INVESTIGATION OF FEATURE REDUCTION METHODS FOR IMPROVING EMG AND EEG PATTERN RECOGNITION ROBUSTNESS

Article and worksheets

Master's thesis, Spring 2015
Astrid Clausen Nørgaard
Biomedical Engineering and Informatics

AALBORG UNIVERSITY
DENMARK

Aalborg University
Department of Health Science & Technology

Investigation of features reduction methods for improving EMG and EEG pattern recognition robustness

Master's thesis - 10th semester
Biomedical Engineering and Informatics

Student: Astrid Clausen Nørgaard
Group: 15gr1083
Supervisor: Ernest Kamavuako
Number of pages: 54
Completion date: 3th of June 2015

Preface

This project was carried out by Astrid Clausen Nørgaard as her master's thesis on the 10th semester of Biomedical Engineering and Informatics on Aalborg University. The time period of the project was the 1st of February 2015 to the 3th of June 2015.

Acknowledgment

The author would like to thank the subjects who participated in the study, as well as Julie Gade and Rósa Hugosdóttir, whom the experiments was carried out in cooperation with. The author would also like to thank Rasmus Wiberg Nedergaard for kindly handing over his data form his master's thesis, and Ernest Nlandu Kamavuako who supervised this project.

Reading guide

This project contains two parts, an article and worksheets. The article serves as the main documentation, while the worksheets can be read, if the reader wishes to get a supplementary information about the feature reduction methods or the extracted features. The references are specified by the Vancouver System in the article and Harvard Method [last name, year] in the worksheets.

Abstrakt

Reducering af dimensionen af signal egenskaber (RDSE) er et essentielt trin i mønster genkendelse. Kliniske studier lider ofte af et højt antal af signal egenskaber og et lavt antal observationer, hvor RDSE ofte er nødvendigt for at fjerne redundante signal egenskaber og for at undgå overtilpasning af klassifikatøren. Litteratur, som omhandler RDSE, fokuser primært på RDSE's evne til at forbedre klassificeringen. RDSE's evne til at forbedre robusthed inden for mønstergenkendelse bliver derimod ofte overset i litteraturen. Dette er til trods for at litteraturen rapporterer stor variation i biologiske signaler optaget over forskellige dage/sessioner.

Formålet med dette projekt var at undersøge hvilke RDSE-metoder, der medvirker til den mest robuste klassificering over flere dage. Dette blev undersøgt ved at analysere otte RDSE-metoder på to forskellige datasæt: 1) Elektromyografi (EMG)-data optaget over 3 dage, hvor otte forsøgspersoner udførte syv forskellige håndbevægelser, 2) Elektroencefalografi (EEG)-data optaget over 7 dage, hvor syv forsøgspersoner udførte to forskellige dorsalflektioner. Efter filtrering og segmentering af datasættene, blev henholdsvis 90 og 72 signal egenskaber udtrukket af EMG og EEG-datasættet. Dimension af signal egenskaberne blev herefter reduceret med følgende otte RDSE-metoder, som blev udvalgt på baggrund af en litteratur gennemgang:

- Principal component analysis (PCA)
- Fisher discriminant analysis (FDA)
- Kernel principal component analysis (KPCA)
- Nonparametric discriminant analysis (NDA)
- Independent component analysis (ICA)
- Nonparametric weighted feature extraction (NWFE)
- Neighbourhood components analysis (NCA)
- Maximally collapsing metric learning (MCML)

RDSE-metoder blev evalueret med klassifikatørerne linear discriminant analysis (LDA) og Support vector machine (SVM). Robustheden af RDSE-metoderne blev evalueret gennem to senarier, ét senarie hvor RDSE-projektionen og klassifikatørerne blev trænet hver dag, og ét senarie hvor RDSE-projektionen og klassifikatørerne kun blev trænet den første dag.

Resultaterne for EMG, viste at NCA havde en høj klassifiseringsnø-

jagtighed og var den mest robuste RDSE-metode, for begge senarier af træning. Resultaterne for EMG, viste at KPCA havde den højest klassifiseringsnøjagtighed og var blandt de mest robuste RDSE-metoder, for begge senarier af træning.
Det kan gennem dette projekt konkluderes, at RDSE-metoder kan forbedre robustheden. Ved implementering af RDSE-metoder i et klassifiseringsystem, anbefales det, at man tester forskellige RDSE-metoder, for at finde den metode der passer det givne signal bedst.

Article

Investigation of feature reduction methods for improving EMG and EEG pattern recognition robustness

Author: Astrid Clausen Nørgaard

Abstract

Robustness of pattern recognition receives little attention in literature dealing with feature reduction, despite that current literature reports inconsistency and day-to-day / session-to-session variation in biomedical signals. This article aims to investigate the robustness of eight feature reduction methods for data recorded over multiple days. The feature reduction methods were tested on two dataset: 1) Electromyography (EMG)-data recorded during three days, where eight subjects performed seven different hand movements and 2) Electroencephalography (EEG)-data recorded during seven days, where seven subjects performed two different dorsiflexions. The results show that features reduction has a great impact on the performance and robustness of EMG and EEG classification. For EMG, Nonparametric discriminant analysis (NDA) showed high classification accuracies and was the most robust feature reduction method. For EEG, Kernel principal component analysis (KPCA) showed the highest classification accuracies and was among the most robust feature reduction methods. In conclusion, feature reduction must be included, when designing a classification system that is robust over time, but it is recommended to test the different methods for feature reduction, to find the method that fits the given data the best.

Keywords: Feature reduction · Dimension reduction · Robustness · EEG · EMG · LDA · SVM · Pattern recognition

1 Introduction

Feature reduction is an essential step in biomedical pattern recognition [19, 17]. Clinical studies are often hampered by a large number of features and low number of observations, also known as the *curse of dimensionality*. Therefore reduction of features is often necessary to remove redundant features and to avoid overfitting. Furthermore it has been shown that feature reduction can improve the classification accuracy, when comparing with no feature reduction [19].
Some of the most commonly used feature reduction methods for biomedical signals includes Principal Component Analysis (PCA) and Fishers Discriminant Analysis (FDA). These methods are widely tested in the literature, and are often used as a benchmark when testing a new method for feature reduction [24, 18, 7]. Current literature dealing with feature reduction mainly focuses on feature reduction's ability to improve the classification accuracy [24]. However, when dealing with EMG classification, the following three properties has been suggested to ensure a high quality feature space [2, 22, 5, 4, 28, 16]:

1. *Maximum class separability*: A high quality feature space should have maximum class separability or minimum overlap, to ensures high classification accuracy.
2. *Robustness*: A high quality feature space should be able to adapt time-varying changes.
3. *Complexity*: The computational complexity of the feature space should be kept low.

Maximum class separability and *complexity* are well studied in current literature [24, 3, 29]. However, *robustness* receives little attention in literature dealing with feature reduction. This is despite that current literature reports inconsistency and day-to-day / session-to-session variation in biological signals [27, 1]. An investigation of how well feature reduction methods can handle these inconsistency, and make the classification more robust is therefore needed [1].

1.1 Related Work

Literature dealing with EMG-classification is often broken down in the three signal processing components: *feature extraction*, *dimensionality reduction* and *classification*. Studies by Kaufmann et al. [12] and Phinyomark et al. [23] have investigated robustness for classifier and feature extraction respectively.
Kaufmann et al. recorded EMG data during 21 days. Five different classifiers (k-nearest-neighbor, linear discriminant analysis, decision trees, artificial neural networks and support vector machines) were compared when classifying ten different hand movements. The results show

that the classification accuracies gradually decrease during the 21 days, if the classifier was not retrained with current data. However, LDA only dropped 3.6 % during the 21 days and was found to be the most robust classifier [12].
Phinyomark et al. [23] used the same EMG data as Kaufmann et al. to investigate the robustness of 50 time-domain and frequency-domain features. Sample entropy was the most robust feature and showed a classification accuracy at 93.37 % when the classifier (LDA) was not retrained with current data. This was only 2.45 % lower compared to when the classifier was retrained [23].
Studies dealing with dimensionality reduction and robustness, are not investigated in any current literature.

1.2 Aim

The aim of this study is to investigate which of eight feature reduction method that produces the most robust performance. This will be studied though the following four objectives:

1. Investigate robustness of eight feature reduction methods across multiple days, when the classifier and feature projection are retrained.
2. Investigate robustness of eight feature reduction methods across multiple days, when the classifier and feature projection are not retrained.
3. Investigate the general performance and the robustness eight feature reduction methods compared to the original feature space.
4. Investigate the number of features needed for the feature reduction methods to show the highest performance.

The aim will be investigated by use of two datasets:

- EMG-data from three different days, where eight subjects performed seven different hand movements.
- EEG-data from seven different days, where seven subjects performed two different dorsiflexions.

2 Methods

2.1 EMG experiment

The experiment was performed over three separate days with two and four days in between.

Subjects
The EMG data were collected from eight healthy volunteers (three women and five men) with mean age of 25 $\pm$ 1 year. All subjects were right handed and none of the subjects had any known sensory-motor deficits. All subjects gave their written informed consent to participate in the study.

Recording
The EMG signals were recorded with an analog EMG amplifier (AnEMG12, OT Bioelettronica, Italy) at a frequency of 2 kHz. The signals were digitalized using a 16-bit ADC and recorded by the software, Mr. Kick (Knud Larsen, SMI, Aalborg University).

Experimental procedure
After preparation with electrode gel, one pair of Ag/AgCl surface electrodes (Ambu Neuroline 720) was placed on the following five positions:

1. The pronator teres muscle
2. The flexor digitorum superficialis muscle and flexor carpi radialis muscle
3. The flexor carpi ulnaris muscle
4. The extensor digitorum muscle
5. The extensor carpi ulnaris muscle and extensor carpi radialis muscle

The positions of the electrodes is shown on Figure 1.
Furthermore, a wrist-band was placed around the subject's wrist as reference electrode. Data was recorded during a steady-state medium contraction with the right hand of the following seven hand movements: hand closing (HC), hand opening (HO), wrist flexion (WF), wrist extension (WE), wrist supination (WS), wrist pronation(WP) and pinch grip (PG). Further data was recorded during rest. The total number of classes was thereby eight. The hand movements are shown on Figure 2. Each movement was performed four times.

