury," *Med. Eng. Phys.*, vol. 36, no. 7, pp. 975–980, 2014.

[6] N. E. Bunderson and T. A. Kuiken, "Quantification of feature space changes with experience during electromyogram pattern recognition control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 239–246, 2012.

[7] M. Ortiz-Catalan, R. Brånemark, and B. Håkansson, "BioPatRec: A modular research platform for the control of artificial limbs based on pattern recognition algorithms.," *Source Code Biol. Med.*, vol. 8, no. 1, p. 11, 2013.

[8] A. G. Nerlich, A. Zink, U. Szeimies, and H. G. Hagedorn, "Ancient Egyptian prosthesis of the big toe," *Lancet*, vol. 356, no. SUPPL., pp. 2176–2179, 2000.

[9] A. Roche, H. Rehbaum, D. Farina, and O. Aszmann, "Prosthetic Myoelectric Control Strategies: A Clinical Perspective," *Curr. Surg. Reports*, vol. 2, no. 3, pp. 1–11, 2014.

[10] D. S. Childress and D. Ph, "Historical Aspects of Powered Limb Prostheses," *Clin. prosthetics Orthot.*, vol. 9, no. 1, pp. 2–13, 1985.

[11] M. Ortiz-Catalan, B. Håkansson, and R. Brånemark, "Real-time and simultaneous control of artificial limbs based on pattern recognition algorithms," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 4, pp. 756–764, 2014.

[12] M. B. I. Raez, M. S. Hussain, F. Mohd-Yasin, M. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: detection, processing, classification and applications.," *Biol. Proced. Online*, vol. 8, no. 1, pp. 11–35, 2006.

[13] E. Spinelli and M. Haberman, "Insulating electrodes: a review on biopotential front ends for dielectric skin-electrode interfaces.," *Physiol. Meas.*, vol. 31, no. 10, pp. S183–S198, 2010.

[14] S. Micera, J. Carpaneto, and S. Raspopovic, "Control of hand prostheses using peripheral information," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 48–68, 2010.

[15] B. Gerdle, S. Karlsson, S. Day, and M. Djupsjobacka, "Acquisition, processing and analysis of the surface electromyogram," *Mod. Tech. Neurosci.*, pp. 705–755, 1999.

[16] M. Ortiz-Catalan, R. Brånemark, B. Håkansson, and J. Delbeke, "On the viability of implantable electrodes for the natural control of artificial limbs: Review and discussion," *Biomed. Eng. Online*, vol. 11, no. 1, p. 33, 2012.

[17] M. Ortiz-Catalan, B. Hakansson, and R. Branemark, "An osseointegrated human-machine gateway for long-term sensory feedback and motor control of artificial limbs," *Sci. Transl. Med.*, vol. 6, no. 257, pp. 257re6–257re6, 2014.

[18] N. Bu, O. Fukuda, and T. Tsuji, "EMG-based motion discrimination using a novel recurrent neural network," *J. Intell. Inf. Syst.*, no. 1978, pp. 113–126, 2003.

[19] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, 2012.

[20] M. Ortiz-Catalan, "Cardinality as a highly descriptive feature in myoelectric pattern recognition for decoding motor volition," *Front. Neurosci.*, vol. 9, no. OCT, pp. 1–7, 2015.

[21] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, 1993.

[22] M. Ortiz-Catalan, R. Brånemark, and B. Håkansson, "Biologically inspired algorithms applied to prosthetic control," *IASTED Int. Conf. Biomed. Eng.*, no. BioMed, pp. 7–15, 2012.

[23] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition.," *J. Neuroeng. Rehabil.*, vol. 7, no. 1, p. 21, 2010.

[24] Z. Yan, "The Study on Feature Selection Strategy in EMG Signal Recognition," pp. 711–716, 2013.

[25] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005.

[26] J. Liu, "Feature dimensionality reduction for myoelectric pattern recognition: A comparison study of feature selection and feature projection methods," *Med. Eng. Phys.*, vol. 36, no. 12, pp. 1716–1720, 2014.

[27] Q. Cheng, H. Zhou, and J. Cheng, "The fisher-markov selector: Fast selecting maximally separable feature subset for multiclass classification with applications to high-dimensional data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1217–1233, 2011.

[28] T. Christiano Silva and L. Zhao, *Machine Learning in Complex Networks*.

2016.

[29] T. A. Kuiken, B. A. Lock, R. D. Lipschutz, L. A. Miller, K. A. Stubblefield, and K. B. Englehart, "Targeted Muscle Reinnervation for Real-time Myoelectric Control of Multifunction Artificial Arms," *JAMA*, vol. 301, no. 6, pp. 619–628, 2016.

[30] B. Karlık, "Machine Learning Algorithms for Characterization of EMG Signals," *Int. J. Inf. Electron. Eng.*, vol. 4, no. 3, pp. 189–194, 2014.

[31] P. Kaufmann, K. Englehart, and M. Platzner, "Fluctuating EMG signals: Investigating long-term effects of pattern matching algorithms," *2010 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC'10*, pp. 6357–6360, 2010.

[32] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2. pp. 179–188, 1936.

[33] J. H. Friedman, "Regularized Discriminant Analysis," *J. Am. Stat. Assoc.*, vol. 84, no. 405, pp. 165–175, 1989.

[34] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 4, no. 4. 2006.

[35] J. U. Chu, I. Moon, and M. S. Mun, "A real-time EMG pattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2232–2239, 2006.

[36] M. Ortiz-Catalan, "BioPatRec," 2016. [Online]. Available: https://github.com/biopatrec/biopatrec/. [Accessed: 16-Jun-2016].

[37] K. K. Htike and D. C. Hogg, "Efficient non-iterative domain adaptation of pedestrian detectors to video scenes," *Proc. - Int. Conf. Pattern Recognit.*, pp. 654–659, 2014.

[38] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proc. 27th Int. Conf. Mach. Learn.*, no. 3, pp. 807–814, 2010.

Appended Papers

Paper I

Estimates of Classification Complexity for Myoelectric Pattern Recognition

Niclas Nilsson
Department of Signals and Systems
Chalmers University of Technology
Gothenburg, Sweden
Email: nniclas@student.chalmers.se

Max Ortiz-Catalan
Chalmers University of Technology
Gothenburg, Sweden
Sahlgrenska University Hospital
Gothenburg, Sweden
Integrum AB
Mölndal, Sweden
Email: maxo@chalmers.se

***Abstract*—Myoelectric pattern recognition (MPR) can be used for intuitive control of virtual and robotic effectors in clinical applications such as prosthetic limbs and the treatment of phantom limb pain. The conventional approach is to feed classifiers with descriptive electromyographic (EMG) features that represent the aimed movements. The complexity and consequently classification accuracy of MPR is highly affected by the separability of such features. In this study, classification complexity estimating algorithms were investigated as a potential tool to estimate MPR performance. An early prediction of MPR accuracy could inform the user of faulty data acquisition, as well as suggest the repetition or elimination of detrimental movements in the repository of classes. Two such algorithms, Nearest Neighbor Separability (NNS) and Separability Index (SI), were found to be highly correlated with classification accuracy in three commonly used classifiers for MPR (Linear Discriminant Analysis, Multi-Layer Perceptron, and Support Vector Machine). These Classification Complexity Estimating Algorithms (CCEAs) were implemented in the open source software BioPatRec and are available freely online. This work deepens the understanding of the complexity of MPR for the prediction of motor volition.**

I. Introduction

Myoelectric Pattern Recognition (MPR) has been shown to have great potential as part of the control strategy for a number of clinical applications, such as upper-limb prostheses control [1], phantom limb pain treatment [2] and rehabilitation after stroke [3]. Electromyography (EMG) is commonly acquired using surface electrodes (SEs) that are sensitive to changes in environmental conditions and motion artifacts [4], which makes frequent calibration or training of the applied pattern recognition algorithm (PRA) necessary. In order to acquire the data needed for such calibration or training, EMG is recorded while the patient performs muscle contractions relevant to the desired movements. Such recordings might be affected by errors due to the surface electrodes instability but also by human factors.

Reaz *et al.* suggested the analysis of important EMG attributes, such as the signal to noise ratio, in order to enable high MPR accuracy [5]. However, analyzing data based on these attributes requires experience and time. The literature on automated data analysis methods is limited despite the well-known consequences of using low quality recordings.

Apart from a few exceptions most studies on MPR use features that are extracted from raw EMG [6]. PRAs classification accuracy is highly dependent on the feature sets used as input, and therefore studies have been conducted on the performance of a variety of EMG features, as well as on the selection of such features [7], [8]. Liu *et al.* applied two feature selecting algorithms, Minimum Redundancy and Maximum Relevance (mRMR) and Markov Random Fields (MRF), to an electrode array setup [9]. The Kullback-Leibler Divergence was used in mRMR to rate relevance and redundancy of features and channels, which were ranked and selected into sets according to these ratings [10]. MRF was employed similarly to mRMR, but the features and channels were rated based on inter and intra class scatters, as well as total data scatter [11]. Bunderson *et al.* defined three data quality indices, namely Repeatability Index (RI), Mean Semi-principal Axis (MSA) and Separability Index (SI). These indices were used to rate the subjects ability to increase data quality when EMG was recorded repeatedly over several days [12]. Even though none of the studies above aimed to estimate classification complexity, they suggested useful ways to draw information from EMG when predicting classification accuracy.

Studies on Classification Complexity Estimates (CCEs) are more common outside the MPR field. Two nonparametric multiresolution complexity measures, Nearest Neighbor Separability (NNS) and Purity, were defined by Singh in 2003 [13]. These CCEs showed promising results but were not evaluated using EMG. Singh compared his results with a number of statistical similarity measures which were also potentially adequate CCEs. Among them were Kullback-Leibler Divergence, Bhattacharyya distance [14] and Mahalanobis distance [15]. In the present study these algorithms were used to describe the complexity of MPR to decode motor volition.

The aforementioned Classification Complexity Estimating Al-

gorithms (CCEAs) were implemented in BioPatRec, which is an open source tool for the development and benchmarking of algorithms for advanced bioelectric control [16]. BioPatRec enables recording, preprocessing, feature extraction, pattern recognition and real-time control of artificial limbs using bioelectric signals. In the work presented here we evaluated the outcome of CCEAs and compared it with the accuracy of a number of classifiers. The resulted correlations provided evidence of CCEAs suitability to inform on MPR complexity. All code and data used in this study is available online [17].

II. Methods

A. Data Set

The data set used in this study is included in the BioPatRec data repository and is available online [17]. EMG was recorded in 20 subjects who performed 11 movements (Hand open/close, wrist flexion/extension, pro/supination, side grip, fine grip, agree or thumb up, pointer or index extension and rest) [16]. Disposable Ag/AgCl electrodes (⊘ = 1 cm) were place over the skin in bipolar configurations with 2 cm inter-electrode distance. The first channel was placed along the extensor carpi ulnaris mucle, and the rest (three) were equally spread around the most proximal third of the forearm. The more proximal electrode of every bi-pole was connected to the positive terminal of the amplifier.

B. Recording and Pre-Processing

The subjects were requested to perform each movement 3 times and rest in between. The movement was held during 3 seconds (contraction time) and the resting time was 3 seconds. To avoid inactivity periods being considered as movement related information due to delay between request and reaction, only 70 % of the contraction time was used. This percentage of the contraction time has been found to exclude inactive periods while keeping the dynamic portion of the contraction [16].
Sliding time windows of 200 ms with a 50 ms increment were used to extract a variety of signal features. The feature vectors were randomly distributed into sets for training (40%), validation (20%) and testing (40%) before training the classifiers. No data from the testing set was used during training and validation of the classifier.

C. Classification Complexity Estimating Algorithms

The CCEAs were designed to accept different numbers of channels and features, which allows for the estimation of classification complexity for individual and sets of features.

1) Separability Index: Separability Index (SI) for one class is defined as half the Mahalanobis distances (in features space) between the class and the center (mean of all dimensions) of its nearest class [12]. The distance in a two dimensional feature space is illustrated in Fig. 1 A.

The SI for a complete data set is computed by the average over all classes. See equation 1.

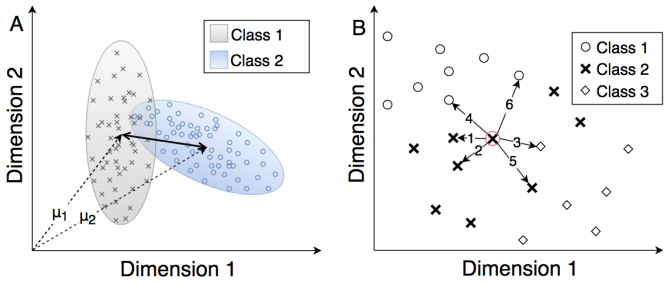


Fig. 1. Inset A shows the distance between the center points (mean of both dimensions) of two classes in a two dimensional feature space. The ellipses are constructed to represent the covariance of the classes. When the Separability Index is extracted, the distance is weighted by the covariance of the classes being compared. The bigger the weighted distance, the more separable the classes [12]. Inset B shows the six Nearest Neighbors for the target point marked by the red circle. By evaluating the dominance of Nearest Neighbors from the same class as the target point, an estimation of class separability can be established using all points in a data set as targets. Higher dominance equals higher separability [13].

$$SI = \frac{1}{c}\sum_{i=1}^{c}\left(\min_{j=1,\ldots,i-1,i+1,\ldots,c}\frac{1}{2}\times \times\sqrt{(\mu_i-\mu_j)^T S_i^{-1}(\mu_i-\mu_j)}\right) \quad (1)$$

Where c is the number of classes, μ_i and μ_j are vectors of mean values, one for every dimension, and S_i is the covariance matrix of for the class i.
However, the Mahalanobis distance does not take the variance of the nearest class into account [15]. In order to investigate if this was limiting the algorithm, a list of commonly used statistical similarity measures was implemented as distance definition for SI. Their names and equations for multivariate normal distributions are listed below, starting with the original:

- Mahalanobis distance [15]

$$D_M = \sqrt{(\mu_1-\mu_2)^T S_1^{-1}(\mu_1-\mu_2)} \quad (2)$$

- Bhattacharyyas distance [14]

$$D_B = \frac{1}{8}(\mu_1-\mu_2)^T S^{-1}(\mu_1-\mu_2) + + \frac{1}{2}ln\left(\frac{detS}{\sqrt{detS_1 detS_2}}\right) \quad (3)$$

- Kullback-Leibler Divergence [14]

$$D_{KL} = \frac{1}{2}\left(tr(S_1^{-1}S_2)+(\mu_1-\mu_2)^T S_1^{-1}(\mu_1-\mu_2)- - k + ln\left(\frac{detS_1}{detS_2}\right)\right) \quad (4)$$

- Mahalanobis distance modified to take both covariance matrices into account:

$$D_{MM} = \sqrt{(\mu_1-\mu_2)^T S^{-1}(\mu_1-\mu_2)} \quad (5)$$

By this definition SI is related to the overlap shown in Fig. 1 A. Note the similarities with equation 3. The first term is completely included while the second is left out. The second term compares the shapes of the distributions. This is relevant for similarity measures but not for separability. *E.g.* if μ_1 = μ_2, the second term could still give a high value but class separation would be impossible. This distance definition will be referred to as Modified Mahalanobis.

For all equations μ_1 and μ_2 are vectors of mean values, one for every dimension, and

$$S = \frac{S_1 + S_2}{2} \tag{6}$$

where S_1 and S_2 are covariance matrices. Subscript 1 and 2 labels the two classes being compared.

2) Nearest Neighbor Separability: Nearest Neighbor Separability (NNS) measures how well the class of a data set is represented among their nearest neighbors (NNs) in feature space [13]. Fig. 1 B show the six NNs of a target member in a two dimensional feature space. Proportions of NNs from the same class as the target were calculated. The contribution of each NN was weighted differently depending on its proximity to the target point by calculating the average of the aforementioned proportions for all numbers of NNs 1,2,...,k, where k is the number of NNs taken into account. Equation 7 shows this step for the target member in Fig. 1 B.

$$\left(\frac{1}{1} + \frac{2}{2} + \frac{2}{3} + \frac{2}{4} + \frac{3}{5} + \frac{3}{6}\right) \Big/ 6 = 0.71 \tag{7}$$

The final result was the average over all members. In the original algorithm this was repeated for a set of different resolutions [13]. In this study only the resolution 1 was used, *e.i.* feature space was not divided into hyper cuboids.

D. Features

The following features were used in this study. In time domain; mean absolute value (tmabs), standard deviation (tstd), variance (tvar), waveform length (twl), RMS (trms), zero-crossing (tzc), slope changes (diff) (tslpch), power (tpwr), difference abs. mean (tdam), max fractal length (tmfl), fractal dimension Higuchi (tfdh) fractal dimension (tfd), cardinality (tcard) and rough entropy (tren). In frequency domain; waveform length (fwl), mean (fmn) and median (fmd).

E. Classifiers

The classifiers used in this study were LDA, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM)(quadratic), and Regulatory Feedback Networks (RFN). These classifiers were used as implemented in BioPatRec [16] (code available online [17]), where LDA and SVM were implemented using pre-defined functions in Matlab.

F. Evaluation and Comparison

The data set was used in two ways. First, CCEs using individual features were compared with the resulting accuracy using only that feature. This served not only to obtain a wide range of CCEs, but also to rate the features adequacy as classifier inputs. Both accuracy and CCE were calculated for all classes and all subjects. Results given by use of individual classes are referred as individual results. Rating classes individually provided a wide CCE range as well as information about the EMG acquisition, i.e low separability for one class means high influence of error in that class. Averages over all classes were included in the result and referred as average results. Sets of 2-4 features were selected by ranking the results given by SI (Modified Mahalanobis) and NNS (k = 20) for all possible feature combinations including the feature with the highest CCE from individual feature evaluations, further referred to as *best sets*. Ortiz-Catalan proposed sets of 2-4 features found to be highly performing by a genetic algorithm [7]. These sets containing two and three features, and the Hudgins set, which is four out of five features introduced by Hudgins in [18], were used as benchmarking *reference sets*:

- **Ref 2:** tstd, trms [7]
- **Ref 3:** tstd, fwl, fmd [7]
- **Ref 4:** tmabs, twl, tslpch, tzc [18]

The *best* and *reference sets* of equal number of features were compared using different classifiers. Statistical significance was calculated with Wilcoxon signed-rank test ($p <= 0.05$). Since there was no clear linearity in the dependencies between accuracies and any CCEA, correlations were calculated using Spearmans rho.

III. Results

A. Separability Index

Correlations between SI and LDA accuracy when using the different statistical similarity measures as distance definitions are presented in Table I. Because the Modified Mahalanobis had higher correlation, and a more cohesive distribution with less outliers (see Fig. 2), it was selected as the distance definition for SI in following experiments.

TABLE I
SEPARABILITY INDEX CORRELATION WITH ACCURACY

	Individual Results	Average Results
Bhattacharyyas	0.79	0.83
Kullback-Leibler	0.78	0.69
Mahalanobis	0.85	0.78
Mahalanobis Modified	0.85	0.93

Correlations, using Spearmans rho, between accuracy and SI when using individual features classified by LDA. The statistical similarity measures in column one are used as distance definitions. Under individual results, correlations were calculated using values for every class while average over classes were used under column three.

In Fig. 3 accuracies using different classifiers are plotted against SI. All data corresponds to individual features. SVM and RFN results are more scattered compared with LDA and MLP results. They also have a considerable number of individual results close to zero accuracy, which seem uncorrelated

with SI. The average results however, show high correlation between SI and accuracy for all classifiers.

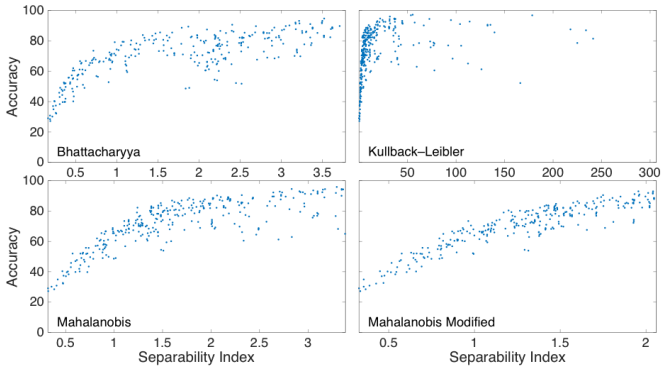


Fig. 2. Plots of accuracy against SI using individual features fed separately to LDA. The statistical similarity measure given in the plot is used as distance definitions. Accuracy and SI are average result over all classes.

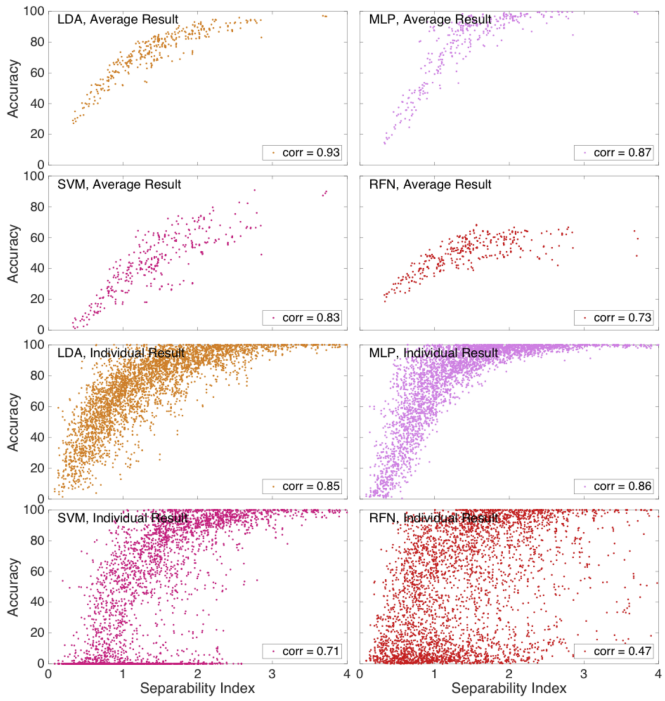


Fig. 3. Plots of accuracy against SI using individual features fed separately to the different classifiers. The top four show results that are average over all classes while the bottom four show result for classes individually. Correlation was calculated using Spearmans rho.

B. Nearest Neighbor Separability

The consequence of increasing the parameter k is that the algorithm becomes more sensitive to overlapping classes but it takes more iterations to compute. See how higher k increases correlation with accuracy along with the relative computation time in Table II. SI with Modified Mahalanobis uses 5.2 % of the computation time used by NNS with $k = 20$.

Plots of NNS results for different classifiers are shown in Fig. 4. Again individual results for SVM and RFN are

TABLE II
NEAREST NEIGHBOR SEPARABILITY CORRELATION WITH ACCURACY AND COMPUTATION TIME

	Individual Results	Average Results	Relative Time
k = 20	0.84	0.87	1
k = 40	0.86	0.88	1.37
k = 60	0.88	0.90	1.75
k = 80	0.89	0.91	2.14

Correlations, using Spearmans rho, between accuracy and NNS when using individual features fed separately to LDA and the values of k in the first column. Correlations under individual result are calculated using values for every class while average result is used for column three. Forth column shows the computation time relative to the fastest, k = 20.

widely scattered, but high correlation can still be found for all classifiers looking at the average result. However, NNS is spreading more for higher accuracies. This is especially clear in the average results for LDA where the plot is sun fan shape above 60 % accuracy and clustered around a line otherwise.

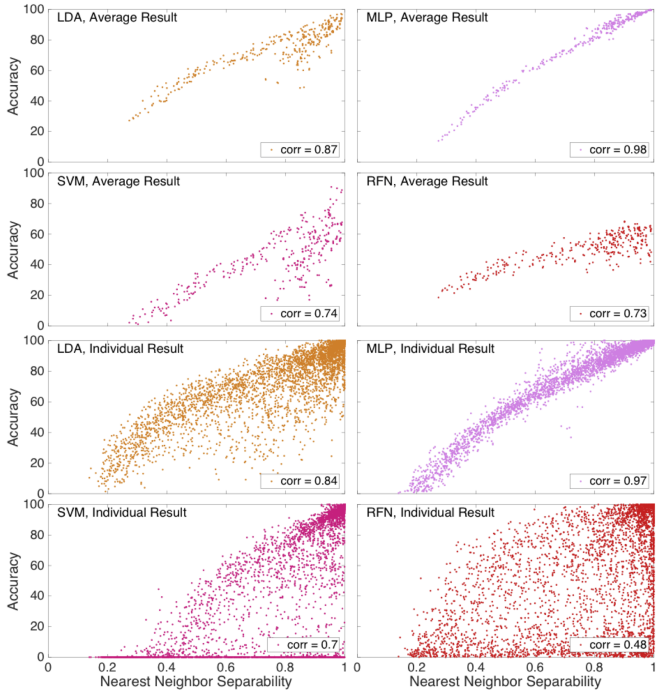


Fig. 4. Plots of accuracy against NNS using individual features fed separately to the different classifiers. The top four shows results that are average over all classes while the bottom four shows result for classes individually. Correlation was calculated using Spearmans rho.

C. Feature Sets

The classification accuracy from all classifiers when fed by the *best* and *reference sets* was used as evaluation method for performance. These results are illustrated in Fig. 5. Statistical significance is indicated by *. Correlation between CCEs and accuracies from using the *best* and *reference sets* are show in Fig. 6. Comparing Fig. 6 with Fig. 3 and 4 the results for LDA and MLP are similar except that accuracies are generally higher for feature sets over individual features. The SVM results are

less scattered and neither RFN or SVM results are clustered at zero accuracy as in Fig. 3 and 4.

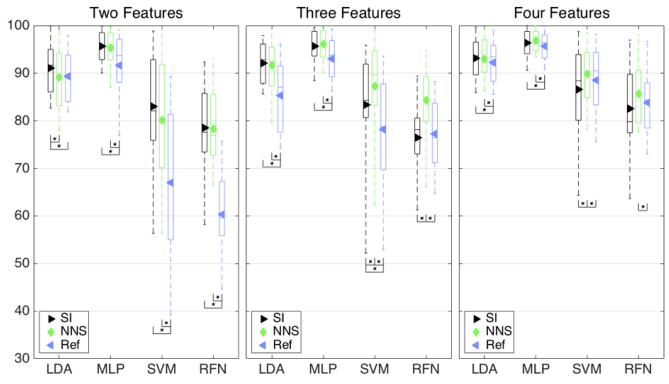


Fig. 5. Accuracies using the *best sets* from SI and NNS compared with the *reference sets*. The center line in the box is the median value, the marker is the mean value and the top and bottom are 25th and 75th percentiles respectively. The total data range is shown by the whiskers. Statistical significance using significance level 5 % is marked with *.

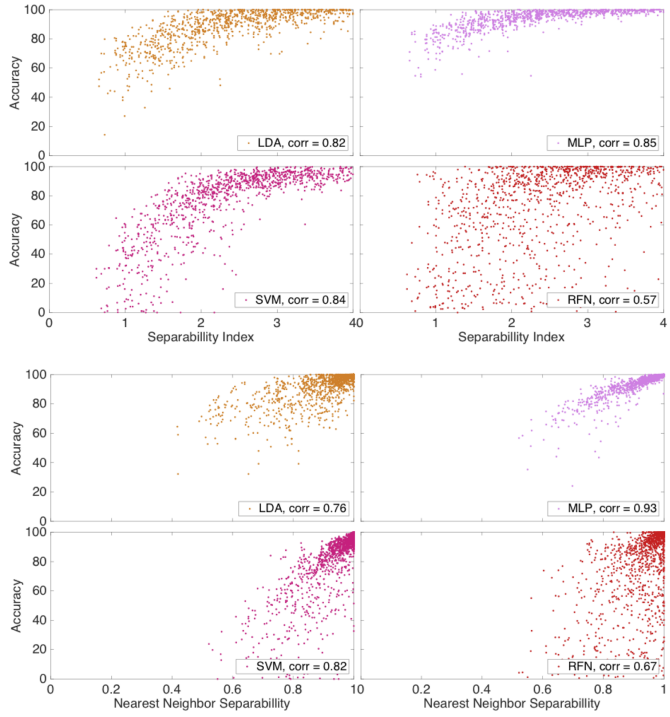


Fig. 6. Accuracy plotted against SI (top four insets) and against NNS (bottom four insets). The *best* and *reference sets* were used separately to extract feature vectors for all subject. One feature vector was fed to the classifier to create one point. Correlation was calculated using Spearmans rho.

D. Features

The rating of features and feature sets in this study has provided information on the features general performance. Fig. 7 shows the five features with highest and lowest average accuracy when using individual features. LDA and MLP results are represented. The five most selected features for the *best sets* are *tcard, fmn, tvar, tstd* and *tpwr* for SI and *tcard, tdam, tstd, twl* and *fmn* for NNS. The features are ranked in the order they are written, with the most selected feature first. It is worthy of notice that one of the top feature, cardinality, was recently found to be a highly performing feature in MPR [19].

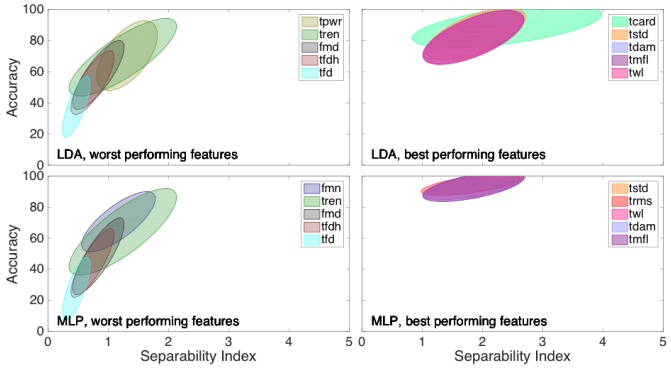


Fig. 7. Ellipses representing feature scatters of Accuracy against SI plots. The ellipses are drawn to represent the covariance matrix. The left insets show the features with the lowest average accuracy and the right insets show the features with the highest.

IV. Discussion

A. Accuracy Prediction

This study shows that NNS and SI can provide useful information when predicting accuracy for MLP and LDA based on the high correlation between classification accuracies and the CCEs, for both individual features and feature sets. This was also supported by the statistically significant improvement on MPR accuracy of the *best sets* over the *reference sets*. The only exception found was the *best set* of two features selected by NNS and fed to LDA which yielded lower accuracy than the *reference set*. However, the fact that these sets perform similarly is an indication that NNS provides relevant information for the problem at hand.

The SVM accuracy had low correlation with SI and NNS, compared to LDA and MLP, for individual features. However, correlation improved when feature sets were used, and the *best sets* are all yielded higher accuracy than the *reference sets*. How relevant the CCEs are for SVM seems to change with the number of inputs to the classifier.

All CCEAs evaluated in this study result in low correlation with RFN accuracy. However, all the *best sets* from NNS yielded significantly higher accuracy than the *reference sets*, which supports the use of this method for feature selection.

B. Consistency Over Change in Number of Dimensions

Consistency of SI and NNS for MLP and LDA can be appreciated by comparing the corresponding plots for individual features and feature sets. The clusters are forming similar patterns even though different number of dimensions were used. This is not true for SVM and RFN, where many of the individual results from individual features has accuracy close to zero. This makes these plots inconsistent with the corresponding patterns found for feature sets.

C. Limitations and Future Challenges

The CCEAs correlate differently with different classifiers accuracies. One reason is that neither SI or NNS describe the limitation of LDA due to its dependency on linearly separable classes. MLP is a nonlinear classifier and its accuracy is consequently more accurately estimated by the two CCEAs. Another reason is that SI assumes normality of the feature distributions while NNS does not. On the other hand, SI is much less computationally demanding. How PRAs and CCEAs are combined must be considered as CCEAs are implemented in more specific applications.

The data set for this study is limited to individual movements, four EMG channels and only offline accuracy was considered. To really evaluate how relevant the CCEAs are for MPR, further tests are needed with more diverse data sets.

This study shows that the information given by CCEAs can be used in feature selection for MPR. However, the feature sets in this study had maximum four features and there are only four recorded channels. The low number of dimensions makes brute-force search possible but the complexity will increase rapidly as the number of channels, features and/or classes increase. MFL and mRMR are two feature selection strategies already used for MPR, and are both examples of what is likely to be an important part of MPR in the future. The CCEAs in this study will have to be used strategically in a similar way to be efficient for feature selection.