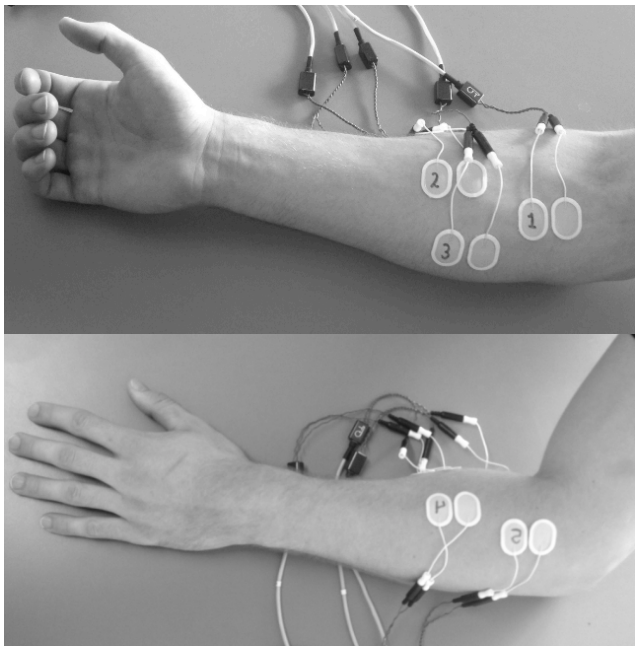

Figure 1: *The position of the electrodes.*

The position of the electrodes was marked after each session, to ensure identical placement of the electrodes at each day.

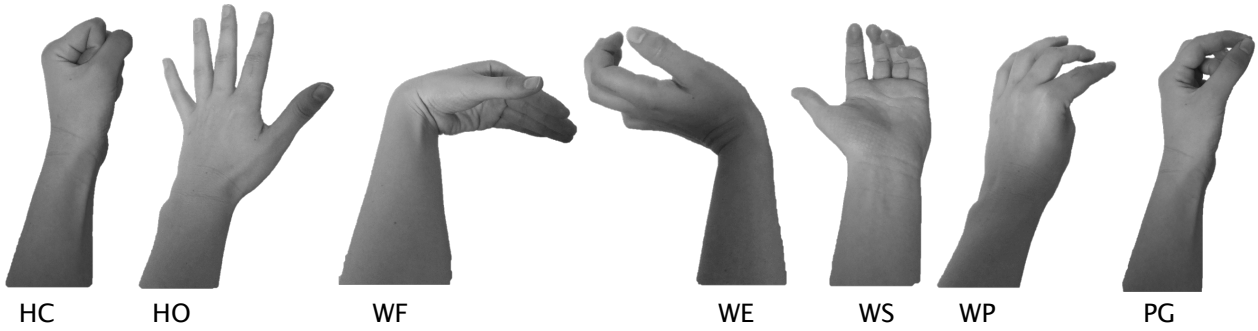

Figure 2: *Hand movements: hand closing (HC), hand opening (HO), wrist flexion (WF), wrist extension (WE), wrist supination (WS), wrist pronation (WP) and pinch grip (PG). Selection of hand movements are inspired by [23]*

2.2 EEG experiment

In order to make a more general conclusion of robustness of the feature reduction methods, an EEG dataset was analysed in this study. The experiment was performed two times per week for four weeks and one session at week eight. This makes it possible to analyse robustness for an extended period of time. The EEG experiment was conducted by Rasmus Wiberg Nedergaard [20] and data was used with permission from Rasmus Wiberg Nedergaard. Further description of the experiment can be seen in his master thesis [20].

Subjects
The EEG data were collected from seven healthy volunteers (one woman and six men), with mean age of 26 $\pm$ 1 year. None of the subjects had any known neurological disorders and disorders of their right foot or ankle. All subjects gave their written informed consent before participation [20].

Recording
The EEG signals were recorded with an EEG amplifier (Nuamps Express, Neuroscan) and a 32 channel Quick-Cap (Neuroscan) at a frequency of 500 Hz. The signals were digitally converted with 32 bits accuracy. Furthermore, force was sampled with 2000 Hz from a force transducer mounted on a foot pedal, and displayed by the software Mr. Kick (Knud Larsen, SMI, Aalborg University) [20].

Experimental procedure
The electrodes were placed at F3, F4, C3, C4, Cz, P3, P4 and Pz according to the 10-20 system [20]. A reference was placed on the right mastoid bone and the ground electrode was placed at the nasion. The impedance of the electrodes were kept below 5 kΩ. Three MVC forces of a dorsiflexion of the right ankle was initially recorded at each session. The subjects performed two kind of dorsiflexions at force of 20 % of the highest MVC. 1) a fast movement, reaching the target force after 0.5 s, and 2) a slow movement, reaching the target force after 3 s. This study used the part of the experiment, where the subjects performed 2 x 30 movements of fast and slow dorsiflexions in randomised order. A trigger was sent at the beginning of each movement to be able to split the continuous recording into epochs.
Data from week 2 was excluded in this study due to technical errors in the recordings. Data from the remaining seven days was included in the study [20].

2.3 Data analysis

Preprossesing of EMG data
The data was bandpass filtered using a fourth order Butterworth bandpass filter with cut-off frequencies at 20 and 400 Hz. Furthermore, the data was filtered with a narrow notch bandstop to remove 50-Hz noise. This was followed by a windowing with a segment length of 250 ms with an overlap of 150 ms.

Feature extraction of EMG data
The following features were extracted from the filtered EMG data from all five channels:

1. Mean Absolute Value
2. Wilson Amplitude, $threshold = 10mV$
3. Zero Crossing, $threshold = 10mV$
4. Slope Sign Changes, $threshold = 16mV$
5. Variance Of EMG
6. Wave Length
7. Root Mean Square
8. Mean Frequence
9. Mean Power
10. Median Frequence
11. 6 Autoregressive coefficients, $order = 6$
12. Sample Entropy, $m = 2, r = 0.2 \times \sigma$ [1]
13. Approximate Entropy $m = 2, r = 0.2 \times \sigma$ [2]

The dimension of the original feature space was thereby 90.
1-11 are all common used features in EMG classification [21, 23]. Sample entropy and approximate entropy are extracted due to their robustness found in the study by Phinyomark et al. [23].
The value of the parameters are based on the suggestion

[1] m=embedded dimension, r=tolerance. See worksheets, chapter 4
[2] m=embedded dimension, r=tolerance. See worksheets, chapter 4

in the literature [23].
More information about the extracted EMG features can be found in the worksheets in chapter 4.

Preprossesing of EEG data
The data was bandpass filtered using a fourth order Butterworth bandpass filter with cut-off frequencies at 0.5 and 5 Hz. The signals were split into epochs with a segment length of 5 s via the trigger.

Feature extraction of EEG data
Features was extracted from the movement-related cortical potentials (MRCP), which is a low frequency negative shifts, that are associated with planning and execution of a voluntary movement [10]. The mean MRCP for the two movements, across channels and subject, can be seen on Figure 3. Time domain features like mean, max, slope and intersection are often extracted from the MRCP [10]. The chosen features are based on visual inspection of Figure 3.

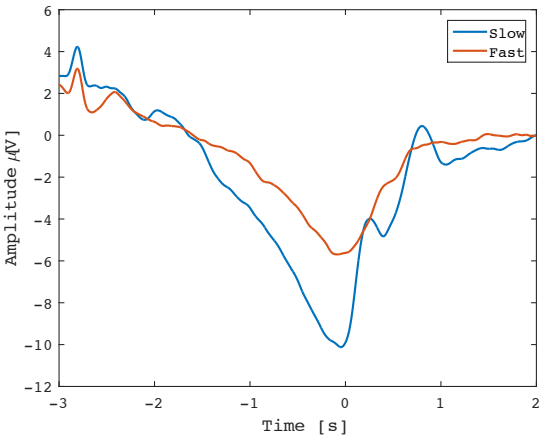

Figure 3: *Mean MRCP from two movements, slow and fast. 0 s is the time of the movement onset.*

The following features were extracted from the filtered EEG data from all nine channels:

1. Mean amplitude from -0.5 s to 0.5 s
2. Mean amplitude from -1 s to 0 s
3. Point of maximum negativity
4. Maximum negativity
5. Slope of a linear regression from -1 s to 0 s
6. Intersection of a linear regression from -1 s to 0 s
7. Slope of a linear regression from 0 s to 1 s
8. Intersection of a linear regression from 0 s to 1 s

The dimension of the original feature space was thereby 72.

2.4 Feature reduction

The following nine methods were used to reduce the features. The selection of methods is based on the literature review in the worksheets (chapter 2). The worksheets further contain a mathematical approach of the used methods (chapter 3).

PCA
Principal Component Analysis (PCA) is one of the most popular methods for dimensionality reduction [18]. PCA seeks to maximise the variance in the data by mapping the data into a linear subspace, containing the principal components. The first principal components describes the most variance in the data, and so forth, and is found on the basis of the eigenvectors and eigenvalues [6].

FDA
Fisher discriminant analysis (FDA) is another popular method for feature reduction. FDA is a supervised method that seeks to maximise the between-class scatter matrix, and minimise the within-class scatter matrix [6].

KPCA
Kernel principal component analysis (KPCA) is a variant of PCA that uses a nonlinear kernel function, rather than the original linear function, before finding the eigenvectors and the eigenvalues of the kernel matrix [8]. Different kernels was tested on both datasets, and a Gaussian radial basis function ($\sigma = 30$) showed the best results on average, and was therefore implemented.

NDA
Nonparametric discriminant analysis (NDA) is similar to FDA as it also relies on the scatter matrixes. NDA however defines a nonparametric between-class scatter matrix [11].

ICA
Independent Component Analysis (ICA) is a blind source separation technique, that separates a dataset into independent, non-Gaussian subcomponents [19, 3]. The goal of ICA is to find the features that are most independent from each other [8] and is in this study implemented by the FastICA method.

NWFE
Nonparametric Weighted Feature Extraction (NWFE) is a feature reduction method, which idea is to compute the *weighted mean* by weighing every sample differently. On the basis of the *weighted mean* the nonparametric between-class and within-class scatter matrices are defined [13].

NCA
Neighborhood Components Analysis (NCA) is a supervised method, that seeks to find a Mahalanobis distance metric for k-nearest-neighbours (kNN) that optimizes the leave-one-out error on the training set [17]. The optimization problem are non-convex and relies on a gradient based iterative algorithm,

MCML
Maximally Collapsing Metric Learning (MCML) is a supervised method, which is similar to NCA, and also relies on the Mahalanobis distance metric for k-nearest neighbours. MCML differs from NCA, as the optimization problem is convex for MCML [7].

2.5 Evaluation of feature space

The features space was evaluated by calculating the classification accuracies using Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM).

LDA
LDA was chosen because of its robustness found by Kaufmann et al. [12]. Additionally LDA is a simple classifier, that is computationally efficient, and it does not require any adjustment of parameter [23]. However, LDA is limited when the number of features are high compared to the number of samples, often referred to as "*large p small n*" [25]. Hence, it is not possible to evaluate the full original feature space with LDA, and LDA will only be tested with the projected features.

SVM
SVM are well known for being able to deal with a high dimensional feature space [24]. Furthermore, it was found to be the second most robust classifier in the study by Kaufmann et al. [12]. SVM, one-against-all with a linear kernel, will be used to evaluate the full original feature space as well as the projected feature space.

2.6 Evaluation of robustness

The measurement for robustness was chosen to be the *standard deviation* between days, as suggested in [9]. The robustness across the multiple days was tested with and without retraining of the classifier and the feature reduction projection.

Retraining:
Four-fold cross-validation was used, when the classifier and the feature reduction projection was retrained. The data was partitioned into four equal folds, e.g. the EMG data was split into four folds containing eight samples each. In each fold all eight classes was represented. Three of the four fold was used as training data, and the cross-validation process was performed by testing all four different combinations of training and test data for each of the four folds. The classification and the feature reduction projection were therefore performed four times with four different combinations of training and test data. The mean of the four classification errors was presented.

No retraining:
When testing the robustness without retaining of the classifier and the feature reduction projection, day 1 acted as training data, and day 2 and day 3 etc. acted as test data.

2.7 Evaluation of number of features

The number of features was optimised, so that the number of features that gave the highest classification accuracy, was chosen. This was chosen rather than using a fixed number of features, e.g. based on a certain percentage of the explained variability for PCA, as this may not necessarily show the highest classification accuracy.

Statistics:
To test if the results were robust across days, the non-parametric Friedman test was used, as the assumptions for ANOVA (equal variance and sphericity) was not met for all data. For p-values below 0,05 a Bonferroni Post Hoc test was used. Three Friedman tests were performed, with the following three aims:

- Test for differences between the days.
- Test for differences between the feature reduction methods.
- Test for differences between retrain and no retrain.

3 Results

3.1 EMG

Retrain
The average EMG classification accuracies for LDA and SVM across subjects, are presented in Table 1. These results are obtained by using the retrained data and 4-fold cross validation. On average, it is seen that NDA shows the highest classification accuracy when classifying with LDA. It is also seen that NDA is the most robust method, with the lowest standard deviation. PCA also shows the high classification accuracy when classifying with both LDA and SVM, but the robustness of PCA is lower than most of the other methods.
It should also be noted that SVM shows a lower classification accuracies for all methods and also shows a lower robustness.
The original feature space was only outperformed by three of the feature reduction methods (PCA, NWFE and NDA) when classifying with SVM.
A significant change was found between the days for ICA classified with SVM. The post hoc test showed a statistical significant difference between day 2 and day 3 ($p = 0.03$).
Furthermore, a significant difference between the methods was found, where p was < 0.01 for both LDA and SVM. The post hoc test showed the following statistical significant differences:

- ICA was significant different from NCA when classifying with LDA ($p = 0.02$)

- ICA was significant different from NCA when classifying with SVM (p = 0.03)
- ICA was significant different from PCA when classifying with SVM (p = 0.02)

Table 1: *Mean classification accuracies across subjects, with retrained EMG data.*

LDA					
Method	**Day 1**	**Day 2**	**Day 3**	**Mean ± std**	**p**
PCA	89.8	**95.7**	94.5	93.4 ± 3.1	0.09
FDA	68.0	75.4	73.0	72.1 ± 3.8	0.35
KPCA	87.5	85.5	83.6	85.5 ± 2.0	0.27
NDA	**93.4**	95.3	**94.9**	**94.5 ± 1.0**	0.84
ICA	44.1	45.3	42.6	44.0 ± 1.4	0.37
NWFE	79.3	84.8	81.3	81.8 ± 2.8	0.38
NCA	87.9	90.2	89.5	89.2 ± 1.2	0.76
MCML	89.1	92.2	89.1	90.1 ± 1.8	0.26
SVM					
Method	**Day 1**	**Day 2**	**Day 3**	**Mean ± std**	**p**
PCA	**83.2**	78.5	**75.8**	**79.2 ± 3.8**	0.34
FDA	58.2	59.8	62.9	60.3 ± 2.4	0.61
KPCA	55.5	56.3	49.2	53.6 ± 3.9	0.64
NDA	79.3	**79.7**	73.8	77.6 ± 3.3	0.24
ICA	26.2	22.7	28.5	25.8 ± 2.9	*0.03
NWFE	77.0	72.7	70.7	73.4 ± 3.2	0.30
NCA	65.6	64.8	60.5	63.7 ± 2.7	0.52
MCML	68.8	64.5	61.7	65.0 ± 3.5	0.42
Orig. feat.	69.9	68.8	64.5	67.7 ± 2.9	0.39

No retrain

Table 2 shows the average EMG classification accuracies across subjects when LDA and SVM and the feature projection was not retained. NWFE shows an average classification accuracy at 94.3 % when classifying with LDA. The classification accuracies obtained by SVM are again lower than for LDA, and the highest classification accuracy for SVM is NDA (77.7 %). NDA also shows a very robust performance, both when classifying with LDA and SVM.

Beside ICA, all feature reduction methods show higher classification accuracies than the original feature space. The feature reduction methods also improved the robustness.

In the statistical tests, no significant changes were found between the days.

Furthermore, no significant difference between the methods was found when classifying with LDA (p=0.06). However, a significant difference between the methods was found, when classifying with SVM (p=0.04). The post hoc test showed a statistical significant difference between ICA and NDA (p = 0.05).

Table 2: *Mean classification accuracies across subjects, where day 1 acted as training data, and day 2 and 3 as test data.*

LDA				
Method	**Day 2**	**Day 3**	**Mean ± std**	**p**
PCA	87.9	85.2	86.5 ± 1.9	0.41
FDA	78.5	72.7	75.6 ± 4.1	0.16
KPCA	88.7	81.6	85.2 ± 5.0	0.06
NDA	89.1	89.1	89.1 ± 0.0	0.56
ICA	45.3	39.8	42.6 ± 3.9	0.48
NWFE	**96.5**	**92.2**	**94.3 ± 3.0**	0.32
NCA	87.5	81.6	84.6 ± 4.1	0.41
MCML	85.9	83.6	84.8 ± 1.7	0.65
SVM				
Method	**Day 2**	**Day 3**	**Mean ± std**	**p**
PCA	75.8	74.2	75.0 ± 1.1	0.41
FDA	54.7	58.6	56.6 ± 2.8	0.10
KPCA	73.4	69.5	71.5 ± 2.8	0.48
NDA	**78.1**	**77.3**	**77.7 ± 0.6**	0.32
ICA	41.0	35.5	38.3 ± 3.9	0.10
NWFE	64.8	57.0	60.9 ± 5.5	0.18
NCA	77.3	73.4	75.4 ± 2.8	0.48
MCML	77.3	75.4	76.4 ± 1.4	0.48
Org. feat.	60.9	50.8	55.9 ± 7.2	0.16

Retrain vs. no retrain

The mean differences between retrain and no retrain are seen in Table 3. A negative value indicates lower performance when not being retrained and vice versa. The results from the statistical tests show that most methods have a significant differences between retrain and no retrain. Some of the methods showed a higher performance when not being retrained, e.g. NWFE with LDA as classifier showed 11.3% higher classification accuracy, than when being retrained.

Table 3: *Mean difference between retrain and no retrain, EMG*

Method	**LDA**		**SVM**	
	Mean diff. betw. retrain & no retrain	**p**	**Mean diff. betw. retrain & no retrain**	**p**
PCA	-8.6	*0.01	-2.1	0.26
FDA	1.4	0.71	-4.7	0.06
KPCA	0.6	1.00	18.8	*0.03
NDA	-6.1	0.10	1.0	0.48
ICA	-1.4	1.00	12.7	*0.00
NWFE	11.3	*0.00	-10.7	*0.01
NCA	-5.3	0.06	12.7	*0.03
MCML	-5.9	*0.03	13.3	*0.01
Org. feat.			-10.7	0.16

Number of features

Table 4 shows the median of required features to obtain the highest EMG classification accuracies across subjects. NWFE is the method that requires the lowest number of features, both for LDA and SVM, and when being retained and not retrained. The number of features using NWFE are reduced from 90 features to 6 and 7.

Table 4: *Median of the required features across subjects to obtain the presented EMG classification accuracies.*

Method	LDA		SVM	
	Retrain	No retrain	Retrain	No retrain
PCA	**6**	9	**6**	8
FDA	**6**	**6**	10	17
KPCA	9	12	**6**	20
NDA	**6**	9	**6**	13
ICA	10	19	**6**	27
NWFE	**6**	**6**	**6**	**7**
NCA	7	8	7	17
MCML	7	13	7	20

3.2 EEG

Retrain

The average EEG classification accuracies across subjects, for LDA and SVM are presented in Table 5. On average, KPCA shows the highest classification accuracy for both LDA and SVM. KPCA also shows a robust performance, as the standard diviations are among the lowest.
The original feature space is only outperformed by KPCA and NWFE.
In the statistical tests, significant changes between the days were found at three of the tests. The post hoc test showed the following statistical significant differences:

- NCA using SVM: week 1 - day 2 and week 4 - day 1 (p = 0.05)
- MCML using SVM: week 3 - day 2 and week 4 - day 2 (p = 0.05)
- NDA using SVM: week 1 - day 2 and week 3 - day 2 (p = 0.04)

Furthermore, a significant difference between the methods was found, where p was < 0.01 for both LDA and SVM. The post hoc test showed the following statistical significant differences:

- KPCA was significant different from FDA (p=0.02), PCA (p=0.04) and ICA(p=0.02) when classifying with LDA.
- KPCA was significant different from FDA (p=0.01), ICA (p=0.01) and MCML (p=0.05) when classifying with SVM.
- FDA was significant different from KPCA (p=0.01), NWFE (p=0.05) and orig. feat. (p=0.05) when classifying with SVM.