Estimating accuracy without implementing a classifier allows early evaluation of recorded data. Classes with low separability can be recorded and evaluated again until desired separability can be confirmed. More research needs to be done on how to interpret CCEs when they are used for MPR.

V. Conclusions

Two Classification Complexity Estimating Algorithms, namely Separability Index and Nearest Neighbor Separability, were found adequate for MLP and LDA based on their high correlation with classification accuracy for both individual features and feature sets. High correlation with SVM accuracy was also observed for feature sets.

The two CCEAs were also found useful for EMG feature selection for all three aforementioned classifiers as feature sets selected based on level of separability given by the CCEAs resulted in higher or similar accuracies when compared with *reference sets*.

Acknowledgments

This study is supported by VINNOVA, Promobilia, and MedTech West.

References

[1] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use," *Journal of rehabilitation research and development*, vol. 48, no. 6, pp. 643–660, 2011.

[2] M. Ortiz-Catalan, N. Sander, M. B. Kristoffersen, B. Håkansson, and R. Brånemark, "Treatment of phantom limb pain (PLP) based on augmented reality and gaming controlled by myoelectric pattern recognition: A case study of a chronic PLP patient," *Frontiers in Neuroscience*, vol. 8, no. 8 FEB, pp. 1–7, 2014.

[3] X. Zhang and P. Zhou, "High-Density Myoelectric Pattern Recognition Toward Improved Stroke Rehabilitation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1649–57, 2012. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6172561 http://www.ncbi.nlm.nih.gov/pubmed/22453603

[4] D. Tkach, H. Huang, and T. a. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition." *Journal of neuroengineering and rehabilitation*, vol. 7, p. 21, 2010.

[5] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: detection, processing, classification and applications," *Biological Procedures Online*, vol. 8, no. 1, pp. 11–35, 2006. [Online]. Available: http://www.springerlink.com/index/10.1251/bpo115

[6] N. Bu, O. Fukuda, and T. Tsuji, *A Recurrent Probabilistic Neural Network for EMG Pattern Recognition*. IGI Global, 2006, pp. 130–153.

[7] M. Ortiz-Catalan, R. Brånemark, and B. Håkansson, "Biologically inspired algorithms applied to prosthetic control," *The IASTED International Conference, Biomedical Engineering,*, no. BioMed, pp. 7–15, 2012.

[8] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2012.01.102

[9] J. Liu, X. Li, G. Li, and P. Zhou, "EMG feature assessment for myoelectric pattern recognition and channel selection: a study with incomplete spinal cord injury." *Medical engineering & physics*, vol. 36, no. 7, pp. 975–80, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350453314000988

[10] H. C. Peng, F. H. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[11] C. Qiang, Z. Hongbo, and C. Jie, "The Fisher-Markov Selector: Fast Selecting Maximally Separable Feature Subset for Multiclass Classification with Applications to High-Dimensional Data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 6, pp. 1217–1233, 2011.

[12] N. E. Bunderson and T. a. Kuiken, "Quantification of feature space changes with experience during electromyogram pattern recognition control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 3, pp. 239–246, 2012.

[13] S. Singh, "Multiresolution Estimates of Classification Complexity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1534–1539, 2003.

[14] G. Nagy and X. Zhang, *Simple Statistics for Complex Feature Spaces*. London: Springer London, 2006, pp. 173–195.

[15] S. Raudys, *Measures of Data and Classifier Complexity and the Training Sample Size*. London: Springer London, 2006, pp. 60–68.

[16] M. Ortiz-Catalan, R. Brånemark, and B. Håkansson, "BioPatRec: A modular research platform for the control of artificial limbs based on pattern recognition algorithms." *Source code for biology and medicine*, vol. 8, no. 1, p. 11, 2013.

[17] M. Ortiz-Catalan, "Biopatrec," Januari 2016. [Online]. Available: https://github.com/biopatrec/biopatrec

[18] B. Hudgins, P. Parker, and R. Scott, "A new strategy for multifunction myoelectric control," *Biomedical Engineering, IEEE Transactions on*, vol. 40, no. 1, pp. 82–94, Jan 1993.

[19] M. Ortiz-Catalan, "Cardinality as a Highly Descriptive Feature in Myoelectric Pattern Recognition for Decoding Motor Volition," *Frontiers in Neuroscience*, vol. 9, no. 416, 2015. [Online]. Available: http://www.frontiersin.org/neural technology/10.3389/fnins.2015.00416/abstract

Paper II

Electromyography Data Analysis For Myoelectric Pattern Recognition

Niclas Nilsson[1], Bo Håkansson[1], and Max Ortiz-Catalan[1,2,3]
1) Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden.
2) Centre for the Advanced Reconstruction of Extremities (C.A.R.E.), Sahlgrenska University Hospital, Gothenburg, Sweden.
3) Integrum AB, Mölndal, Sweden.

Abstract

Background

Limb prosthetics, exoskeletons, and neurorehabilitation devices can be intuitively controlled using myoelectric pattern recognition (MPR) to decoded the subject's intention of movement. In conventional MPR, descriptive electromyography (EMG) features representing the aimed movement are feed to a classification algorithm. The separability of the different movements in the feature space highly affects the classification complexity. Classification Complexity Estimating Algorithms (CCEAs) were studied in this work in order to improve feature selection, predict MPR performance, and inform on faulty data acquisition.

Methods

CCEAs such as Nearest Neighbor Separability (NNS), Purity, Repeatability Index, and Separability Index (SI) were evaluated based on their correlation with classification accuracy, and on the their suitability to produce highly performing feature sets. SI was evaluated using alternatives to the original distance definition, Mahalanobis Distance, including Bhattacharyyas Distance, Hellingers Distance, Kullback Leiblers Convergence and a modified version of Mahalanobis Distance. Three commonly used classifiers in MPR were used to compute classification accuracy (Linear Discriminant Analysis (LDA), Multi-Layer Perceptron (MLP), and Support Vector Machine (SVM)). The algorithms and analytic graphical user interfaces produced in this work are freely available in BioPatRec.

Results

NNS and SI were found to be highly correlated with classification accuracy (correlations up to 0.98 for both algorithms), as well as capable to yield highly descriptive feature sets. Additionally, the experiments revealed how the level of correlation between the inputs of the classifiers influence classification accuracy, and emphasizes the classifiers sensitivity to such redundancy. They also inform on how dimensionality of the classification task influences the outputs of the CCEAs.

Conclusions

This study deepens the understanding of the classification complexity in prediction of motor volition based on myoelectric information. In addition, it provides researcher with tools to analyze myoelectric recordings in order to improve classification performance.

Background

Decoding of motor volition via Myoelectric Pattern Recognition (MPR) has many clinical applications such as prosthetic control [1], phantom limb pain treatment [2], and rehabilitation after stroke [3]. Research on MPR has focused on classifiers [4], pre-processing algorithms [5], and electromyography (EMG) acquisition [6] among other factors that influence the classification outcome. Reaz *et al.* studied different attributes of EMG signals, such as signal-to-noise ratio, that decrease the complexity of MPR [7]. However, limited studies have been done on the complexity of the classification task itself.

Most algorithms for MPR, except for a few exceptions [8], use EMG features extracted from overlapping time windows as the classifier input. The resulting classification accuracy is therefore dependent on the features used to described the EMG signals. The performance of a variety of such features, and feature selection algorithms, have been studied previously [9], [10]. Two feature selecting algorithms, namely Minimum Redundancy and Maximum Relevance (mRMR) [11], and Markov Random Fields (MRF) [12], were applied to an electrode array by Liu *et al.* [13], who used Kullback-Leibler Divergence and Feature Scatter to rate the relevance and redundancy of features. The features were then ranked and selected into sets according to these ratings. Similarly, Bunderson *et al.* defined three data quality indices, namely Repeatability Index (RI), Mean Semi-principal Axis (MSA), and Separability Index (SI) to evaluate the changes in data quality over repeated recordings of EMG [14]. Classification complexity estimation was not investigated in the aforementioned studies, but suggestions were made on such task could be achieved.

Classification complexity has been studied outside the field of MPR. Singh suggested two nonparametric multiresolution complexity measures, namely Nearest Neighbor Separability (NNS) and Purity [15]. These complexity measures were compared with common statistical similarity measures, such as Kullback-Leibler Divergence, Bhattacharyya Distance, and Mahalanobis Distance, and were found to yield a higher correlation with classification accuracy. These Classification Complexity Estimating Algorithms (CCEAs) along with Hellinger Distance were investigated in the present study with focus on their relevance for MPR.

In the present study, CCEAs were evaluated based on their correlation with offline classification accuracy and real-time classification performance. As a consequence, different attributes were reveled about the CCEAs, classification algorithms, and features descriptiveness. Two of such attributes, namely channel correlation dependency and dimensionality dependency, were investigated further. The CCEAs which were found to yield high correlation with classification accuracy (NNS and SI) were then used for feature selection and benchmarked against features sets found in the literature.

The result of these experiments provided evidence on the suitability of CCEAs to predict MPR performance. The algorithms used in this work were implemented and made freely available in BioPatRec, an open source platform for development and benchmarking of algorithms used in advanced myoelectric control [16], [17].

Methods

Data Sets

Two data sets were used in this study; both recorded on healthy subjects. The first set contained individual movements (*IM* data): 20 subjects, 4 EMG channels, 14 bits Analog to Digital Conversion (ADC), and 11 classes (hand open/close, wrist flexion/extension, pro/supination, side grip, fine grip, agree or thumb up, pointer or index extension, and rest or no movement) [16]. The second set contained individual and simultaneous movements (*SM data*): 17 subjects, 8 EMG channels, 16 bits ADC, and 27 classes (hand open/close, wrist flexion/extension, pro/supination, and all their possible combinations) [18]. Disposable Ag/AgCl (Ø = 1 cm) electrodes in a bipolar configuration (2 cm inter-electrode distance) were used in both sets. The bipoles were evenly spaced around the most proximal third of the forearm, with the first channel placed along the extensor carpi ulnaris. The data sets along with details on demographics and acquisition hardware are available online as part of BioPatRec [17]. Table 1 summarizes these data sets.

Table 1 - Summery of data sets

Table summering the data sets used in the experiments of this study. The reference column contains the name used when referring to that data set through out the report.

Reference	Movements	Subjects	Channels	ADC (bits)	Classes
IM data	Individual	20	4	14	11
SM data	Simultaneous	17	8	16	27

Signal Acquisition, Pre-processing and Feature extraction

BioPatRec recording routines guided the subjects to perform each movement three times with resting periods in between. The instructed contraction time, as well as the resting time was three seconds. The initial and final 15% of each contraction was discarded as this normally corresponds to delayed response and anticipatory relaxation by the subject, while the remaining central 70% still preserves portions of the dynamic contraction [16].

Time windows of 200 ms were extracted from the concatenated contraction data using 50 ms time increment. Features were then extracted from each time window and distributed in sets used for training (40 %), validation (20 %), and testing (40 %) of the classifiers. The testing sets were never seen by the classifier during training or validation. A 10-fold cross-validation was performed by randomizing the feature vectors between the three sets before training and testing.

The following EMG signal features were used as implemented in BioPatRec [16], [17], [19]. In time domain; mean absolute value (tmabs), standard deviation (tstd), variance (tvar), waveform length (twl), RMS (trms), zero-crossing (tzc), slope sign changes (tslpch), power (tpwr), difference abs. mean (tdam), max fractal length (tmfl), fractal dimension Higuchi (tfdh), fractal dimension (tfd), cardinality (tcard), and rough entropy (tren). In frequency domain; waveform length (fwl), mean (fmn) and median (fmd).

Classification Complexity Estimating Algorithms

The Classification Complexity Estimating Algorithms (CCEAs) were designed to return Classification Complexity Estimates (CCEs) for each movement separately (*individual result*), and averaged over all movements (*average results*). The CCEAs used were:

Separability Index

Separability Index (SI) was implemented as introduced by Bunderson *et al., i.e.* the average of the distances between all movements and their most conflicting neighbor [14]. Figure 1A

illustrates the distance and conflict between two classes in an exemplary two dimensional feature space.

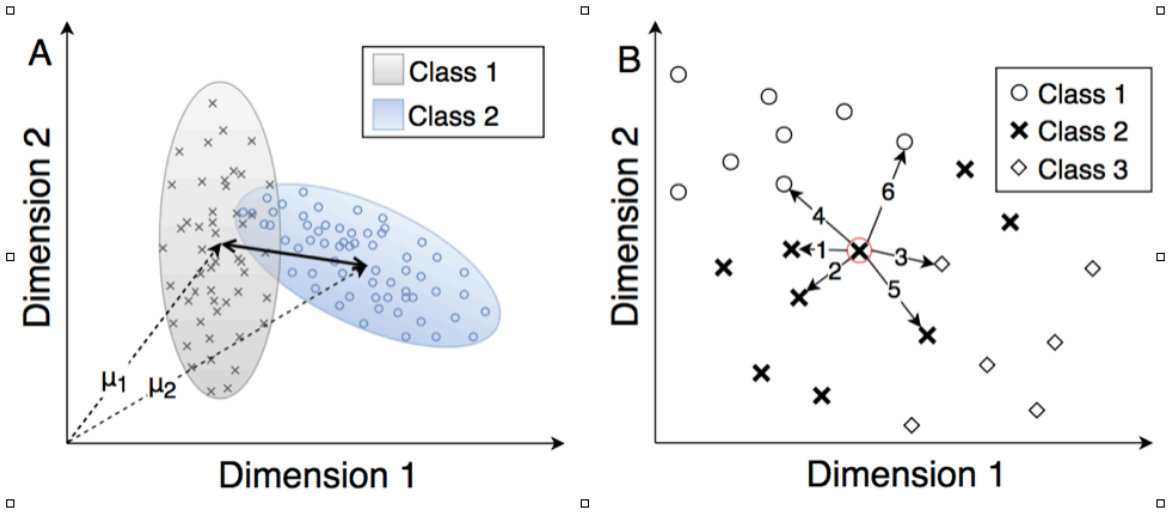


Figure 1 Illustration of a two dimensional feature space.

Inset A shows the distance between two classes in a two dimensional features space. The ellipses representing the classes are constructed according to the covariance of the two dimensional data. The figure emphasizes the overlap of classes, which is a big challenge in pattern recognition. Inset B shows the 6 nearest neighbors of the marked target data point. Nearest Neighbor Separability is based on the fraction of the neighbors from the same class as the target point.

The aforementioned distance was defined by Bunderson *et al.* to be half the *Mahalanobis Distance*, resulting in the following equation:

$$SI = \sum_{i=1}^{K} \left(\min_{j=1,\ldots,i-1,i+1,\ldots,K} \frac{1}{2} \sqrt{(\mu_i - \mu_j)^T S_i^{-1} (\mu_i - \mu_j)} \right)$$

where K is the number of classes or movements, and μ_x and S_x are mean vectors and covariance matrices for class x, respectively.

This definition only considers the covariance of the target movement (S_i), and not that of the comparing movement (i.e. S_j). We considered this particular formulation as a potential limitation, and therefore we introduced additional *Distance Definition*s. The *Distance Definitions* were used under the assumption of normality as *Mahalanobis Distance* was defined under the same assumption [20]. The introduced *Distance Definitions* are described in Table 2**.**

Table 2 - Distance Definitions for Separability Index

Table of Distance Definitions used to compute Separability Index including their names, definitions and how the were implemented in the present study.

Distance Definition	**Description**
Mahalanobis Distance	*Mahalanobis Distance* was designed to measure the distance between a distribution and a single point [20]. Half the *Mahalanobis Distance* will be the value referred to as *Mahalanobis Distance* hereafter because that is how it was originally used in SI [14]. *Mahalanobis Distance* for multivariate normal distributions is defined as: $$\frac{D_M}{2} = \frac{1}{2}\sqrt{(\mu_1 - \mu_2)^T S_1^{-1}(\mu_1 - \mu_2)}$$
Bhattacharyyas Distance	*Bhattacharyyas Distance* is a measurement of statistical similarity between two distributions based on the *Bhattacharyyas Coefficient* (BC) [21]. Unlike *Mahalanobis Distance*, *Bhattacharyyas Distance* take both the distance and similarity in covariance between the distributions into account. In this study the square root of *Bhattacharyyas Distance* was used to equate the formulation of *Mahalanobis Distance* and facilitate comparison. *Bhattacharyyas Coefficient* for the continues probability distributions p and q is defined as: $$BC = \int \sqrt{p(x)q(x)}dx$$ *Bhattacharyyas Distance* as function of *Bhattacharyyas Coefficient:* $$\sqrt{D_B} = \sqrt{-\frac{1}{2}\ln(BC)}$$ *Bhattacharyyas Distance* for multivariate normal distributions (square root) [22]: $$\sqrt{D_B} = \sqrt{\frac{1}{8}(\mu_1 - \mu_2)^T S^{-1}(\mu_1 - \mu_2) - \frac{1}{2}\ln\left(\frac{detS}{\sqrt{detS_1 detS_2}}\right)}$$
Kullback-Leibler Divergence	*Kullback-Leibler Divergence* is a well-known statistical similarity measure that is typically used to determine if an observed distribution, *Q*, is a sample of a true distribution, *P* [23]. *Kullback-Leibler Divergence* for multivariate normal distributions is defined as [22]: $$D_{KL} = \frac{1}{2}\left(tr(S_1^{-1}S_2) + (\mu_1 - \mu_2)^T S_1^{-1}(\mu_1 - \mu_2) - k + ln\left(\frac{detS_1}{detS_2}\right)\right)$$
Hellingers Distance	*Hellingers Distance* is related to *Bhattacharyyas Distance* as it is also based on the *Bhattacharyyas Coefficient* [24]. The square of *Hellingers*

	Distance was used in this study to avoid complex numbers appearing where the assumption of normality fails, and this will be the value referred to as *Hellingers Distance* here. *Hellingers Distance* as function of *Bhattacharyyas Coefficient* is defined as: $$D_H^2 = 1 - BC$$ *Hellingers Distance* for multivariate normal distributions: $$D_H^2 = 1 - \frac{(detS_1)^{\frac{1}{4}}(detS_2)^{\frac{1}{4}}}{(detS_1)^{\frac{1}{2}}} * exp\left\{-\frac{1}{8}(\mu_1 - \mu_2)^T S^{-1}(\mu_1 - \mu_2)\right\}$$
Modified Mahalanobis	This measure of statistical similarity is equal to the aforementioned *Mahalanobis Distance*, except that it takes the covariance matrix of both distribution being compared into account. The algorithm is related to *Bhattacharyyas Distance* but is only focused on the distance between the distributions. This CCEA is here referred to as *Modified Mahalanobis* and is defined for multivariate normal distributions as: $$\frac{D_{MM}}{2} = \frac{1}{2}\sqrt{(\mu_1 - \mu_2)^T S^{-1}(\mu_1 - \mu_2)}$$
Explanations	For all equations above index 1 and 2 are appointed the considered movement and the compared movement respectively, and $$S = \frac{S_1 + S_2}{2}$$

Nearest Neighboor Separability

Nearest Neighbor Separability (NNS) was inspired by the algorithm with the same name defined by Singh [15]. It is based on the dominance of nearest neighbors, in feature space, belonging to the same class (movement) as a target data point. The contribution of the nearest neighbors are weighted by their proximity to the target point and the result is normalized to be values between 0-1. Let

$$b(p_t, p_i) = \begin{cases} 1, & if\ p_t, p_i \in C \\ 0, & if\ p_t \in C, p_i \notin C \end{cases}$$

Where p_t is the target point, p_i is p_t:s *i-th* nearest neighbor and C is a class. The aforementioned dominance is then defined as:

$$d_t = \left(\sum_{i=1}^{k} \frac{1}{i}\right)^{-1} \sum_{i=1}^{k} \frac{b(p_t, p_i)}{i}$$

A target point and its 6 nearest neighbors are illustrated in Figure 1B.
The end result is the average dominance:

$$NNS = \frac{1}{N}\sum_{i=1}^{N} d_i$$

Where N is the total number of samples.
Unless stated otherwise, the parameter k is set to 120, which is the maximum number of nearest neighbors from the same class for the data sets of this study.

Purity

Purity was computed by dividing the feature hyperspace into smaller hyper cuboids called cells [15]. The cells were rated individually and high dominance of one class in one cell meant high Purity for that cell. The final Purity of a data set was the average over all cells and different cell resolutions.

Repeatability Index

Repeatability Index (RI) measured how much individual classes varies between different occurrences using *Mahalanobis Distance* [14]. The three repetitions during the recording session were the occurrences that were evaluated. The end result is the average *Mahalanobis Distance* between the first repetition and the following ones for all movements.

Classifiers and Topologies

Three common classifiers for MPR were used in this study, Linear Discriminant Analysis (LDA), Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM). A quadratic kernel function was used for SVM. The classifiers were utilized as implemented in BioPatRec [16] (code available online [17]), where LDA and SVM were implemented using Matlab's statistic toolbox. MLP and SVM are inherently capable of simultaneous classification when provided with the feature vectors of mixed (simultaneous) outputs, hereafter referred as "MIX" output configuration, *i.e.* there is one output for every individual movement and combinations of movements produce the corresponding mix of outputs to be turned on. LDA's output is computed by majority voting and therefore it cannot produce simultaneous classification by creating a mixed output. However, classifiers like LDA can still be used for simultaneous classification using the *label power set* strategy where the classifier is constructed having the same number of outputs as the total number of classes. This configuration is referred here as "All Movements as Individual" (AMI). Ortiz-Catalan *et al.* showed that AMI could also favor classifiers capable of mixed outputs [18], and therefore MLP and SVM were evaluated in both MIX and AMI configurations for simultaneous predictions. In addition, LDA was also used in the One-Vs-One topology (OVO), as this has been shown to improve classification accuracy for individual movements [18], [25].

Evaluation and Comparison

In order to evaluate the correlation between Classification Complexity Estimates (CCEs) and classification accuracy, all features were used individually to classify all movements from each subject in both data sets, which provided a wide range of classification accuracies and their related CCEs. Correlations were then calculated considering the classification of each movements individually (*individual results*), or the average over all movements (*average results*).

In order to evaluate the influence of dimensionality change (*i.e.*, change in number of channels and/or features), the feature were used in sets of two or three features and then feed to each MPR algorithm. The sets were selected so that one member varied between all features (equivalent to when features were used individually) and the other one (or two) were the feature(s) resulting in the lowest classification accuracy when used individually. This was based on the assumption that the additional features had consequently low influence on the feature sets performance, and the resulting classification would be closer to that of using individual features.

The CCEAs were further used to select sets of two, three, and four features by brute force search through all possible combinations based on their CCEs. The selected sets are referred hereafter as the *best sets* and were obtained using the *IM data* set.

Ortiz-Catalan *et al.* used a genetic algorithm to find optimal feature sets of two, three, and four features based on classification performance [9]. Their proposed sets of two and three features were used as benchmarking sets in this study along with the commonly used four-feature set proposed by Hudgins *et al.* [26]. These sets are referred in this study as *reference sets*:

- **Ref 2:** tstd, trms [9]
- **Ref 3:** tstd, fwl, fmd [9]
- **Ref 4:** tmabs, twl, tslpch, tzc [26]

The *best* and *reference sets* of equal number of features were compared to each other based on the resulting classification accuracy as given by the three different classifiers. Classification accuracy corresponds to offline computations unless otherwise stated. Real-time testing was done using the *Motion Tests* as implemented in BioPatRec [16], [27]. CCEAs proficiency on predicting real-time performance was evaluated by their correlation with the *Completion Time* obtained from *Motion Tests*, which is the time from the first prediction not equal to *rest* until 20 correct predictions are achieved. If the number of correct predictions were under 20 after 20 seconds, the *Completion Time* was set to 20 seconds. The real-time results were obtained from *IM data* set and related *Motion Test*s [27].

Wilcoxon signed-rank test ($p <= 0.05$) was used to evaluate statistical significant differences. Correlations were calculated using Spearmans rho, since there was no clear linearity in the dependencies between accuracy and CCE.

Results

Separability Index

The correlations found between classification accuracy and *Separability Index* using different *Distance Definitions* are summarized in Table 3, where the highest value for every classifier is highlighted.

Table 3 - Correlations for the different Distance Definitions

Correlations under "individual results" were calculated using classification accuracies and Separability Indices from every individual movement, subject and feature, while those under "average result" were derived using the average Separability Index and classification accuracy per subject and feature. Both methods provide one correlation, although "individual results" use more data. Classifiers were configured using "All-Movements-as-Individual" (AMI) or "Mixed outputs" (MIX). Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO). The highest correlation values per column are highlighted in bold. All correlations were found statistically significant at p<0.01. The MIX configuration is not applicable (NA) for individual movements since there is not mixed outputs.

	AVERAGE RESULT			**INDIVIDUAL RESULTS**			
	LDA (AMI) Single/OVO	**MLP AMI/MIX**	**SVM AMI/MIX**	**LDA (AMI) Single/OVO**	**MLP AMI/MIX**	**SVM AMI/MIX**	**Data set**
MAHALANOBIS	0.72/0.91	0.90/0.91	0.79/0.80	0.81/**0.92**	0.84/0.85	0.70/0.68	SM
	0.78/0.88	0.86/NA	0.71/NA	**0.85**/0.91	0.80/NA	0.60/NA	IM
BHATTACHARYYA	0.74/**0.97**	**0.98**/**0.97**	0.79/0.82	0.69/0.91	0.93/**0.91**	0.66/0.65	SM
	0.83/0.96	0.96/NA	0.68/NA	0.79/0.89	**0.94**/NA	0.68/NA	IM
KULLBACK-LEIBLER	0.60/0.88	0.93/0.90	0.65/0.70	0.54/0.76	0.84/0.82	0.63/0.60	SM
	0.51/0.72	0.80/NA	0.32/NA	0.65/0.75	0.87/NA	0.65/NA	IM
HELLINGER	0.68/0.94	**0.98**/0.96	0.75/0.77	0.69/0.90	0.93/**0.91**	0.66/0.65	SM
	0.80/0.95	0.97/NA	0.66/NA	0.79/0.89	**0.94**/NA	0.68/NA	IM
MODIFIED MAHALANOBIS	0.92/**0.97**	0.92/0.95	**0.94**/**0.95**	0.79/0.91	0.88/0.89	**0.74**/**0.71**	SM
	0.93/0.94	0.87/NA	0.83/NA	**0.85**/0.90	0.86/NA	0.71/NA	IM

Figure 2 and Figure 3 shows plots of *average result* for *IM* and *SM data* sets respectively with the most correlating *Distance Definition* highlighted for classifiers individually. Table 3, Figure 2 and Figure 3 indicates that the best *Distance Definitions* vary for each classifiers.

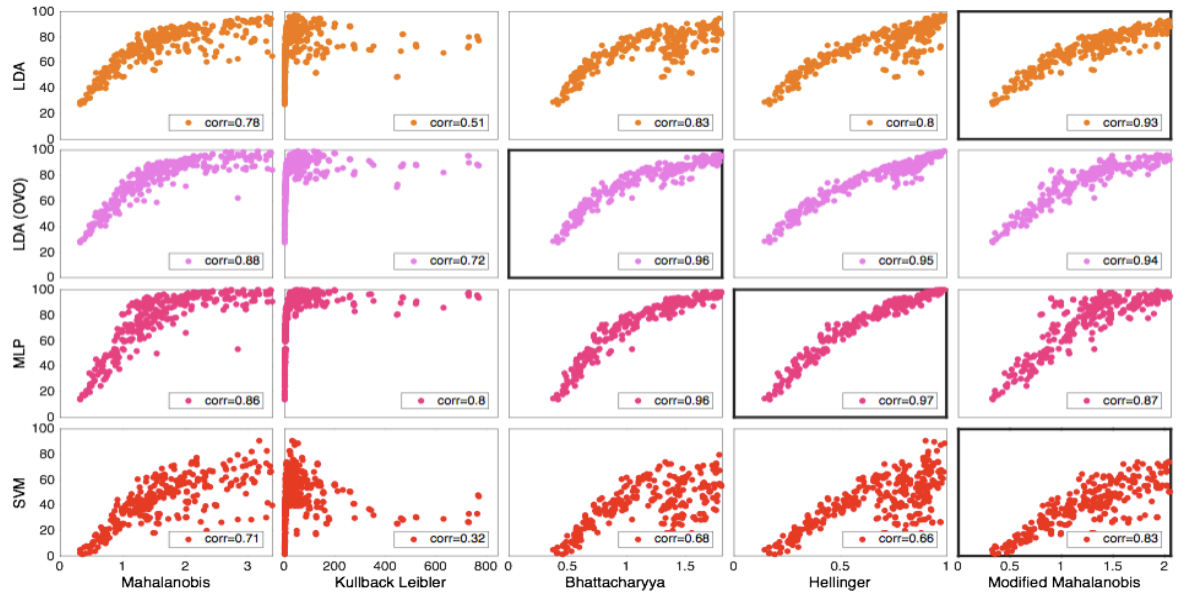


Figure 2 The distribution of Distance Definitions and classifiers data using individual movement (IM set)

Plot matrix where the insets show classification accuracy plotted against Separability Index for the individual movements data set. One marker represents the average over all movements for one subject and one feature. The classifiers are grouped in rows and the Distance Definitions for Separability Index are group in columns. Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO). All correlations were found statistically significant at p<0.01. Classifiers and Distance Definitions are stated at the left side and the bottom of the plot matrix respectively. Highest correlating Distance Definition for every classifier is marked by a thicker frame around the plot.

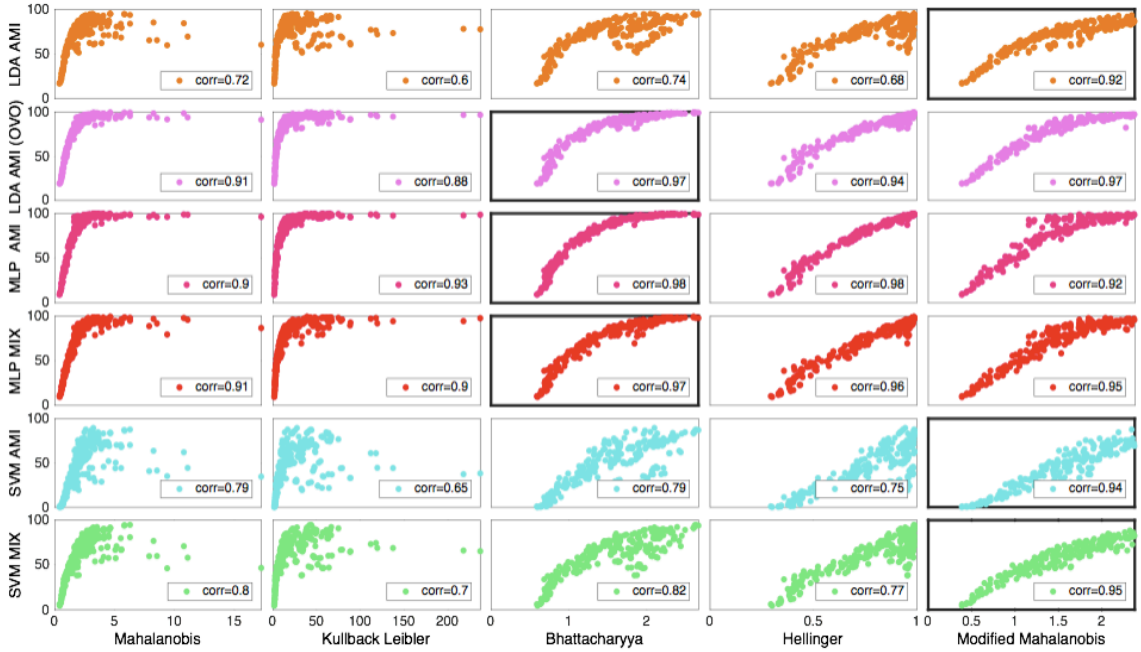


Figure 3 The distribution of Distance Definitions and classifiers data using simultaneous movements (SM set)

Plot matrix where the insets shows classification accuracy plotted against Separability Index for the simultaneous movements data set. One marker represents the average over all movements for one subject and one feature. The classifiers are grouped in rows and the Distance Definitions for Separability Index are group in columns. Classifiers were configured using "All-Movements-as-Individual" (AMI) or "Mixed outputs" (MIX). Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO). All correlations were found statistically

significant at p<0.01. Classifiers and Distance Definitions are stated at the left side and the bottom of the plot matrix respectively. Highest correlating Distance Definition for every classifier is marked by a thicker frame around the plot.