No retrain

Table 6 shows the average EEG classification accuracies, when LDA and SVM and the feature projection was not retained. KPCA shows the best average classification accuracy and the most robust performance, for both LDA and SVM. All methods, expect ICA, show a higher average classification accuracy than the original feature space.
In the statistical tests, significant changes between the days were found at three of the test. The post hoc test showed the following statistical significant differences:

- NCA using LDA: week 4 - day 1 and week 8 (p = 0.04)
- NWFE using SVM: week 1 - day 2 and week 8 (p = 0.03)
- NCA using SVM: week 4 - day 1 and week 8 (p = 0.05)
- MCML using SVM: week 4 - day 1 and week 8 (p = 0.02)

Furthermore, a significant difference between the methods was found, where p was < 0.01 for both LDA and SVM. The post hoc test showed the following statistical significant differences:

- KPCA was significant different from FDA (p=0.01), ICA (p=0.02) and NWFE (p=0.03) when classifying with LDA.
- PCA was significant different from FDA (p=0.03), ICA (p=0.04) when classifying with LDA.
- KPCA was significant different from ICA (p=0.02) and orig. feat. (p=0.03) when classifying with SVM.
- PCA was significant different from ICA (p=0.05) and orig. feat. (p=0.05) when classifying with SVM.

Retrain vs. no retrain

The mean differences between retrain and no retrain are seen in Table 7. A negative value indicates lower performance when not being retrained and vice versa. The results from the statistical tests show that non of the methods have a significant differences between retrain and no retrain.

Table 7: *Mean difference between retrain and no retrain, EEG*

Method	LDA		SVM	
	Mean diff. betw. retrain & no retrain	p	Mean diff. betw. retrain & no retrain	p
PCA	1.2	0.71	1.2	0.71
FDA	2.1	0.71	1.4	0.71
KPCA	-4.7	0.06	-3.3	0.26
NCA	-2.4	0.26	-1.3	0.06
ICA	-1.1	0.71	0.4	0.71
NWFE	-5.0	0.06	-4.4	0.26
MCML	-2.0	0.26	-0.6	0.71
NDA	-0.6	0.71	-0.1	0.26
Orig. feat.			-6.8	0.26

Number of features

Table 8 shows the median of required features to obtain the highest EEG classification accuracies across subjects. Similar to the results found for EMG, NWFE is the method that on average requires the lowest number of features.

Table 5: *Mean classification accuracies across subjects, with retrained EEG data.*

LDA									
Method	**Week 1 Day 1**	**Week 1 Day 2**	**Week 3 Day 1**	**Week 3 Day 2**	**Week 4 Day 1**	**Week 4 Day 2**	**Week 8 Day 1**	**Mean ± std**	**p**
PCA	64.9	72.3	65.8	62.9	65.2	71.2	72.4	67.8 ± 4.0	0.11
FDA	63.7	64.8	67.1	62.7	62.9	62.9	65.3	64.2 ± 1.6	0.42
KPCA	**74.4**	**83.1**	**75.4**	**76.3**	**74.8**	**77.2**	**77.3**	**76.9 ± 2.9**	0.62
NDA	65.7	74.5	64.6	65.6	64.8	73.5	71.2	68.6 ± 4.3	0.06
ICA	64.9	68.0	62.9	66.3	63.5	64.0	65.8	65.1 ± 1.8	0.52
NWFE	65.9	75.8	66.6	66.8	71.2	69.3	68.4	69.2 ± 3.5	0.51
NCA	65.4	72.7	67.2	62.9	67.1	71.5	72.0	68.4 ± 3.7	0.06
MCML	65.2	73.5	66.3	66.1	63.3	71.8	71.6	68.2 ± 4.0	0.07
SVM									
Method	**Week 1 Day 1**	**Week 1 Day 2**	**Week 3 Day 1**	**Week 3 Day 2**	**Week 4 Day 1**	**Week 4 Day 2**	**Week 8 Day 1**	**Mean ± std**	**p**
PCA	64.5	71.8	67.1	62.0	64.2	71.8	70.9	67.5 ± 4.1	0.19
FDA	59.1	64.7	64.5	63.3	62.8	64.0	67.0	63.5 ± 2.4	0.42
KPCA	**73.2**	**82.0**	**73.7**	**75.5**	**73.5**	**75.9**	**74.6**	**75.5 ± 3.1**	0.31
NDA	65.7	74.3	65.0	63.8	64.4	72.2	69.6	67.8 ± 4.2	*0.02
ICA	64.2	63.1	61.4	64.1	64.5	64.0	63.4	63.5 ± 1.0	0.81
NWFE	67.8	75.9	67.5	68.8	73.0	71.5	69.0	70.5 ± 3.1	0.63
NCA	64.4	73.3	65.3	63.9	63.5	69.1	69.3	67.0 ± 3.7	*0.02
MCML	63.6	71.4	63.7	62.0	63.2	73.0	71.4	66.9 ± 4.8	*0.01
Orig. feat.	66.8	78.4	68.5	68.7	69.3	70.1	70.8	70.4 ± 3.8	0.11

Table 6: *Mean classification accuracies across subjects, where day 1 acted as training data, and day 2, day 3 ect. as test data.*

LDA								
Method	**Week 1 Day 2**	**Week 3 Day 1**	**Week 3 Day 2**	**Week 4 Day 1**	**Week 4 Day 2**	**Week 8 Day 1**	**Mean ± std**	**p**
PCA	**74.3**	68.2	71.5	65.4	68.5	69.0	69.5 ± 3.1	0.07
FDA	67.7	61.1	62.4	64.1	62.9	62.8	63.5 ± 2.3	0.67
KPCA	72.3	**72.4**	**74.4**	**71.1**	**73.5**	**72.5**	**72.7 ± 1.1**	0.86
NDA	72.4	68.0	69.4	65.0	67.2	68.5	68.4 ± 2.4	0.74
ICA	68.5	62.2	63.8	62.7	61.2	65.5	64.0 ± 2.7	0.30
NWFE	70.2	62.5	64.6	63.4	65.1	62.6	64.7 ± 2.9	0.13
NCA	68.9	65.7	63.1	63.9	66.7	70.7	66.5 ± 2.9	*0.02
MCML	68.8	65.3	64.6	66.1	65.6	70.3	66.8 ± 2.3	0.19
SVM								
Method	**Week 1 Day 2**	**Week 3 Day 1**	**Week 3 Day 2**	**Week 4 Day 1**	**Week 4 Day 2**	**Week 8 Day 1**	**Mean ± std**	**p**
PCA	72.9	69.0	67.6	65.9	68.6	71.1	69.2 ± 2.5	0.14
FDA	72.1	64.5	63.8	63.8	63.8	66.9	65.8 ± 3.3	0.58
KPCA	**74.4**	**72.8**	**73.2**	**69.2**	**74.2**	**71.6**	**72.5 ± 1.9**	0.39
NDA	71.5	67.6	66.9	65.7	68.2	68.9	68.1 ± 2.0	0.30
ICA	66.7	61.5	64.3	62.9	61.5	65.9	63.8 ± 2.2	0.21
NWFE	70.9	64.6	67.0	64.8	67.5	64.6	66.6 ± 2.5	*0.03
NCA	66.7	65.4	63.1	63.6	68.0	69.9	66.1 ± 2.6	*0.01
MCML	69.8	67.2	62.9	63.8	65.9	71.3	66.8 ± 3.3	*0.03
Orig. feat.	69.3	63.5	62.8	60.9	65.1	63.3	64.1 ± 2.8	0.13

Table 8: *Median of the required features across subjects to obtain the presented EEG classification accuracies.*

Method	LDA		SVM	
	Retrain	No retrain	Retrain	No retrain
PCA	12	22	12	23
FDA	13	20	10	20
KPCA	16	17	12	18
ICA	13	18	11	**15**
NWFE	**2**	**7**	**9**	16
NCA	14	13	11	18
MCML	16	18	13	16
NCA	15	22	10	21

4 Discussion

The four aforementioned objectives will be discussed along with a discussion of the methodology.

1. Robustness of feature reduction methods, retrain
For EMG, NDA showed the highest and most robust classification accuracy obtained by LDA. This is similar to results found in previous literature, where NDA showed to outperform e.g. FDA and PCA [15, 11]. The high robustness for NDA has however never been reported before.
Also PCA showed high classification accuracy for both LDA and SVM. This was unexpected, as several studies report that PCA shows lower classification accuracy when comparing to e.g. FDA, NWFE and ICA [5, 28, 24].
For EEG the results differ from EMG, as KPCA showed high and robust classification accuracy, and KPCA was significant different from many of the other methods.
It is seen from the results, that the choice of feature reduction, can be a trade off between robustness and classification accuracy. For instance, PCA showed one of the high average performances for both EMG and EEG, but also showed a poor robustness. A general conclusion of which feature reduction method that shows the most robust classification accuracy when retraining cannot be be drawn.

2. Robustness of feature reduction methods, no retrain
For EMG, NWFE showed a high classification accuracy at 94.3 % when classifying with LDA. This is 11.3 % higher compared to the retrain-test. Compared to studies that investigated robustness of EMG classification, drops of 3.6 % for the most robust classifier and 2.45 % for the most robust feature are reported [12, 23]. These studies was however recorded during 21 days.
NWFE was not the most robust feature reduction within the no retrain-test, but is still within an acceptable range. NDA was the most robust feature reduction method, just like it also was seen in the retrain-test.
For EEG, KPCA showed the highest classification accuracy for both LDA and SVM (76.9 % and 75.5 %). This is 4.7 % and 3.3 % lower than when comparing to the retrain-test, which is quite similar to the results found in [12, 23].
KPCA was also found to be one of the most robust feature reduction methods within the no retrain-test.
When testing over multiple days, studies report that retraining sessions can be necessary each day to overcome time variations in the signals [14, 26]. In this study, many of the methods showed no significant difference between the retrain-test and the no retrain-test. The results found in this study, thereby indicates, that this retraining session might not be necessary, if an appropriate feature reduction method is used.