Mahalanobis Distance

Mahalanobis Distance was found as the *Distance Definition* that most correlated with LDA in an OVO topology for *individual results* using *SM data*. The corresponding classification accuracy against SI is plot in Figure 4A.

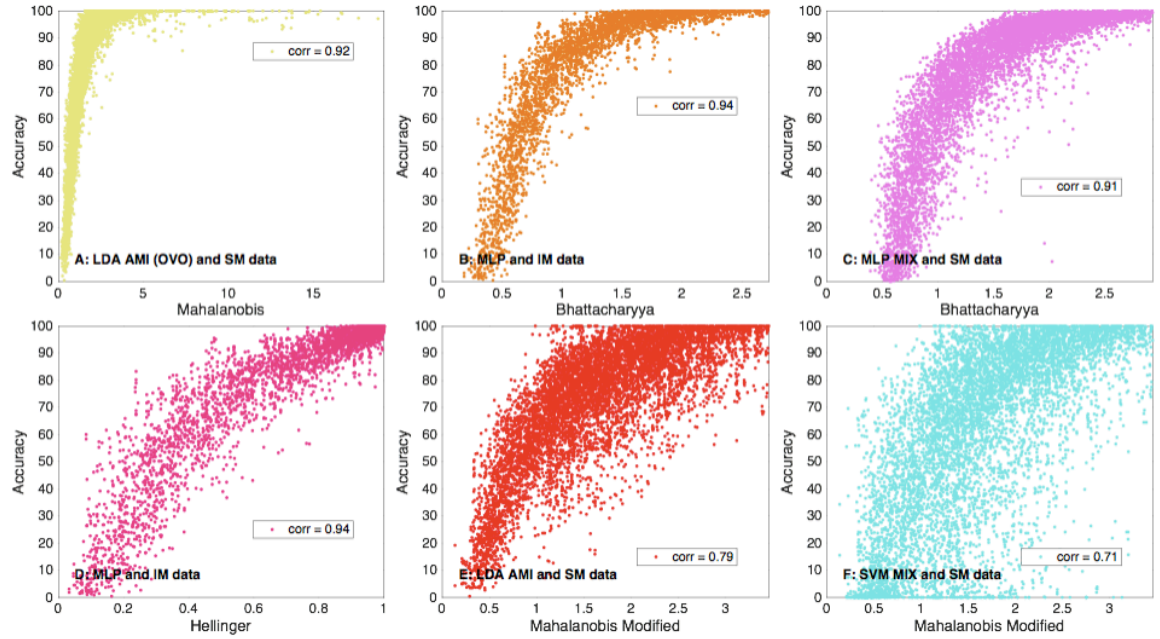


Figure 4 Data distribution for the most correlating Distance Definitions

Plot matrix where the insets show classification accuracy plotted against Separability Index. One dot represents one movement, one subject and one feature, making the number of dots the number of movements times the number of subjects times the number of features. The plots represent the highlighted correlations in Table 1.

Kullback-Leibler Divergence

Kullback-Leibler Divergence was not found to yield higher correlation than any other *Distance Definition* for any of the classifiers, however it was found most correlating with the *average results* of MLP using both topologies. This correlation is visualized in Figure 2 and Figure 3. Owing to its low correlation with classification accuracy, *Kullback-Leibler Divergence* was not used in the reaming experiments..

Bhattacharyyas Distance

Bhattacharyyas Distance was the most correlating *Distance Definition* for MLP in both *AIM* and *MIX* configurations. Plots of classification accuracy for the two classifiers against SI based on *Bhattacharyyas Distance* is shown in Figure 4 insert B and C. *Individual results* are presented and *IM data* and *SM data* are used for *AIM* and *MIX* configurations respectively.

Hellinger Distance

Bhattacharyyas Distance and *Hellingers Distance* are highly related as they are both based on the *Bhattacharyyas Coefficient*. Table 3 confirms their resemblance as the correlations related to the two *Distance Definitions* are very similar in all cases. Naturally, *Hellingers Distance* is together with *Bhattacharyyas Distance* as the *Distance Definitions* most correlating with MLP MIX and AMI for *individual result,* and with MLP *AMI* for *average result*. MLP *AMI* classification

accuracy is plotted against *Hellingers Distance* based SI in Figure 4E, where *individual results* using *IM data* is represented.

Modified Mahalanobis

Modified Mahalanobis was found as the *Distance Definition* most correlating with *average results* of LDA and SVM classification accuracy for all topologies and configurations. The same is true for *individual results,* except for LDA in an OVO topology. Figure 4 inset E and F show LDA *AMI* and SVM *MIX* classification accuracy plotted against SI based on *Modified Mahalanobis. Modified Mahalanobis* was the version of *Mahalanobis Distance* used in the remaining results because of its overall higher correlation with classification accuracy.

Nearest Neighbor Separabillity

A summary of correlations with all classifiers for both data sets is presented in Table 4.

Table 4 - Correlations between classification accuracy and Nearest Neighbor Separability

The correlation between classification accuracy and Nearest Neighbor Separability with different values of the parameter k. Correlations under "individual results" were calculated using classification accuracies and Nearest Neighbor Separabilities from every individual movement, subject and feature, while those under "average result" were derived using one average Nearest Neighbor Separability and classification accuracy for every subject and feature. Both methods provide one correlation, although "individual results" use more data. Classifiers were configured using "All-Movements-as-Individual" (AMI) or "Mixed outputs" (MIX). Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO). The highest correlation values per column are highlighted in bold. All correlations were found statistically significant at p<0.01. The MIX configuration is not applicable (NA) for individual movements since there is not mixed outputs.

	AVERAGE RESULT			**INDIVIDUAL RESULTS**			
	LDA (AMI) Single/OVO	**MLP AMI/MIX**	**SVM AMI/MIX**	**LDA (AMI) Single/OVO**	**MLP AMI/MIX**	**SVM AMI/MIX**	**Data set**
K = 20	0.86/**0.98**	0.96/**0.97**	0.90/0.90	0.83/0.93	0.92/0.92	0.72/0.72	SM
	0.86/0.97	0.98/NA	0.74/NA	0.84/0.92	**0.97**/NA	0.70/NA	IM
K = 120	**0.90**/0.97	0.92/0.95	**0.92**/**0.92**	0.87/0.90	0.87/**0.89**	**0.73**/0.73	SM
	0.90/**0.98**	**0.97**/NA	0.78/NA	**0.89**/**0.93**	0.94/NA	**0.73**/NA	IM

Table 4 also shows the influence of the parameter *k*. Figure 5 and Figure 6 shows plots of *average result* for the *IM* and *SM data* respectively.

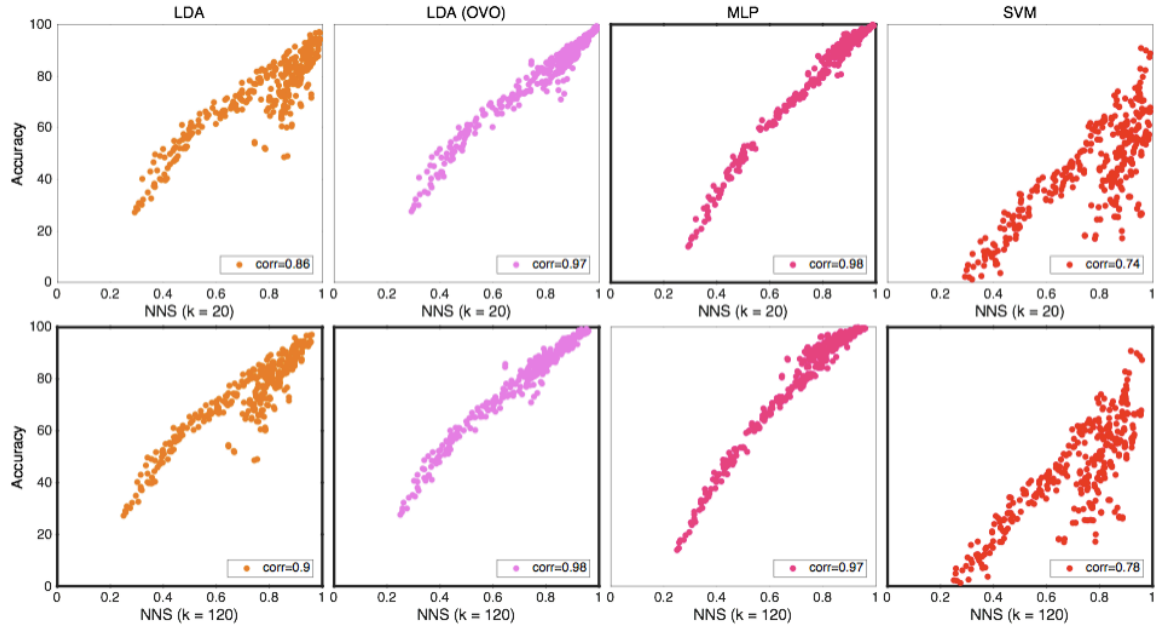


Figure 5 The distribution of data from individual movement for Nearest Neighbor Separability and all classifiers

Plot matrix where the insets shows classification accuracy plotted against Nearest Neighbor Separability for the individual movements data set. One marker represents the average over all movements for one subject and one feature. Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO). All correlations were found statistically significant at p<0.01. The classifiers are grouped in columns and the results for different values of the parameter k are group in rows. The highest correlation values per column are highlighted by a thicker frame.

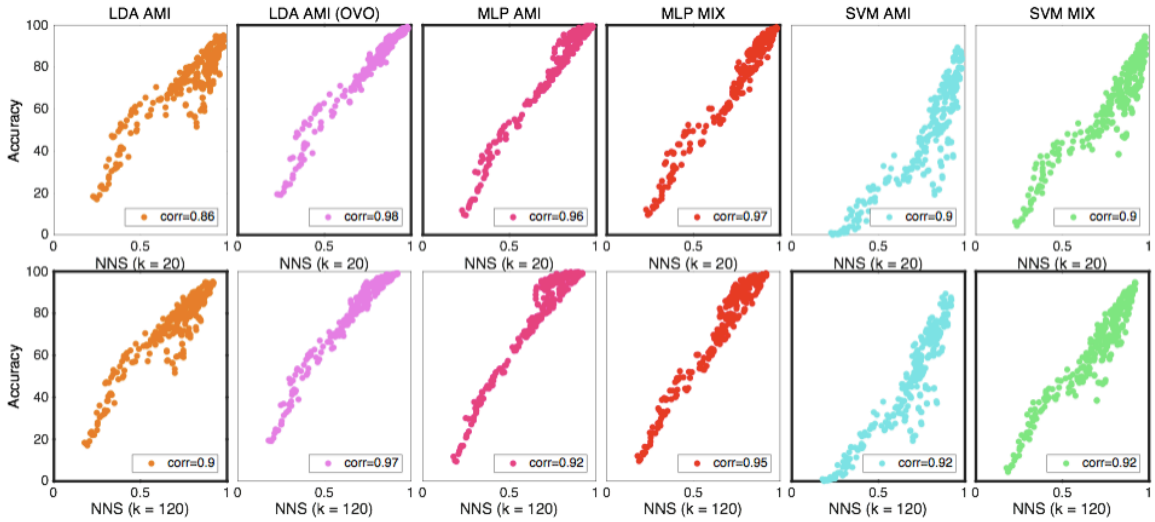


Figure 6 The distribution of data from simultaneous movement for Nearest Neighbor Separability and all classifiers

Plot matrix where the insets shows classification accuracy plotted against Nearest Neighbor Separability for the simultaneous movements data set. One marker represents the average over all movements for one subject and one feature. Classifiers were configured using "All-Movements-as-Individual" (AMI) or "Mixed outputs" (MIX). Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO). The highest correlation values per column are highlighted by a thicker frame. All correlations were found statistically significant at p<0.01. The classifiers are grouped in columns and the results for different values of the parameter k are group in rows.

NNS is most correlated with LDA in an OVO topology, which is equivalent to the results obtained by SI based on *Bhattacharyyas Distance* for the same classifier. The *individual results* for LDA using OVO are plotted for both data sets in Figure 7.

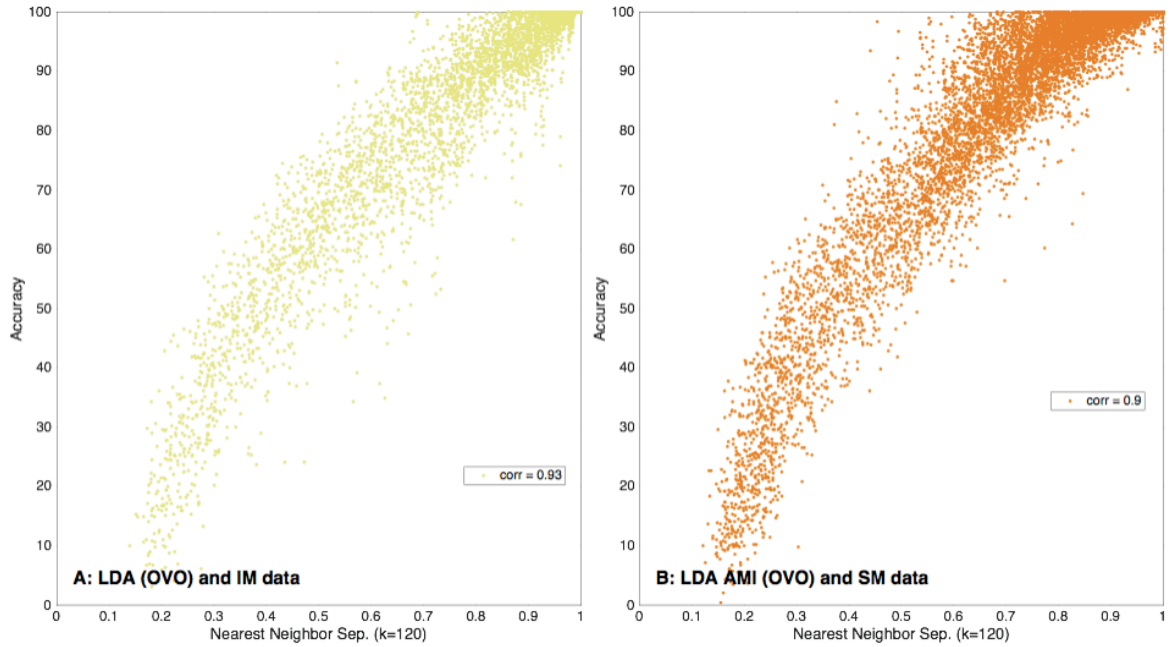


Figure 7 Highest correlation for Nearest Neighbor Separability

LDA (OVO) classification accuracy plotted against Nearest Neighbor Separability for individual result. One dot represents one movements, one subject and one feature, making the number of dots the number of movements times the number of subjects times the number of features. The plots illustrate the highest correlation from Table 2.

Purity and Repeatability Index

Purity and Repeatability Index resulted in low correlation with classification accuracy for all classifiers. The correlations for *IM data* can be found in
Table 5. Figure 8 shows *Individual results* of MLP for the two algorithms and the aforementioned data set. Because of the low correlation, Purity was excluded from the following experiments, and RI from the *Feature Sets* experiment.

Table 5 - Correlation for Purity and Repeatability Index regarding classification accuracy

The correlation with classification accuracy for Purity and Repeatability Index. Correlations under "individual results" were calculated using classification accuracies and Classification Complexity Estimates from every individual movement, subject and feature, while those under "average result" were derived using one average Classification Complexity Estimate and classification accuracy for every subject and feature. Both methods provide one correlation, although "individual results" use more data. Classifiers were configured using "All-Movements-as-Individual" (AMI). All correlations were found statistically significant (p<0.05).

	AVERAGE RESULT			INDIVIDUAL RESULTS			
	LDA (AMI) Single/OVO	MLP AMI	SVM AMI	LDA (AMI) Single/OVO	MLP AMI	SVM AMI	Data set
PURITY	0.31/0.0062	-0.14	0.51	0.3/0.15	0.14	0.54	IM
REPEATABILITY	0.64/0.8	0.85	0.57	0.23/0.36	0.45	0.16	IM

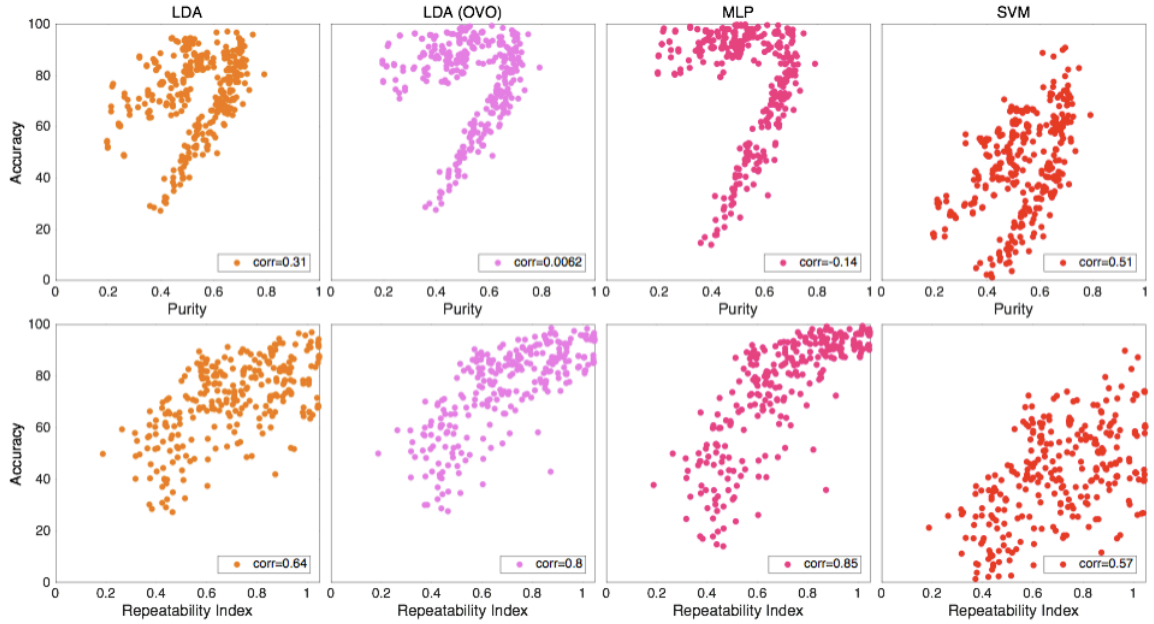


Figure 8 The distribution of data from individual movement for all classifiers with Purity and Repeatability

Plot matrix where the insets shows classification accuracy plotted against Purity for row one and Repeatablity for row two. The result is for the individual movements data set. One marker represents the average over all movements for one subject and one feature. Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO). The classifiers are grouped in columns.

Feature Sets

In this section the *best sets* are compared with each other and the *reference sets.* In Figure 9, the *best sets* corresponding to the *Distance Definitions* of SI are compared. The *Modified Mahalanobis* sets are significantly higher than the other *Distance Definitions* sets in 8 out of 12 cases, and averagely higher in all but the case where MLP is used with sets of three features. In that case *Bhattacharyyas Distance* and *Hellingers Distance* sets performing higher average classification accuracy.

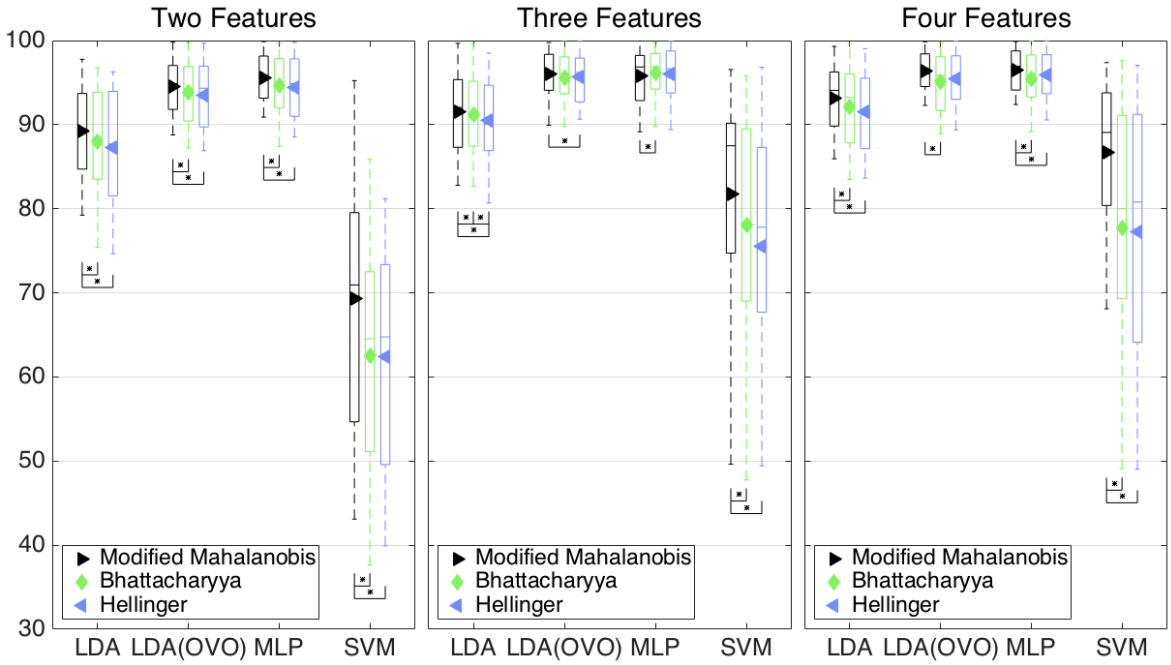


Figure 9 Classification accuracy for the best sets corresponding to Distance Definitions of Separability Index

Boxplot of average classification accuracy over all movements when using the best sets representing the Distance Definitions found in the legends. The middle line of the box is the median, the marker is the mean and the box extend to the 25 and the 75 percentile for the bottom and the top respectively. The different insets compare sets of different number of features. The result is derived from the IM data set. Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO).

The influence of parameter k of the NNS algorithm is shown in Figure 10 by comparing the *best sets* for k = 120 and k = 20. The higher value of k leads to higher average classification accuracy in all cases. However, it is statistical significant for SVM and three features only.

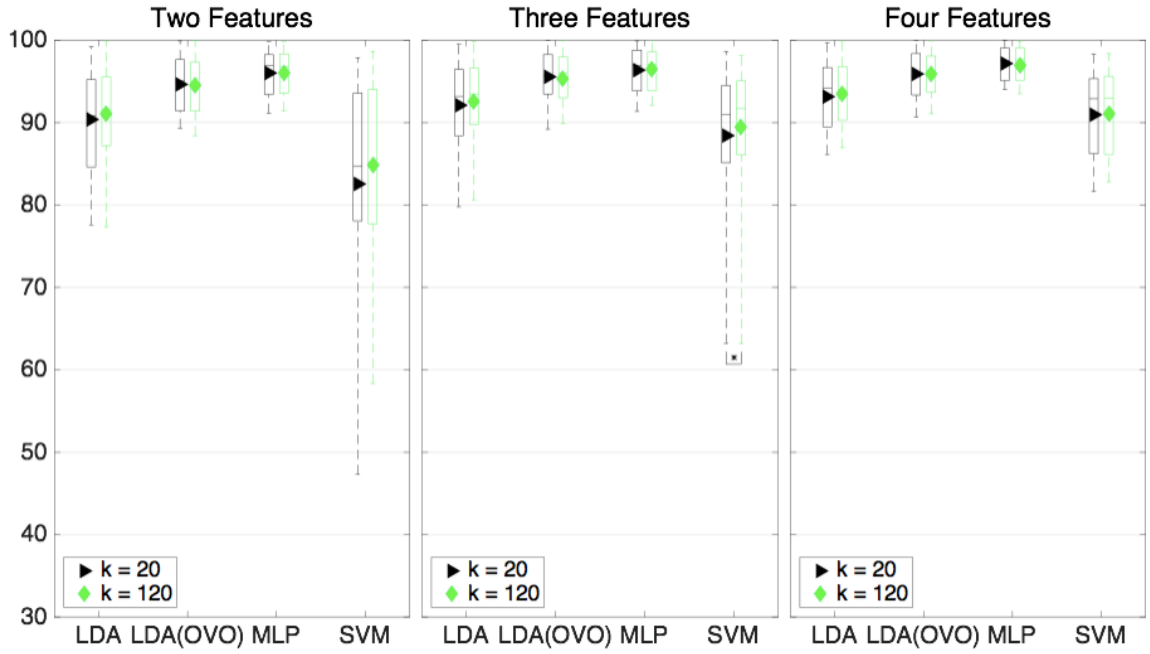


Figure 10 Classification accuracy for the best sets corresponding to Distance Definitions of Separability Index

Boxplot of average classification accuracy over all movements when using the best sets representing the Distance Definitions found in the legends. The middle line of the box is the median, the marker is the mean and the box extend to the 25 and the 75 percentile for the bottom and the top respectively. The different insets compare sets of different number of features. The result is derived from the IM data set. Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO).

The member with the highest average classification accuracy were selected from Figure 9 and Figure 10, *Modified Mahalanobis* and k = 120 respectively, to be compared with the reference sets in Figure 11. The NNS sets leads to significantly higher classification accuracy then the reference in all but one case, while *Modified Mahalanobis* is significantly higher for 9 out of 12. The average classification accuracy for the NNS sets are higher than *Modified Mahalanobis* for all classifiers but for LDA in an OVO topology, where *Modified Mahalanobis* is consistently higher.

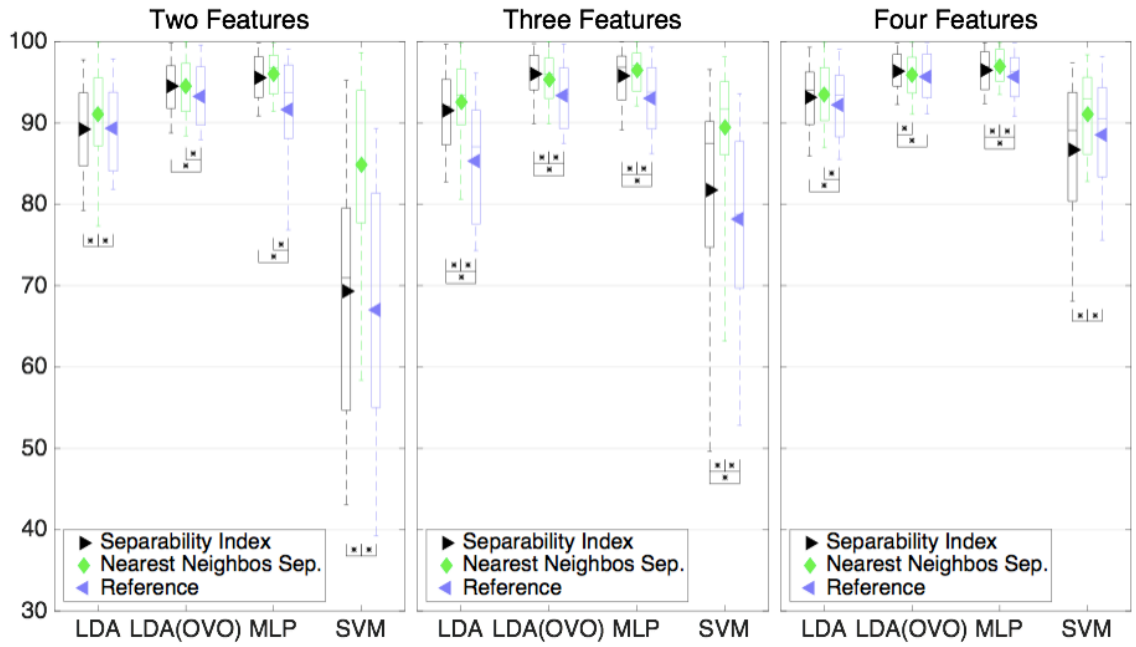


Figure 11 Classification accuracy for the best sets compared to the reference sets

Boxplot of average classification accuracy over all movements when using the best sets representing Separability Index with Modified Mahalanobis as Distance Definition, Nearest Neighbor Separability with k = 120 and the reference sets. The value of k is found in the legend. The middle line of the box is the median, the marker is the mean and the box extend to the 25 and the 75 percentile for the bottom and the top respectively. The different insets compare sets different number of feature algorithms. The result is derived from the IM data set. Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO).

Real-Time

Figure 12 summarizes the correlations between the *Motion Test* result *Completion Time* and CCEs corresponding to RI, NNS and SI based on *Modified Mahalanobis* and *Bhattacharyyas Distance*. Statistical significant correlations ($p < 0.001$) are highlighted by a darker frame.

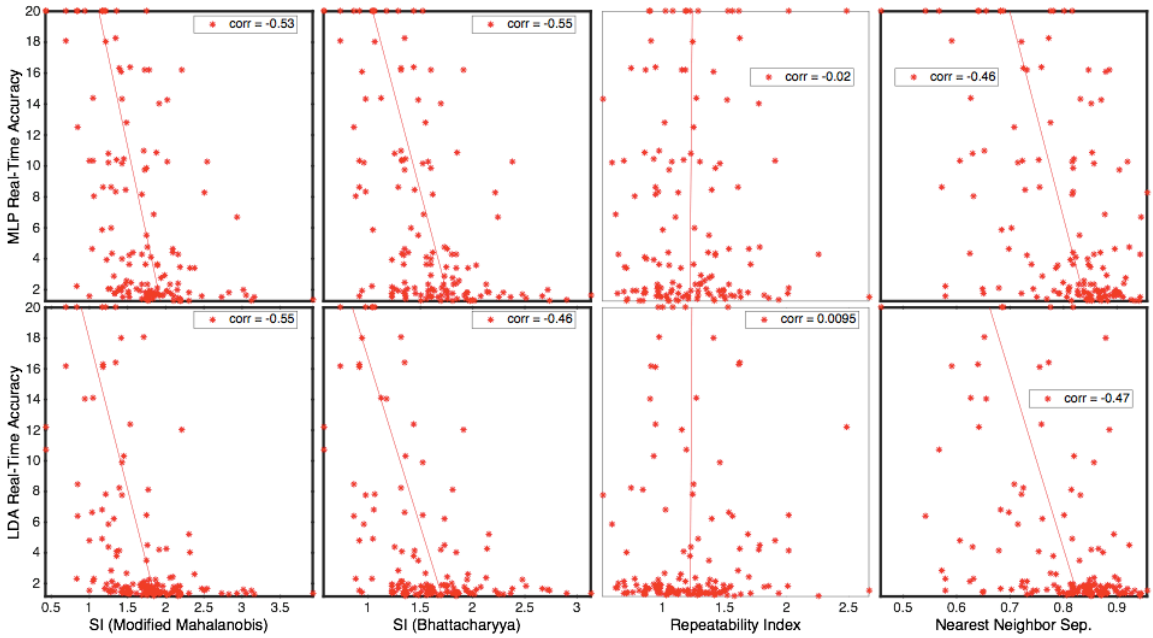


Figure 12 Real-time correlation for Classification Complexity Estimations

Plot matrix where the insets are Completion Time plotted against Classification Complexity Estimates. Significant correlation ($p < 0.001$) is highlighted with bold frame.

Feature Attributes

As the correlations used to evaluate the CCEAs was derived by use of one feature at the time, attributes of features individually were revealed. Examples of such attributes are average classification accuracy and classification accuracy variance. These two attributes are illustrated in Figure 13 and Figure 14 for *IM* and *SM data* respectively. Figure 13 shows the 5 features resulting in the highest and lowest average classification accuracy for classifiers separately.