3. Robustness and performance of feature reduction methods compared to the original feature space
For EMG, retrain-test, PCA, NWFE and NDA showed better performance than the original features, but did not tend to improve the robustness.
For EMG, no retrain-test, all feature reduction methods, except ICA, showed higher performance than the original feature space. Also, the robustness for the no retrain-test, was improved for all feature reduction methods.
For EEG, most feature reduction methods did not show to have the same positive impact on the results, when comparing to the original feature space. A significant difference between the original features and KPCA for the no retrain-test was however found.

4. Dimension
It was found that NWFE needed the fewest number of features for both EMG and EEG. The number of features for NWFE was reduced from 90 to 6-7 features for EMG, and from 72 to 2-16 for EEG.
These results are similar to the results found in previous literature [16]. It was found that NWFE needed the lowest number of features compared to PCA and FDA [16]. NWFE might thereby be able to overcome the *curse of dimensionality*-phenomena.

5. Methodology
It should be considered if an experiment recorded over three days are enough to evaluate the robustness. Ideally, the EMG experiment should have been recorded over an extended period of time, to draw a more certain conclusion about the robustness.
Also, the methodology for the no retrain-test should be considered. Four-fold cross validation was not applied in the no retrain-test. The number for training samples for e.g. EMG was therefore increased from 24 to 32. This might be the reason for big differences between the retrain-test and no-retrain test, e.g. for NWFE which showed an increase of 11.3 % in the no retrain-test for EMG.

6. Conclusion
This study was the first of its kind to investigate robustness of feature reduction methods. The aim of this study was to investigate eight feature reduction methods and their ability to produces robust per-

formance. Feature reduction shows to have a great impact on the performance and robustness of EMG and EEG classification. For EMG, NDA showed high classification accuracies and was the most robust feature reduction methods. For EEG, KPCA showed the highest classification accuracies and was among the most robust feature reduction methods.
In order to make a classification system that is robust over time and can adapt time-varying changes, feature reduction must be included. However, it is recommended to test the different methods for feature reduction, to find the method that fits the given data best, as the results were highly dependent of the signal and the classifier.

References

[1] AlZoubi, O., Fossati, D., D'Mello, S., and Calvo, R. A. Affect detection from non-stationary physiological data using ensemble classifiers. *Evolving Systems* (2014), 1–14.

[2] Boostani, R., and Moradi, M. H. Evaluation of the forearm emg signal features for the control of a prosthetic hand. *Physiological measurement 24*, 2 (2003), 309.

[3] Cao, L., Chua, K., Chong, W., Lee, H., and Gu, Q. A comparison of pca, kpca and ica for dimensionality reduction in support vector machine. *Neurocomputing 55*, 1 (2003), 321–336.

[4] Chan, A. D., and Green, G. C. Myoelectric control development toolbox. In *Proceedings of 30th Conference of the Canadian Medical & Biological Engineering Society* (2007), vol. 1, pp. M0100–1.

[5] Chu, J.-U., Moon, I., Lee, Y.-J., Kim, S.-K., and Mun, M.-S. A supervised feature-projection-based real-time emg pattern recognition for multifunction myoelectric hand control. *Mechatronics, IEEE/ASME Transactions on 12*, 3 (2007), 282–290.

[6] Giri, D., Acharya, U. R., Martis, R. J., Sree, S. V., Lim, T.-C., Ahamed, T., and Suri, J. S. Automated diagnosis of coronary artery disease affected patients using lda, pca, ica and discrete wavelet transform. *Knowledge-Based Systems 37* (2013), 274–282.

[7] Globerson, A., and Roweis, S. T. Metric learning by collapsing classes. In *Advances in neural information processing systems* (2005), pp. 451–458.

[8] Huang, K., Velliste, M., and Murphy, R. F. Feature reduction for improved recognition of subcellular location patterns in fluorescence microscope images. In *Biomedical Optics 2003* (2003), International Society for Optics and Photonics, pp. 307–318.

[9] Jin, Y., and Sendhoff, B. Trade-off between performance and robustness: an evolutionary multiobjective approach. 237–251.

[10] Jochumsen, M., Niazi, I. K., Mrachacz-Kersting, N., Farina, D., and Dremstrup, K. Detection and classification of movement-related cortical potentials associated with task force and speed. *Journal of neural engineering 10*, 5 (2013), 056015.

[11] Kamavuako, E. N., Scheme, E. J., and Englehart, K. B. Nonparametric discriminant projections for improved myoelectric classification. 68–69.

[12] Kaufmann, P., Englehart, K., and Platzner, M. Fluctuating emg signals: Investigating long-term effects of pattern matching algorithms. In *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2010), pp. 6357–6360.

[13] Kuo, B.-C., and Landgrebe, D. A. Nonparametric weighted feature extraction for classification. *Geoscience and Remote Sensing, IEEE Transactions on 42*, 5 (2004), 1096–1105.

[14] Li, X., Blanton, R., Grover, P., and Thomas, D. E. Ultra-low-power biomedical circuit design and optimization: Catching the don't cares. In *Integrated Circuits (ISIC), 2014 14th International Symposium on* (2014), IEEE, pp. 115–118.

[15] Li, Z., Lin, D., and Tang, X. Nonparametric discriminant analysis for face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 31*, 4 (2009), 755–761.

[16] Lin, C.-T., Lin, K.-L., Ko, L.-W., Liang, S.-F., Kuo, B.-C., Chung, I.-F., et al. Nonparametric single-trial eeg feature extraction and classification of driver's cognitive responses. *EURASIP Journal on Advances in Signal Processing 2008*, 1 (2008), 849040.

[17] Manit, J., and Youngkong, P. Neighborhood components analysis in semg signal dimensionality reduction for gait phase pattern recognition. In *Broadband and Biomedical Communications (IB2Com), 2011 6th International Conference on* (2011), IEEE, pp. 86–90.

[18] Martis, R. J., Acharya, U. R., and Min, L. C. Ecg beat classification using pca, lda, ica and discrete wavelet transform. *Biomedical Signal Processing and Control 8*, 5 (2013), 437–448.

[19] Mwangi, B., Tian, T. S., and Soares, J. C. A review of feature reduction techniques in neuroimaging. *Neuroinformatics 12*, 2 (2014), 229–244.

[20] Nedergaard, R. W. Online detection and classification of movement-related cortical potentials on healthy volunteers and patients with a stroke affecting their motor cortex. http://projekter.aau.dk/projekter/files/198493376/14gr1072bed_mmelse2014.pdf, 2014, Master Thesis, Aalborg University.

[21] Phinyomark, A., Limsakul, C., and Phukpattaranont, P. A novel feature extraction for robust emg pattern recognition. *arXiv preprint arXiv:0912.3973* (2009).

[22] Phinyomark, A., Phukpattaranont, P., and Limsakul, C. Feature reduction and selection for emg signal classification. *Expert Systems with Applications 39*, 8 (2012), 7420–7431.

[23] Phinyomark, A., Quaine, F., Charbonnier, S., Serviere, C., Tarpin-Bernard, F., and Laurillau, Y. Emg feature evaluation for improving myoelectric pattern recognition robustness. *Expert Systems with Applications 40*, 12 (2013), 4832–4840.

[24] Subasi, A., and Gursoy, M. I. Eeg signal classification using pca, ica, lda and support vector machines. *Expert Systems with Applications 37*, 12 (2010), 8659–8666.

[25] Tai, F., and Pan, W. Incorporating prior knowledge of gene functional groups into regularized discriminant analysis of microarray data. *Bioinformatics 23*, 23 (2007), 3170–3177.

[26] Wang, W., Collinger, J. L., Degenhart, A. D., Tyler-Kabara, E. C., Schwartz, A. B., Moran, D. W., Weber, D. J., Wodlinger, B., Vinjamuri, R. K., Ashmore, R. C., et al. An electrocorticographic brain interface in an individual with tetraplegia. *PloS one 8*, 2 (2013), e55344.

[27] Wheeler, K. R. Device control using gestures sensed from emg. In *Soft Computing in Industrial Applications, 2003. SMCia/03. Proceedings of the 2003 IEEE International Workshop on* (2003), IEEE, pp. 21–26.

[28] Yang, P., Xing, K., Huang, J., and Wang, Y. A novel feature reduction method for real-time emg pattern recognition system. In *Control and Decision Conference (CCDC), 2013 25th Chinese* (2013), IEEE, pp. 1500–1505.

[29] Zhao, J., Yu, P. L., and Kwok, J. T. Bilinear probabilistic principal component analysis. *Neural Networks and Learning Systems, IEEE Transactions on 23*, 3 (2012), 492–503.

Worksheets

Contents

List of abbreviations 1

Table 1.1: List of abbreviations, feature reduction methods

PCA	Principal component analysis
FDA	Fisher discriminant analysis
KPCA	Kernel principal component analysis
NDA	Nonparametric discriminant analysis
ICA	Independent component analysis
NWFE	Nonparametric weighted feature extraction
NCA	Neighbourhood components analysis
MCML	Maximally collapsing metric learning
NLDA	Nonlinear discriminant analysis
SOFM	Self-organizing feature maps
NFA	Nonparametric feature analysis
NLPCA	Nonlinear principal component analysis
LPP	Locality preserving projection
LLE	Locally linear embedding
CCA	Canonical correlation analysis
DBFE	Decision boundary feature extraction
BDFS	Bhattacharyya distance feature selection
CMM	Correlative Matrix Mapping
NMCML	Non Convex MCML
KLDA	Kernel LDA
BM	Bayesian method

Table 1.2: List of abbreviations, classifiers

SVM	Support vector machine
kNN	k-nearest neighbors
NBC	Naive Bayes classifier
MLP	Multilayer perceptron
GMM	Gaussian mixture model
NN	Neural network
PNN	Orobabilistic neural network
LMkNN	Local mean k-nearest neighbors
RF	Random forest
DT	Decision Trees
MCS	Multiple classifier systems

Literature Review 2

The purpose of this chapter is to give an overview of the current research dealing with feature reduction. The selected feature reduction methods in the article are based on this literature review.