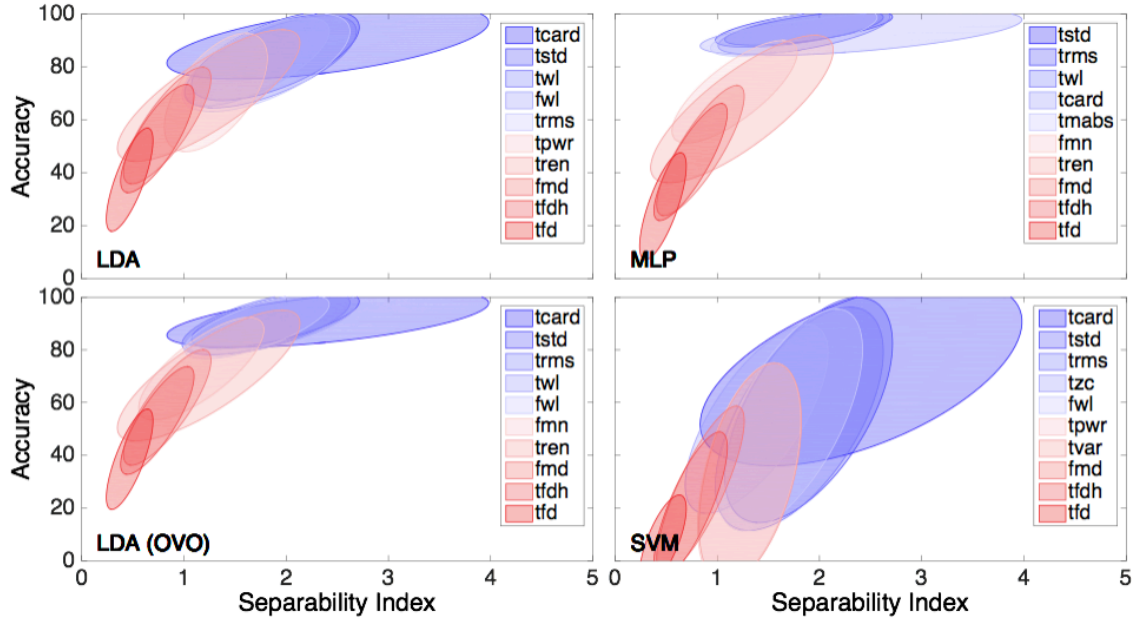


Figure 13 High and Low performing features for the Individual Movements data

Ellipses representing clusters for features in classification accuracy against Separability Index plots for results using the individual movements data set. The Separability Index Distance Definition is Modified Mahalanobis. The ellipses are centered around the means of the feature clusters and constructed according to there covariance matrix. Every inset includes the features with the top 5 and bottom 5 average classification accuracies for the classifier stated in the plot. The ellipses are color coded by red and blue color for low and high average classification accuracy respectively. Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO).

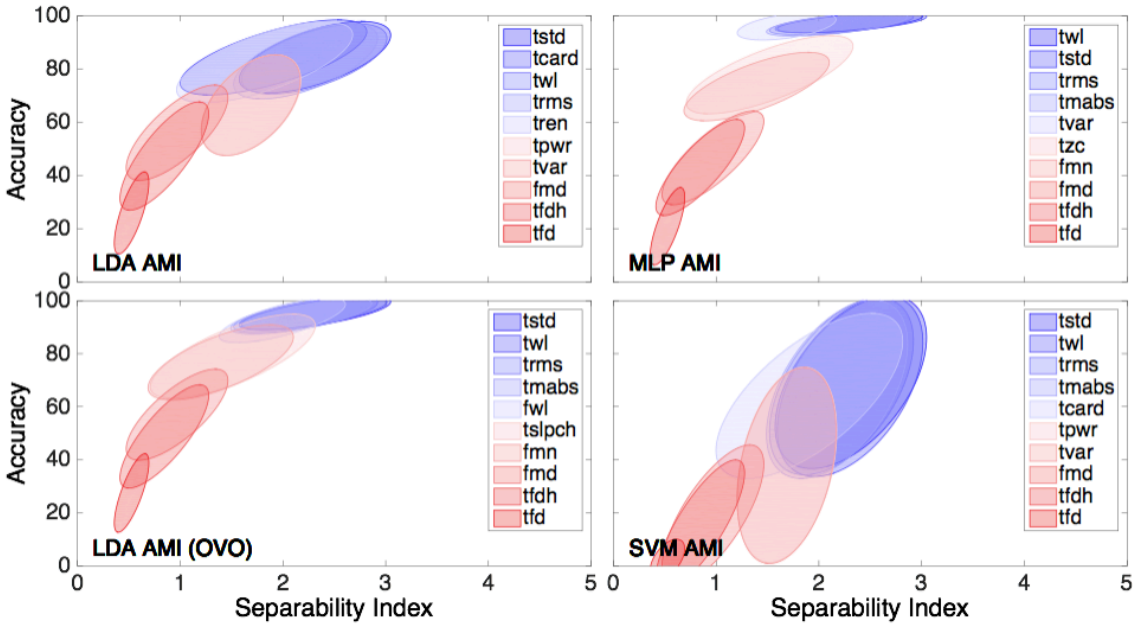


Figure 14 High and Low performing feature algorithms for the Simultaneous Movements data

Ellipses representing clusters for features in a classification accuracy against Separability Index plots for result using simultaneous movements data set. The Separability Index Distance Definition is Modified Mahalanobis. The ellipses are centered around the means of the feature clusters and constructed according to there covariance matrix. Every inset includes the feature algorithms with the top 5 and bottom 5 average classification accuracies for the classifier stated in the plot. The ellipses are color coded by red and blue color for low and high average classification accuracy respectively. Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO).

One attribute that was observed to highly influencing the CCEAs correlation with classification accuracy was channel correlation, *i.e.* correlation between feature sequences extracted from the channels separately using only the feature considered. To illustrate this attribute, average determinants of the channel correlation matrices over all subjects for the different features were extracted from *SM data* and shown in the bar diagram in Figure 15.

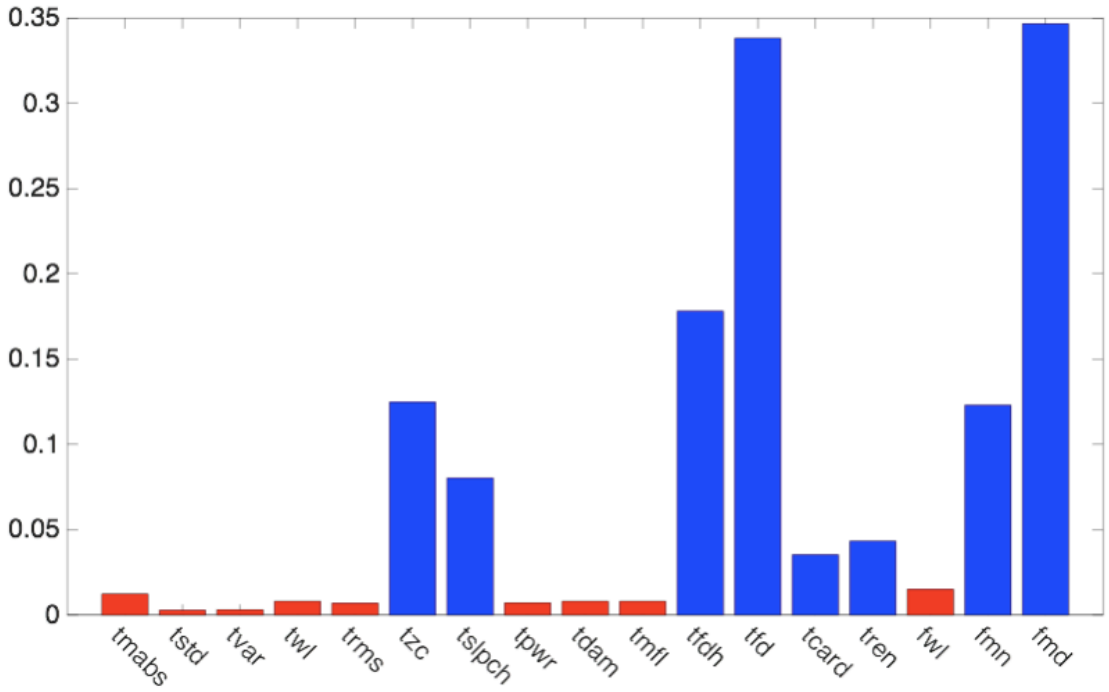


Figure 15 Channel disassociation for features individually

Bar diagram showing the determinant of the correlation matrixes for sets of feature sequences were one feature were extracted from all channels. The values are averages over all subjects in the simultaneous movement data set. The feature algorithm used is stated at the horizontal axis. A high value means low correlation between channels for the specific feature. The features are divided into two groups depending on there channel correlation. Red means high correlation, while blue means low.

The features marked by red color has low average correlation matrix determinants, which means high correlation between channels, while the blue color represents features of low channel correlation. Figure 16 shows how the two groups of features, red and blue from Figure 15, cluster differently in classification accuracy against CCE plots.

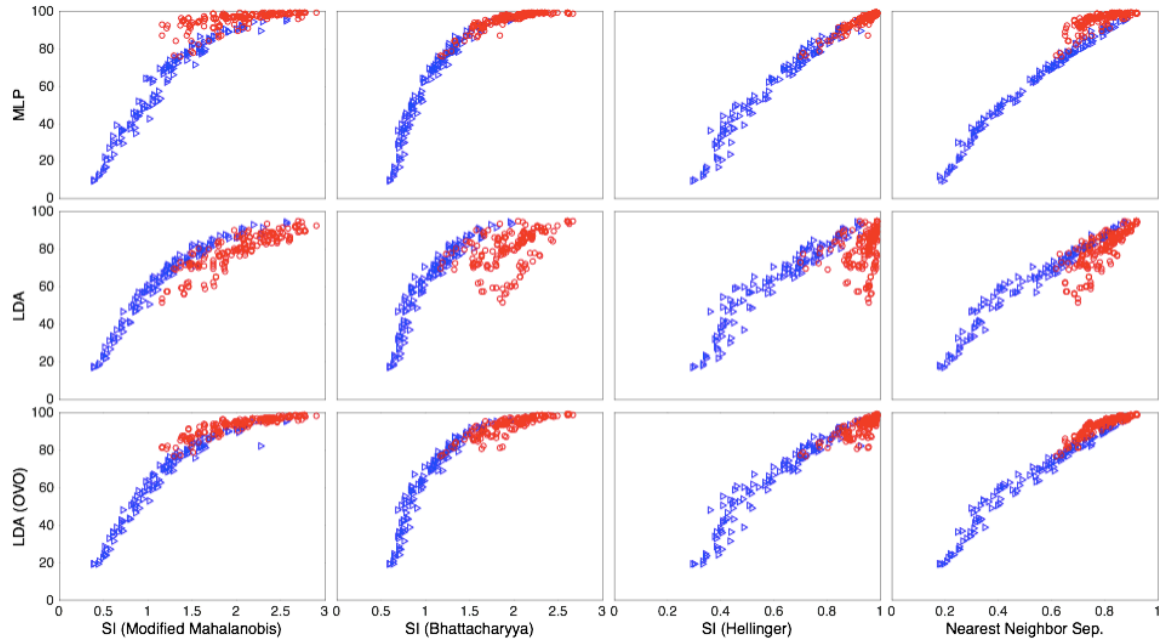


Figure 16 Channel correlation dependency for different Classification Complexity Estimating Algorithms

Illustration of how the two feature groups in Figure 15 cluster in classification accuracy against Classification Complexity Estimate plots. The red and blue group result in high and low channel correlation respectively. Insets in the same row show result from the same Classification Complexity Estimating Algorithm and insets in the same column from the same classifier. Classifiers were used in the conventional "single" topology, aside of LDA which was used in "single" and "One-Vs-One" (OVO).

The *blue* group has similarly dependency on classification accuracy for the three classifiers, while the *red* clearly varies between them.

Dimensionality Dependency

The number of dimensions of a classification task increase as the number of channels and/or features increase. Dimensionality's influence on the correlation with classification accuracy is different for different CCEAs. Figure 17 shows how CCE against classification accuracy clusters shift as the dimensionality change for the CCEAs earlier shown to be relevant for classification accuracy prediction.

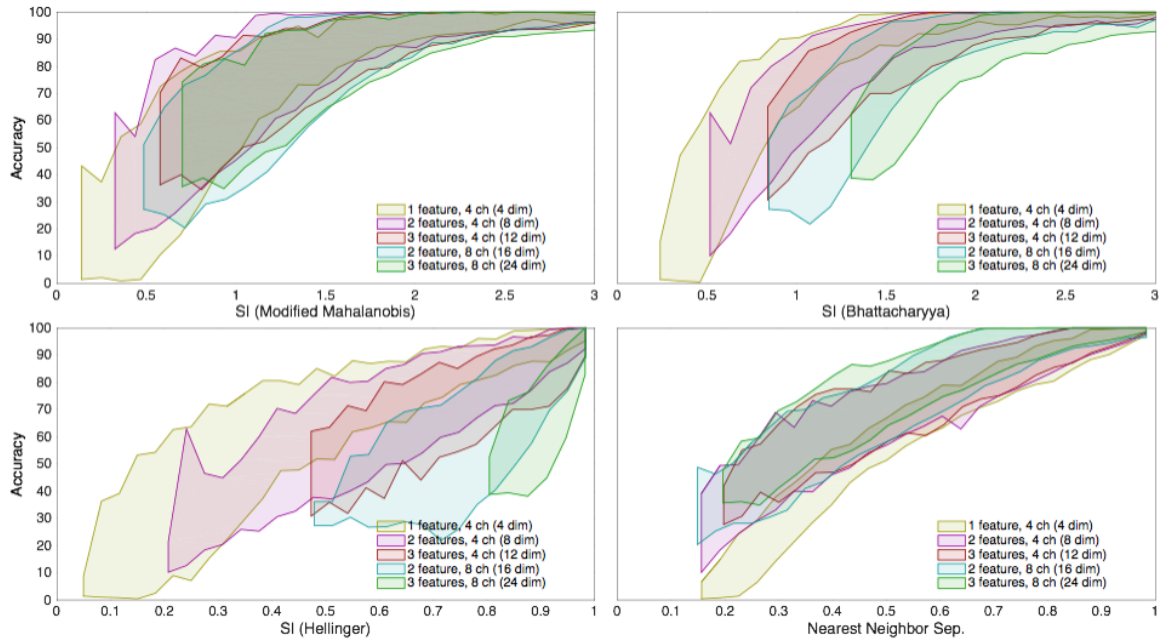


Figure 17 Dimensionality dependency for different Classification Complexity Estimating Algorithms

Surfaces representing clusters in MLP classification accuracy against Classification Complexity Estimates plots. The classification dimensionalities of the clusters are color coded and given in the legends. The different Insets shows the dimensionality dependency for different Classification Complexity Estimating Algorithms.

Both *Bhattacharyyas Distance* and *Modified Mahalanobis* seem to have offset between the clusters of different classification dimensionality, while *Hellingers Distance*'s dimensionality dependency seem more complex. Corresponding clusters for NNS seem to change more depending on the number of features then number of channels, but seems less influenced by dimensionality compared to the other CCEAs.

To compensation for the dimensionality dependency of *Modified Mahalanobis* and *Bhattacharyyas Distance*, average offsets for the dimensionalities $p = 2,3,\ldots,128$ were used. The values represent the average change in CCE when going from the previous dimensionality (p-1) to the considered dimensionality (p). The average offsets were obtained after 100000 repeated calculations of *Modified Mahalanobis* and *Bhattacharyyas Distance* for every *p*, using random feature vectors with members from a 0 to 1 uniform distribution. The results are plotted against *p* in Figure 18, and Figure 19 shows how the distributions in Figure 17 align as they are compensated.

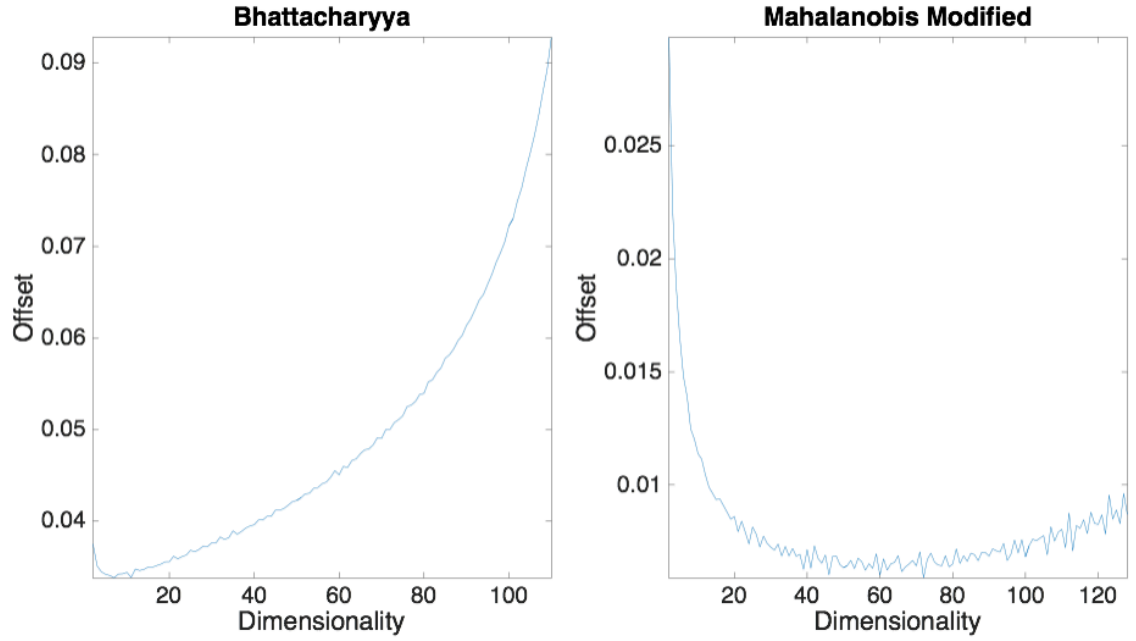


Figure 18 General increase of output due to increased dimensionality

The general increment of Bhattacharyyas Distance (left inset) and Modified Mahalanobis (right inset) do to increment in dimensionality. The general increment for one dimensionality is its average value minus the average value of the directly prior dimensionality after 100000 repeated calculations with 0 - 1 uniform distributed random inputs.