2.1 Methods for the literature review

The following keywords were used during the literature search:

- Feature reduction
- Dimension reduction
- Feature extraction
- Feature projection

Furthermore, chain search was also used, where the references in the already found literature, was investigated.
Only studies that tested two or more methods were included in this review.
The literature will be presented in tables containing:

- The reference
- The data used in the article
- The feature reduction methods used in the article
- The applied classifiers in the article
- A short conclusion of the article

Please notice, that some of the articles reoccur in the tables, e.g. an article dealing with PCA and FDA, will occur in both Table 2.1 and Table 2.2.

2.2 PCA

Principal Component Analysis (PCA) is one of the most popular unsupervised linear method for dimensionality reduction [Martis et al., 2013]. PCA seeks to maximise the variance in the data by mapping the data into a linear subspace, containing the principal components. PCA is often used as a benchmark in the literature, why there are many studies dealing with PCA.
Only literature dealing with biological signals will be presented in this section. Literature dealing with PCA can be seen in Table 2.1.
It is seen that PCA, despite being a popular method, is outperformed by most other methods.

Table 2.1: Literature dealing with PCA

Article	Data	Methods	Classifier	Conclusion
[Subasi & Gursoy, 2010]	EEG	PCA FDA ICA	SVM	PCA was outperformed by FDA and ICA.
[Lin et al., 2008]	EEG	PCA FDA NWFE	KNN NBC	PCA was outperformed by NWFE, but showed better performance than FDA.
[Yang et al., 2013]	EMG	PCA NWFE	SVM	PCA was outperformed by NWFE.
[Chu et al., 2007]	EMG	PCA FCA NLDA SOFM	MLP	PCA showed lower performance than NLDA and FDA, but better than SOFM.
[Giri et al., 2013]	ECG	PCA FDA ICA	SVM GMM PNN kNN	PCA showed the highest average performance across four classifiers, but ICA with GMM as classifier showed the highest performance.
[Martis et al., 2013]	ECG	PCA LDA ICA	SVM NN PNN	PCA with PNN as classifier, showed a higher performance than FDA, but lower performance than ICA.

2.3 FDA

FDA is another popular method for feature reduction. FDA is a supervised method, that seeks to maximise the between-class scatter matrix, and minimising the within-class scatter matrix [Giri et al., 2013]. Only literature dealing with biological signals will be presented for FDA, see Table 2.2.
It is seen that FDA, is outperformed by most other methods.

Table 2.2: Literature dealing with FDA

Article	Data	Methods	Classifier	Conclusion
[Subasi & Gursoy, 2010]	EEG	FDA PCA ICA	SVM	FDA showed a better performance than PCA, but was outperformed by ICA.
[Lin et al., 2008]	EEG	PCA FDA NWFE	kNN NBC	FDA was outperformed by both NWFE and PCA.
[Chu et al., 2007]	EMG	FCA PCA NLDA SOFM	MLP	FDA showed a higher performance than PCA and SOFM, but was outperformed by NLDA.
[Kamavuako et al., 2014]	EMG	FDA NDA NFA	kNN LMkNN	FDA was outperformed by the other methods.
[Giri et al., 2013]	ECG	FDA PCA ICA	SVM GMM PNN kNN	FDA was outperformed by PCA and ICA.
[Martis et al., 2013]	ECG	FDA PCA ICA	SVM NN PNN	FDA was outperformed by PCA and ICA.

2.4 KPCA

Kernel principal component analysis (KPCA) is a variant of PCA that uses a nonlinear kernel function, rather than the original linear [K. Huang et al., 2003]. Literature dealing with ICA can be seen in Table 2.3. In general KPCA shows good performance in the literature, but there was not found any literature that investigates KPCA and biological signals.

Table 2.3: Literature dealing with KPCA

Article	Data	Methods	Classifier	Conclusion
[Castaings et al., 2010]	Image	KPCA PCA NWFE BDFS DBFE	SVM RF	The study tested two different datasets and KPCA was in general outperformed by NWFE and BDFS, but showed better results than PCA and DBFE.
[K. Huang et al., 2003]	Image	KPCA PCA ICA NLPCA	SVM	KPCA outperformed the other tested feature reduction methods. Four feature selection methods was also tested, and they all showed higher performance than KPCA.
[W. Huang & Yin, 2012]	Image	KPCA PCA LPP LLE ISOMAP CCA	kNN soft k-NN LDA SVM	On average, LPP and LLE outperformed the other methods, and KPCA showed similar results to the remaining methods.
[Cao et al., 2003]	Seven various datasets	KPCA PCA ICA	SVM	KPCA showed the highest performance for all tested datasets.

2.5 NDA

NDA is a nonparametric method, that is similar to FDA as it also relies on the scatter matrixes. Literature dealing with NDA can be seen in Table 2.8. NDA shows good performance, when comparing to traditional methods, but is outperformed by NFA.

Table 2.4: Literature dealing with MCML

Article	Data	Methods	Classifier	Conclusion
[Kamavuako et al., 2014]	EMG	NDA FDA NFA	kNN LMkNN	NDA showed better performance than FDA, but was outperformed by NFA
[Li et al., 2009]	Image	NDA PCA FDA BM KLDA	MCS	NDA outperformed the other methods for both of the tested datasets.

2.6 ICA

ICA is a blind source separation technique, that separates a dataset into independent, non-Gaussian subcomponents [Cao et al., 2003; Mwangi et al., 2014].
Literature dealing with ICA can be seen in Table 2.5. ICA outperforms PCA and FDA for the litterature dealing with EEG and ECG, but was outperformed by KPCA in litterature dealing with image and other various datasets.

Table 2.5: Literature dealing with ICA

Article	Data	Methods	Classifier	Conclusion
[Subasi & Gursoy, 2010]	EEG	ICA PCA FDA	SVM	ICA showed higher performance than PCA and FDA.
[Martis et al., 2013]	ECG	ICA PCA FDA	SVM NN PNN	ICA with PNN as classifier showed higher performance than any the other combination of classifier and feature reduction methods.
[Giri et al., 2013]	ECG	ICA PCA FDA	SVM GMM PNN KNN	ICA with GMM as classifier showed higher performance than any the other combinations of classifier and feature reduction methods.
[K. Huang et al., 2003]	Image	ICA PCA KPCA NLPCA	SVM	ICA showed higher performance than NLPCA, but was outperformed by KPCA and PCA.
[Cao et al., 2003]	Seven various datasets	ICA PCA KPCA	SVM	ICA showed higher performance than PCA, but is outperformed by KPCA.

2.7 NWFE

NWFE is a new nonparametric feature reduction method.
Literature dealing with NWFE can be seen in Table 2.6. NWFE shows to outperform many of the other methods.

Table 2.6: Literature dealing with NWFE

Article	Data	Methods	Classifier	Conclusion
[Yang et al., 2013]	EMG	NWFE PCA KPCA	SVM	NWFE outperformed PCA
[Lin et al., 2008]	EEG	NWFE PCA FDA	KNN NBC	NWFE outperformed all the other methods for both of the tested classifiers.
[Castaings et al., 2010]	Image	NWFE PCA KPCA BDFS DBFE	SVM RF	On average NWFE outperformed the other methods, for the two tested datasets.

2.8 NCA

NCA is a supervised method, that seeks to find a Mahalnobis distance metric for kNN that optimises the leave-one-out error on the training set [Manit & Youngkong, 2011].
Literature dealing with NCA can be seen in Table 2.7. NCA shows high performance, and outperforms most methods, except NMCML and MCML in [Globerson & Roweis, 2005].

Table 2.7: Literature dealing with NCA

Article	Data	Methods	Classifier	Conclusion
[Manit & Youngkong, 2011]	EMG	NCA PCA FDA LPP	SVM	NCA outperformed the other methods.
[Soto et al., 2011]	Image	NCA FDA MCML CMM CCA	kNN DT SVM	NCA using kNN-classification was the combination that showed the highest performance
[Goldberger et al., 2004]	Six variuos datasets	NCA PCA FDA	kNN	NCA outperformed the other methods for all datasets.
[Globerson & Roweis, 2005]	Six various datasets	NCA NMCML MCML	kNN	NCA was outperformed by NMCML and MCML on average.

2.9 MCML

MCML is a supervised method, which is similar to NCA, and also relies on the Mahalanobis distance metric for k-nearest neighbours [Globerson & Roweis, 2005].
Literature dealing with MCML can be seen in Table 2.8. MCML shows various results, and is e.g. outperformed by NCA in one study [Soto et al., 2011], and is better than NCA in another study [Globerson & Roweis, 2005].

Table 2.8: Literature dealing with MCML

Article	Data	Methods	Classifier	Conclusion
[Soto et al., 2011]	Image	MCML FDA NCA CMM CCA	kNN DT SVM	In general MCML was outperformed by NCA and CMM.
[Globerson & Roweis, 2005]	Six variuos datasets	MCML NMCML NCA	kNN	MCML and NMCML show similar results, but both methods showed better results than NCA on average.

Mathematical approach for the feature reduction methods 3

This chapter gives an overview of the mathematical approaches used in this study. All methods was implemented in Matlab 2015A.

3.1 PCA

The step by step procedure for PCA is as follows [Giri et al., 2013]:

1. Center the feature dataset by subtracting the mean of the dataset, x.

$$x = x - \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3.1}$$

2. Calculate the covariance matrix (Σ) of the centered dataset, where $\bar{m}$ defines the mean vector and N defines the number of dimensions.

$$\Sigma = \frac{1}{N}\{(x - \bar{m})(x - \bar{m})^T\} \tag{3.2}$$

3. Calculate the eigenvectors (V) and the eigenvalues (D) of the covariance matrix.

$$V \cdot \Sigma = V \cdot D \tag{3.3}$$

4. Sort the eigenvectors according to decreasing eigenvalues.
5. Choose the number of desired principal component.
6. Project the training data by multiplying the centered training data and the eigenvectors.
7. Project the test data by multiplying the centered test data and the eigenvectors.

3.2 FDA

The step by step procedure for FDA is as follows [Giri et al., 2013; Kamavuako et al., 2014]:

1. Calculate the between-class scatter matrix:

$$S_b = \sum_{i=1}^{L} \sum_{x_j \in C_i} (x_j - \bar{m}_i)(x_j - \bar{m}_i)^T \tag{3.4}$$

2. Calculate the within-class scatter matrix:

$$S_w = \sum_{i=1}^{L} n_i (\bar{m}_i - \bar{m})(\bar{m}_i - \bar{m})^T \tag{3.5}$$

3. Calculate the eigenvectors and eigenvalues of $(S_w)^{-1} S_b$
4. Sort the eigenvectors according to decreasing eigenvalues.
5. Project the training data by multiplying the training data and the eigenvectors.
6. Project the test data by multiplying the test data and the eigenvectors.

3.3 KPCA

The step by step procedure for KPCA is as follows [K. Huang et al., 2003; Kuzmin & Warmuth, 2007; Kwok & Tsang, 2004]:

1. Construct the kernel matrix, where x defines the the dataset. The value of σ was chosen to be 30 in this study.

$$K(x, x^T) = exp(-\frac{|x - x^T|^2}{2\sigma^2}) \tag{3.6}$$

2. Center the kernel matrix, where 1_N defines a NxN matrix where each element in the matrix is $1/N$

$$K'_n = K - 1_N \cdot K - K \cdot 1_N + 1_N \cdot K \cdot 1_N \tag{3.7}$$

3. Calculate the eigenvectors (V) and the eigenvalues (D) of the centered kernel matrix, K_n.

$$V \cdot K_n = V \cdot D \tag{3.8}$$

4. Sort the eigenvectors according to decreasing eigenvalues.
5. Choose the number of desired principal component.
6. Project the training data by multiplying the centered kernel matrix with the eigenvectors.
7. Construct a centered kernel matrix of the test data and project the test data by multiplying with the eigenvectors.

3.4 NDA

The step by step procedure for NDA is as follows [Kamavuako et al., 2014]:

1. Calculate the within-class scatter matrix:

$$S_w = \sum_{i=1}^{L} n_i(\bar{m}_i - \bar{m})(\bar{m}_i - \bar{m})^T \tag{3.9}$$

2. Calculate the weighting function $\omega(i, j, l)$

$$\omega(i, j, l) = \frac{min\{d^\alpha(x_l^i, NN_k(x_l^i, i)), d^\alpha(x_l^i, NN_k(x_l^i, i)\}}{d^\alpha(x_l^i, NN_k(x_l^i, i)), d^\alpha(x_l^i, NN_k(x_l^i, i)} \tag{3.10}$$

   where d denotes the Euclidian distance, α controls speed of the changing, regard to the distance ratio and x_l^i denotes the feature vector, l, in class i.
3. Calculate the between-class scatter matrix:

$$S_b = \sum_{i=1}^{c} \sum_{\substack{j=1 \\ j \neq i}}^{c} \sum_{l=1}^{N_i} \omega(i, j, l) \cdot (x_l^i - m_j(x_l^i)) \cdot (x_l^i - m_j(x_l^i))^T \tag{3.11}$$

4. Calculate the eigenvectors and eigenvalues of $(S_w)^{-1} S_b$
5. Sort the eigenvectors according to decreasing eigenvalues.
6. Project the training data by multiplying the training data and the eigenvectors.
7. Project the test data by multiplying the test data and the eigenvectors.

3.5 ICA

ICA is a blind source separation technique, that separates a dataset into independent, non-Gaussian subcomponents [Cao et al., 2003; Mwangi et al., 2014]. ICA assumes that the dataset x is a linear mixture with the source signal, s, and seeks to find this signal:

$$x = A \cdot s \tag{3.12}$$

The step by step procedure for ICA is as follows [Cao et al., 2003; Martis et al., 2013]:

1. Center the feature dataset by subtracting the mean of the dataset, x.

$$x = x - \frac{1}{N} \sum_{i=1}^{N} x_i \tag{3.13}$$

2. Whitening of the dataset, to ensure that the dataset is Gaussian:

$$\tilde{x} = VD^{-1/2}V^T x \tag{3.14}$$

where VDV^T can be obtained by calculating the covariance matrix:

$$\Sigma = VDV^T \tag{3.15}$$

3. Selection of the independence criteria. FastICA was implemented in this study:
   a) Set a random initial weight vector w
   b) Calculate W^+

$$W^+ = E\{xg(W^T x)\} - E\{g'(W^T x)\} \cdot W \tag{3.16}$$

   Where the non-quadratic function for this study was chosen to be $g(u) = u^3$. E denotes the expected value.
   c) Normalise W^+

$$W^+ = W^+ / ||W^+|| \tag{3.17}$$

   d) Repeat until W^+ is converged.
4. When W is converged, its inverse A is calculated.
5. Project the training data by multiplying the whitened training data and the output from the independence criteria.
6. Project the test data by multiplying the whitened test data and the output from the independence criteria.

ICA was implemented by using fastICA.m developed by Hugo Gvert.

3.6 NWFE

The step by step procedure for NWFE is as follows [Kuo & Landgrebe, 2004]:

1. Calculate the distance matrix as follows:

$$w_{lk}^{(i,j)} = \frac{dist(x_t(i), x_k(i)^{-1}}{\sum_{t=1}^{n_j} dist(x_t(i), x_k(i)^{-1}} \tag{3.18}$$

2. Calculate the weighted means $M_j(x_k^{(i)})$ by using the distance matrix $w_{lk}^{(i,j)}$

$$M_j(x_l(i)) = \sum_{k=1}^{N_j} w_{lk}^{(i,j)} x_k^{(j)} \tag{3.19}$$

3. Calculate the weight of the scatter matrix:

$$\lambda_l^{i,j} = \frac{dist(x_l^{(i)}, M_j(x_l(i)^{-1}}{\sum_{k=1}^{N_j} dist(x_t^{(i)}, M_j(x_t(i)^{-1}} \tag{3.20}$$

4. Calculate the nonparametric between-class scatter matrix:

$$S_b = \sum_{i=1}^{L} P_i \sum_{\substack{j=1 \\ j \neq i}}^{L} \sum_{k=1}^{N_i} \frac{\lambda_k^{i,j}}{n_i} \cdot \left(x_k^{(i)} - M_j(x_k^{(i)})\right) \cdot \left(x_k^{(i)} - M_j(x_k^{(i)})\right)^T \tag{3.21}$$

5. Calculate the nonparametric within-class scatter matrix and regularise it:

$$S_w = \sum_{i=1}^{L} P_i \sum_{k=1}^{N_i} \frac{\lambda_k^{i,j}}{n_i} \cdot \left(x_k^{(i)} - M_j(x_k^{(i)})\right) \cdot \left(x_k^{(i)} - M_j(x_k^{(i)})\right)^T \tag{3.22}$$

$$S_w = 0.5 S_w + 0.5 diag(S_w) \tag{3.23}$$

6. Calculate the eigenvectors and eigenvalues of $(S_w)^{-1} S_b$
7. Sort the eigenvectors according to decreasing eigenvalues.
8. Project the training data by multiplying the training data and the eigenvectors.
9. Project the test data by multiplying the test data and the eigenvectors.

3.7 NCA

The step by step procedure for NCA is as follows [Goldberger et al., 2004]:

1. Center the feature dataset by substracting the mean of the dataset, x.

$$x = x - \frac{1}{N} \sum_{o=1}^{N} x_i \tag{3.24}$$

2. Calculate the mahalanobis metrix of the samples $\{x_1, x_2, .., x_N\}$ with the belonging labels $\{y_1, y_2, .., y_N\}$:

$$d(x_i, x_j) = (Ax_i - Ax_j)^T (Ax_i - Ax_j) \tag{3.25}$$

3. NCA is aiming to find A that maximises the nearest neighbor classification. The optimisation criterion is implemented by use of "soft-neighbor"-approach, where p_{ij} must be calculated:

$$p_{ij} = \frac{exp(-||Ax_i - Ax_j||^2}{\sum_{k \neq i} exp(-||Ax_i - Ax_k||^2)}, p_{ii} = 0 \tag{3.26}$$

4. p_i, the probability that a point i will be classified correctly is calculated:

$$p_i = \sum_{j \in C_i} p_{ij} \tag{3.27}$$

$$C_i = \{j | y_j = y_i\} \tag{3.28}$$

5. The optimisation criterion $f(A)$ is calculated as the sum of all the probabilities of a correctly classification:

$$f(A) = \sum_i p_i \tag{3.29}$$

6. A is finally optimised by the gradient rule:

$$\frac{\partial f}{\partial A} = 2A \sum_i (p_i \sum_k p_{ik} x_{ik} x_{ik}^T - \sum_{j \in C_i} p_{ij} x_{ij} x_{ij}^T) \tag{3.30}$$

7. Project the training data by multiplying the centered training data and A.
8. Project the test data by multiplying the centered test data and A.

NCA was implemented by using the Matlab Toolbox for Dimensionality Reduction developed by Laurens van der Maaten.

3.8 MCML

The step by step procedure for MCML is as follows [Globerson & Roweis, 2005]:

1. Center the feature dataset by substracting the mean of the dataset, x.

$$x = x - \frac{1}{N} \sum_{o=1}^{N} x_i \tag{3.31}$$

2. Calculate the mahalanobis metrix of the samples $\{x_1, x_2, .., x_N\}$ with the belonging labels $\{y_1, y_2, .., y_N\}$:

$$d(x_i, x_j | A) = d_{ij}^A = (x_i - x_j)^T A (x_i - x_j) \tag{3.32}$$

where A denotes the PSD matrix.
3. Calculate the conditional probabilities $p^A(j|i)$ and the conditional distribution $p_0(j|i)$

$$p^A(j|i) = \frac{exp - d_{ij}^A}{\sum_{k \neq i} exp - d_{ik}^A}, i \neq j \tag{3.33}$$

$$p_0(j|i) \propto \begin{cases} 1 & y_i = y_j \\ 0 & = y_i \neq y_j \end{cases} \tag{3.34}$$

4. Minimise A by the the Kullback–Leibler (KL) divergence between p_0 and p^A:

$$\min_A = \sum_i KL[p_0(j|i)|p^A(j|i)] \tag{3.35}$$

5. Project the training data by multiplying the centered training data and A.
6. Project the test data by multiplying the centered test data and A.

MCML was implemented by using the Matlab Toolbox for Dimensionality Reduction developed by Laurens van der Maaten.

Feature extraction - EMG 4

This chapter describes the features extracted from the EMG. The features extracted from the EEG will not be described, due to their simplicity. Throughout this chapter x_i denotes the signal in segment i, and N denotes the length of x_i.