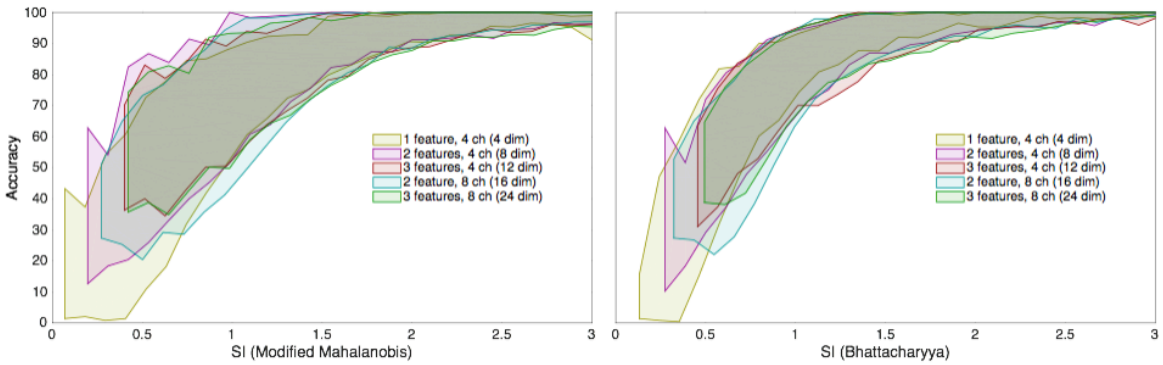


Figure 19 Cluster alignment due to compensation for Classification Dimensionality dependency

Surfaces representing clusters in MLP classification accuracy against Classification Complexity Estimates plots that are compensated for dimensionality dependency. Classification dimensionality of the clusters are color coded and given in the legends as number of features and channels.

Discussion

Offline Results

Separability index

Modified Mahalanobis was the *Distance Definition* that most correlated with classification accuracy (Table 3). However, the *Distance Definitions* based on *Bhattacharyyas Coefficient*, being *Bhattacharyyas Distance* and *Hellingers Distance*, had a higher correlation with MLP's classification accuracy. In the section *Feature Attributes*, it is shown that *Bhattacharyyas Distance* is compensating for the change in dependency to MLP classification accuracy caused by input correlation that is found in the other CCEAs. It should therefore be a more adequate *Distance Definition* for estimation of MLP classification complexity. However, as features are combined into sets, the feature correlation tend to decrease as larger feature vectors are formed using multiple features. This is probably a reason for the absence of significantly higher classification accuracy for *Bhattacharyyas Distance* (Figure 9).

Nearest Neighbor Separability

NNS has high correlation with classification accuracy for all classifiers, as shown in Table 4. Figure 10 shows that the *best sets* corresponding to NNS perform higher overall classification accuracy then both the SI *best sets* and the *reference sets.* The big benefit of NNS is that it does not assume normality of the distribution which makes it more general. There is, however, a dependency to input correlation as can be seen in Figure 16, but just as for *Modified Mahalanobis* this influence will decrease as features are combined into sets and input correlation decrease.

The drawback of NNS compared to SI is that it is more computationally demanding. As implemented for this study, the computation time for NNS using two featuers is approximately 20 and 16 times longer than for SI with *Modified Mahalanobis* as *Distance Definition* using the *IM* and *SM data* respectively.

Purity and Repeatability Index

Purity and RI does not show as high correlation with with classification accuracy as the other CCEAs evaluated in this study and are therefore not included in the feature set experiment. However, the correlation for RI *average result* is relatively high and positive. Interesting is that RI measures the inconsistence during recording. Higher RI means larger swifts in feature space between recording repetitions. As this was expected to limit the classifiers it was also expected that the correlation with classification accuracy would be negative.

Real-Time

The statistically significant correlations with *Completion Time* in Figure 12 argues that both NNS and SI are relevant for prediction of performance in real-time. However, SI with *Modified Mahalanobis* as *Distance Definition* yield higher correlation with *Completion Time* than NNS, while the offline tests show that the NNS *best sets* are performing with higher classification accuracy for both MLP and LDA also represented in the real-time test. The parametric models of the distributions used for SI are probably more robust to changes present in a real-time situation, similar to what is shown for LDA , also dependent on the assumption of normality [28].

We expected consistent within-class distribution in feature space as represented by RI to be beneficial in the real-time tests, but the low correlation with *Completion Time* in Figure 12 does not confirmed that hypothesis.

Even though correlations between the CCEAs and the *Completion Time* are significant for many CCEAs the correlations with offline accuracy are clearly higher. The complexity of real-

time testing is illustrated in Figure 20 where classifier training data is compared to corresponding real-time data for one movement per inset.

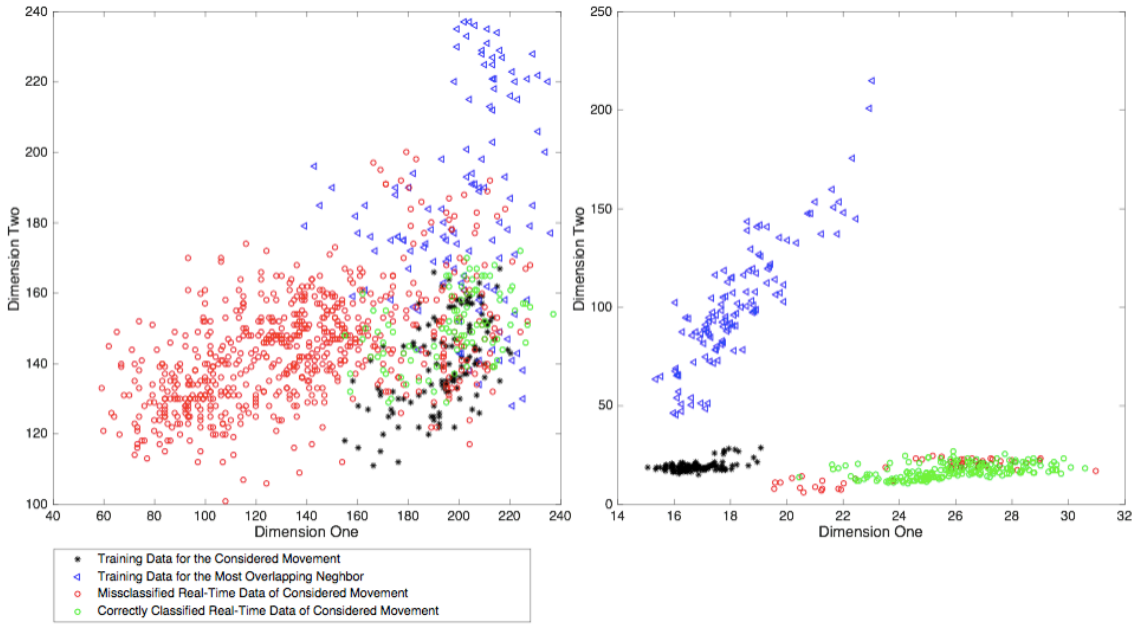


Figure 20 Real-time Classification Complexity illustration

Scatter plots of the classifier training data together the corresponding real-time test data. The inset to the left represents the movement with the highest average completion time of the subject with the overall highest average completion time. The inset to the right represents the movement with the lowest average completion time of the subject with the overall lowest average completion time. The plots also include the movement with the lowest Modified Mahalanobis to the movement being considered. The two dimensions that are used to plot the data were selected from the 16 dimensions of the classification task so that the Modified Mahalanobis between the training data and the real-time data was maximized.

The distribution clearly shifts between the time when training data was recorded and the time when the real-time test was executed.

Dimensionality, Channel Correlation Dependency and Feature Attributes.

For a CCEA to be useful in a channel or feature selection application where sets of different numbers of members will be compared, the dimensionality dependency presented in the result most be considered. The relatively low dependency of NNS and SI using *Modified Mahalanobis* gives them an advantage in such applications. Ways to compensate for dimensionality dependency, as such suggested using common offsets derived by repeated calculations with random inputs, would solve the problem. However, there is a need for further investigation to enable reliable compensation. The reason for this dependency is still unknown to the authors.

The change in dependency between CCEs and classification accuracy due to channel correlation of the features presented in the Channel Correlation Dependency section reveals some interesting attributes of the classifiers. Figure 16 shows that features with high channel correlation result in higher average classification accuracy for MLP compared to LDA, but LDA used in a OVO topology is less influenced by the feature correlation. MLP uses the redundant information in the features more efficient than what is observed for LDA, which suggests that redundancy reduction is of higher importance when selecting both channels and features for a LDA application.

The feature attributes emphasized in Figure 13 and Figure 14 provide information on the performance of the features in different setups. The variation in the top 5 features shows

how dependent the features performance is to others conditions of the classification task, which emphasizes the importance of dynamic feature selection methods for MPR.

Conclusion

This study compared algorithms that estimates the classification complexity of MPR. Two such algorithms, Separability Index (SI) and Nearest Neighbors Separability (NNS), were found to yield high correlation with classification accuracy. There adequacy for MPR was further emphasized by the high classification accuracy yielded by the feature sets selected using the two algorithms. SI was evaluated using different *Distance Definitions,* from which best performance was achieved using a modified version of the *Mahalanobis Distance.* Overall, the offline results indicated that NNS is a more stable CCEA, while SI is less demanding to compute. In addition, feature correlation and dimensionality dependency were found to influence the correlation between CCEs and classification accuracy.

References

[1] E. Scheme and K. Englehart, “Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use,” *J. Rehabil. Res. Dev.*, vol. 48, no. 6, pp. 643–660, 2011.

[2] M. Ortiz-Catalan, N. Sander, M. B. Kristoffersen, B. Håkansson, and R. Brånemark, “Treatment of phantom limb pain (PLP) based on augmented reality and gaming controlled by myoelectric pattern recognition: A case study of a chronic PLP patient,” *Front. Neurosci.*, vol. 8, no. 8 FEB, pp. 1–7, 2014.

[3] X. Zhang and P. Zhou, “High-density myoelectric pattern recognition toward improved stroke rehabilitation,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 6, pp. 1649–1657, 2012.

[4] B. Karlık, “Machine Learning Algorithms for Characterization of EMG Signals,” *Int. J. Inf. Electron. Eng.*, vol. 4, no. 3, pp. 189–194, 2014.

[5] X. Zhang and H. Huang, “A real-time, practical sensor fault-tolerant module for robust EMG pattern recognition,” *J. Neuroeng. Rehabil.*, vol. 12, no. 1, p. 18, 2015.

[6] S. Benatti, F. Casamassima, B. Milosevic, E. Farella, P. Schönle, S. Fateh, T. Burger, Q. Huang, and L. Benini, “A Versatile Embedded Platform for EMG Acquisition and Gesture Recognition,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 5, pp. 620–630, 2015.

[7] M. B. I. Raez, M. S. Hussain, F. Mohd-Yasin, M. Reaz, M. S. Hussain, and F. Mohd-Yasin, “Techniques of EMG signal analysis: detection, processing, classification and applications.,” *Biol. Proced. Online*, vol. 8, no. 1, pp. 11–35, 2006.

[8] N. Bu, O. Fukuda, and T. Tsuji, “EMG-based motion discrimination using a novel recurrent neural network,” *J. Intell. Inf. Syst.*, no. 1978, pp. 113–126, 2003.

[9] M. Ortiz-Catalan, R. Brånemark, and B. Håkansson, “Biologically inspired algorithms applied to prosthetic control,” *IASTED Int. Conf. Biomed. Eng.*, no. BioMed, pp. 7–15, 2012.

[10] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, “Feature reduction and selection for EMG signal classification,” *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, 2012.

[11] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005.

[12] Q. Cheng, H. Zhou, and J. Cheng, “The fisher-markov selector: Fast selecting maximally separable feature subset for multiclass classification with applications to high-dimensional data,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1217–1233, 2011.

[13] J. Liu, X. Li, G. Li, and P. Zhou, “EMG feature assessment for myoelectric pattern recognition and channel selection: A study with incomplete spinal cord injury,” *Med. Eng. Phys.*, vol. 36, no. 7, pp. 975–980, 2014.

[14] N. E. Bunderson and T. A. Kuiken, “Quantification of feature space changes with experience during electromyogram pattern recognition control,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 239–246, 2012.

[15] S. Singh, “Multiresolution Estimates of Classification Complexity,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1534–1539, 2003.

[16] M. Ortiz-Catalan, R. Brånemark, and B. Håkansson, “BioPatRec: A modular research platform for the control of artificial limbs based on pattern recognition algorithms.,” *Source Code Biol. Med.*, vol. 8, no. 1, p. 11, 2013.

[17] M. Ortiz-Catalan, “BioPatRec,” 2016. [Online]. Available:

https://github.com/biopatrec/biopatrec/. [Accessed: 16-Jun-2016].
[18] M. Ortiz-Catalan, B. Håkansson, and R. Brånemark, “Real-time and simultaneous control of artificial limbs based on pattern recognition algorithms,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 4, pp. 756–764, 2014.
[19] M. Ortiz-Catalan, “Cardinality as a highly descriptive feature in myoelectric pattern recognition for decoding motor volition,” *Front. Neurosci.*, vol. 9, no. OCT, pp. 1–7, 2015.
[20] P. Mahalanobis, “On the Generalized Distance in Statistics.” National Institute of Science of India, Calcutta, 1936.
[21] T. Kailath, “The Divergence and Bhattacharyya Distance Measures in Signal Selection,” *IEEE Trans. Comm. Technol.*, vol. 15, pp. 52–60, 1967.
[22] G. Nagy and X. Zhang, “Simple Statistics for Complex Feature Spaces,” in *Data Complexity in Pattern Recognition*, 2006, pp. 173–195.
[23] S. Kullback and R. A. Leibler, “On Information and Sufficiency,” *The Annals of Mathematical Statistics*, vol. 22, no. 1. pp. 79–86, 1951.
[24] D. A. Cieslak, T. R. Hoens, N. V. Chawla, and W. P. Kegelmeyer, “Hellinger distance decision trees are robust and skew-insensitive,” *Data Min. Knowl. Discov.*, vol. 24, no. 1, pp. 136–158, 2012.
[25] E. J. Scheme, S. Member, K. B. Englehart, S. Member, B. S. Hudgins, and S. Member, “Selective Classification for Improved Robustness of Myoelectric Control Under Nonideal Conditions,” vol. 58, no. 6, pp. 1698–1705, 2011.
[26] B. Hudgins, P. Parker, and R. N. Scott, “A new strategy for multifunction myoelectric control,” *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, 1993.
[27] T. A. Kuiken, B. A. Lock, R. D. Lipschutz, L. A. Miller, K. A. Stubblefield, and K. B. Englehart, “Targeted Muscle Reinnervation for Real-time Myoelectric Control of Multifunction Artificial Arms,” *JAMA*, vol. 301, no. 6, pp. 619–628, 2016.
[28] P. Kaufmann, K. Englehart, and M. Platzner, “Fluctuating EMG signals: Investigating long-term effects of pattern matching algorithms,” *2010 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC’10*, pp. 6357–6360, 2010.