4.1 Mean Absolute Value

Mean Absolute Value (MAV) is a frequently used feature within EMG pattern recognition. It is calculated by taking the mean of the absolute amplitude of the signal [Phinyomark et al., 2012]:

$$MAV = \frac{1}{N}\sum_{i=1}^{N}|x_i| \tag{4.1}$$

4.2 Zero Crossing

Zero crossing (ZC) contains information about the frequencies, but is defined in the time domain. It is defined by the number of time the value of the signal crosses a certain threshold. The threshold for this study is $10mV$. It is calculated as follows [Phinyomark et al., 2012]:

$$ZC = \sum_{i=1}^{N-1}[sgn(x_i \cdot x_{i+1} \bigcap |x_i - x_{i+1}| \geq threshold] \tag{4.2}$$

4.3 Wilson Amplitude

Wilson Amplitude (WAMP) also contains information about the frequencies but defined in the time domain. It reflects the contraction force and the firing of motor units. It is defined by the number of time the difference between two amplitudes exceeds a certain threshold. The threshold

for this study is $10mV$. It is calculated as follows [Phinyomark et al., 2012]:

$$WAMP = \sum_{i=1}^{N-1} [f(|x_i - x_{i+1}|)] \tag{4.3}$$

$$f(x) = \begin{cases} 1, & x \geq threshold \\ 0, & otherwise \end{cases} \tag{4.4}$$

4.4 Slope Sign Changes

Slope Sign Changes (SSC) also contains information about the frequencies, but defined in the time domain. It is defines as the number of times the slopes of the signal changes sign, above a certain threshold. The threshold for this study is $threshold = 10mV$. It is calculated as follows [Phinyomark et al., 2012]:

$$SSC = \sum_{i=2}^{N-1} [f([x_i - x_{i-1}) \cdot (x_i - x_{i01})]] \tag{4.5}$$

$$f(x) = \begin{cases} 1, & x \geq threshold \\ 0, & otherwise \end{cases} \tag{4.6}$$

4.5 Variance Of EMG

Variance Of EMG (VAR) is defined as [Phinyomark et al., 2012]:

$$VAR = \frac{1}{N-1} \sum_{i=1}^{N} x_i^2 \tag{4.7}$$

4.6 Wave Length

Wave length is the cumulative length of the signal, and is calculated as follows [Phinyomark et al., 2012]:

$$RMS = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \tag{4.8}$$

4.7 Root Mean Square

Root Mean Square (SMS) is another frequently used feature within EMG pattern recognition, and is calculated as follows [Phinyomark et al., 2012]:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{4.9}$$

4.8 Mean Frequency

Mean frequency (MNF) is a common used frequency domain feature. The mean frequency is defines by [Phinyomark et al., 2012]:

$$MNF = \sum_{j=1}^{M} f_j P_j \Big/ \sum_{j=1}^{M} P_j \tag{4.10}$$

f_j denotes the frequency in the frequency bin j, P_j denotes the power in the frequency bin j and M is the total number of bins.

4.9 Median Frequency

Median Frequency (MDF) is another popular feature from the frequency domain, and is calculated as follows [Phinyomark et al., 2012]:

$$MDF = \frac{1}{2}\sum_{j=1}^{M} P_j \tag{4.11}$$

4.10 Mean Power

Median Frequency (MNP) of the power spectrum is defined as [Phinyomark et al., 2012]:

$$MNP = \sum_{j=1}^{M} P_j M \tag{4.12}$$

4.11 Autoregressive coefficients

Autoregressive (AR) model is defined as follows [Phinyomark et al., 2012]:

$$x_i = \sum_{p=1}^{P} a_p * x_{i-p} + w_i \tag{4.13}$$

where P denotes the order of the model, which was chosen to be 6 in this study. w_i denotes the white noise error.

4.12 Sample Entropy

Sample Entropy (SampEn) can be found as follows [Kumar & Dewal, 2011]:

1. Form a m vector based on the original EMG data ($x_n = x_1, x_2, ..., x(N)$):

$$X_m(i) = [x(i), x(i+1), ..., x(i+m-1)], 1 \leq i \leq N-m+1 \quad (4.14)$$

   where m is defined as 2 in this study.

2. Calculate the distance between $X_m(i)$ and $X_m(j)$ as follows:

$$d[X_m(i), X_m(j)] = max_{k=0,...,m-1}(|x(i+k) - x(j+k)|)^2 \quad (4.15)$$

3. Calculate the Sample Entropy:

$$SampEn = lim\{-ln[\frac{A_r^m}{B_r^m}]\} \quad (4.16)$$

   where A_r^m defines the number for vector pairs having a distance $<$ r of length m+1, and B_r^m defines the number for vector pairs having a distance $<$ r of length m. r is set to $r = 0.2 \times \sigma$ in this study.

4.13 Approximate Entropy

Approximate Entropy (ApEn) can be found as follows [Kumar & Dewal, 2011]:

1. Form a vector of subsequences of $X = [x(1), x(2), .., x(N)]$:

$$x(i) = [x(i), x(i+1), x(i+2), ..., x(i+m-1)], 1 \leq i \leq N-m \quad (4.17)$$

   where m is defined as 2 in this study.

2. Calculate the distance between $X(i)$ and $X(j)$ as follows:

$$d[x(i), x(j)] = max_{k=0,...,m-1}|x(i+k) - x(j+k)| \quad (4.18)$$

3. Find $M^m(i)$, the number of times the distance is above r. r is set to $r = 0.2 \times \sigma$ in this study. Calculate:

$$C_r^m(i) = \frac{M^m(i)}{N-m+1}, for\, i = 1, ..., N-m+1 \quad (4.19)$$

4. Then find the mean logarithm of $C_r^m(i)$:

$$\psi_r^m = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} lnC_r^m(i) \quad (4.20)$$

5. Repeat the calculations for $m+1$.

6. Calculate the ApEn:

$$ApEn = lim(\psi_r^m - \psi_r^{m+1}) \tag{4.21}$$

where A_r^m defines the number for vector pairs having a distance $<$ r of length m+1, and B_r^m defines the number for vector pairs having a distance $<$ r of length m. r is set to $r = 0.2 \times \sigma$ in this study.

Bibliography

Cao, L., Chua, K., Chong, W., Lee, H., & Gu, Q. (2003). A comparison of pca, kpca and ica for dimensionality reduction in support vector machine. Neurocomputing, 55(1), 321–336.

Castaings, T., Waske, B., Atli Benediktsson, J., & Chanussot, J. (2010). On the influence of feature reduction for the classification of hyperspectral images based on the extended morphological profile. International Journal of Remote Sensing, 31(22), 5921–5939.

Chu, J.-U., Moon, I., Lee, Y.-J., Kim, S.-K., & Mun, M.-S. (2007). A supervised feature-projection-based real-time emg pattern recognition for multifunction myoelectric hand control. Mechatronics, IEEE/ASME Transactions on, 12(3), 282–290.

Giri, D., Acharya, U. R., Martis, R. J., Sree, S. V., Lim, T.-C., Ahamed, T., & Suri, J. S. (2013). Automated diagnosis of coronary artery disease affected patients using lda, pca, ica and discrete wavelet transform. Knowledge-Based Systems, 37, 274–282.

Globerson, A., & Roweis, S. T. (2005). Metric learning by collapsing classes. In Advances in neural information processing systems (pp. 451–458).

Goldberger, J., Roweis, S., Hinton, G., & Salakhutdinov, R. (2004). Neighbourhood components analysis.

Huang, K., Velliste, M., & Murphy, R. F. (2003). Feature reduction for improved recognition of subcellular location patterns in fluorescence microscope images. In Biomedical optics 2003 (pp. 307–318).

Huang, W., & Yin, H. (2012). On nonlinear dimensionality reduction for face recognition. Image and Vision Computing, 30(4), 355–366.

Kamavuako, E. N., Scheme, E. J., & Englehart, K. B. (2014). Nonparametric discriminant projections for improved myoelectric classification. , 68–69.

Kumar, Y., & Dewal, M. (2011). Complexity measures for normal and epileptic eeg signals using apen, sampen and sen. Int J Comput Commun Technol, 2, 6–12.

Kuo, B.-C., & Landgrebe, D. A. (2004). Nonparametric weighted feature extraction for classification. Geoscience and Remote Sensing, IEEE Transactions on, 42(5), 1096–1105.

Kuzmin, D., & Warmuth, M. K. (2007). Online kernel pca with entropic matrix updates. In Proceedings of the 24th international conference on machine learning (pp. 465–472).

Kwok, J.-Y., & Tsang, I. W. (2004). The pre-image problem in kernel methods. Neural Networks, IEEE Transactions on, 15(6), 1517–1525.

Li, Z., Lin, D., & Tang, X. (2009). Nonparametric discriminant analysis for face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(4), 755–761.

Lin, C.-T., Lin, K.-L., Ko, L.-W., Liang, S.-F., Kuo, B.-C., Chung, I.-F., et al. (2008). Nonparametric single-trial eeg feature extraction and classification of driver's cognitive responses. EURASIP Journal on Advances in Signal Processing, 2008(1), 849040.

Manit, J., & Youngkong, P. (2011). Neighborhood components analysis in semg signal dimensionality reduction for gait phase pattern recognition. In Broadband and biomedical communications (ib2com), 2011 6th international conference on (pp. 86–90).

Martis, R. J., Acharya, U. R., & Min, L. C. (2013). Ecg beat classification using pca, lda, ica and discrete wavelet transform. Biomedical Signal Processing and Control, 8(5), 437–448.

Mwangi, B., Tian, T. S., & Soares, J. C. (2014). A review of feature reduction techniques in neuroimaging. Neuroinformatics, 12(2), 229–244.

Phinyomark, A., Phukpattaranont, P., & Limsakul, C. (2012). Feature reduction and selection for emg signal classification. Expert Systems with Applications, 39(8), 7420–7431.

Soto, A. J., Strickert, M., Vazquez, G. E., & Milios, E. (2011). Subspace mapping of noisy text documents. In Advances in artificial intelligence (pp. 377–383). Springer.

Subasi, A., & Gursoy, M. I. (2010). Eeg signal classification using pca, ica, lda and support vector machines. Expert Systems with Applications, 37(12), 8659–8666.

Yang, P., Xing, K., Huang, J., & Wang, Y. (2013). A novel feature reduction method for real-time emg pattern recognition system. In Control and decision conference (ccdc), 2013 25th chinese (pp. 1500–1505).