Learning Probabilistic Classifiers from Electromyography Data for Predicting Knee Abnormalities

Lernen von Probabilistischen Klassifizierern von Electromyography Daten um Kniegelenksabweichungen vorherzusagen

Master-Thesis von Jan Kohlschütter
Dezember 2015

1. Gutachten: Dr. Elmar Rueckert
2. Gutachten: Prof. Dr. Jan Peters

TECHNISCHE UNIVERSITÄT DARMSTADT

Learning Probabilistic Classifiers from Electromyography Data for Predicting Knee Abnormalities
Lernen von Probabilistischen Klassifizierern von Electromyography Daten um Kniegelenksabweichungen vorherzusagen

Vorgelegte Master-Thesis von Jan Kohlschütter

1. Gutachten: Dr. Elmar Rueckert
2. Gutachten: Prof. Dr. Jan Peters

Tag der Einreichung:

Erklärung zur Master-Thesis

Hiermit versichere ich, die vorliegende Master-Thesis ohne Hilfe Dritter nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Darmstadt, den 4. Januar 2016

(Jan Kohlschütter)

Abstract

Identifying movement abnormalities from raw Electromyography (EMG) data requires three steps that are the data pre-processing, the feature extraction and training a classifier. As EMG data shows large variation (even for consecutive trials in a single subject) probabilistic classifiers like naive Bayes or probabilistic support vector machines have been proposed. The used feature representations (e.g., PCA, NMF, wavelet transformation) however, can not capture the variation. Here, we propose a fully Bayesian approach where both, the features and the classifier, are probabilistic models. The generative model reproduces the observed variance in the EMG data, provides an estimate of the reliability of the predictions and can be applied in combination with dimensionality reduction techniques such as PCA and NMF. We found the optimal number of components and Gaussians for each model and tuned their meta-parameters. Besides the the focus on the four EMG channels, we tested the knee angle alone and EMG channels with the knee angle. We found that these probabilistic extensions outperforms classical approaches in terms of the prediction of knee abnormalities from few samples. We also show that the robustness against noise of the proposed probabilist model is superior than classical methods.

Zusammenfassung

Das Feststellen von Bewegungsstörungen anhand Elektromyographie-(EMG) Rohdaten benötigt drei grundlegende Schritte. Diese bestehen aus dem Vorverarbeiten der Daten, der Feature-extraction und dem Training eines Klassifizierers. Da die EMG Daten große Schwankungen aufweissen (selbst bei aufeinanderfolgenden Versuchen mit der gleichen Person) werden hierfür probabilistische Klassifizierer wie zum Beispiel Naive Bayes oder die probabilistischen Support Vector Machines angewendet. Die oft verwendeten Feature Representationen (z.b. PCA, NMF, wavelet transformation) können diese Schwankungen jedoch nicht verarbeiten. Deswegen schlagen wir einen vollen Bayes Ansatz vor, in dem sowohl die Features als auch der Klassifizierer probabilistische Modelle sind. Diese generativen Modelle können die festgestellte Varianz der EMG Daten nachbilden und liefert damit eine Abschätzung für die Zuverlässigkeit von Vorraussagen und kann darüber hinaus in Kombination mit dimensions reduzierenden Techniken wie PCA und NMF angewendet werden. Wir haben für jedes Model die optimale Anzahl von Komponenten und Gaussians ermittelt und ihre Meta-Parameter optimiert. Neben dem Fokus auf die vier EMG Kanäle, haben wir auch den Winkel des Kniegelenk alleine und die EMG Kanäle zusammen mit dem Winkel des Kniegelenk untersucht. Wir konnten sehen, dass diese probabilistische Erweiterung besser geeignet ist für die Vorraussage von Kniestörungen als die klassischen Ansätze. Wir konnten auch zeigen, dass das vorgeschlagene probabilistische Model robuster gegenüber Störungen ist als die klassischen Methoden.

Contents

Figures and Tables

List of Figures

List of Tables

1 Introduction

Making medical diagnoses is challenging. Doctors might have a different angle on a certain topic, their examination time is limited or comparative studies are not accessible. If the judgment is wrong, it can cause severe consequences for a patient. To support the decision process computer-aided systems therefore aim at providing additional insights by making use of large *but noisy* data sets. In this work, we propose a *probabilistic* model that can be trained from EMG data, models the noise and is used to predict movement abnormalities.

Electromyography (EMG) signals are recorded electric signals resulting from the activation of muscle cells [1]. There is a wide spread use of EMG data in research disciplines such as muscle surgery [2] by showing changes in abdominal muscle activity after upper abdominal surgery. In neurology [3] in diagnosis of neuromuscular disorders and rehabilitation [4, 5] by building prostheses for amputees and using EMG signals for on-line and realtime control of finger position and force or simply for monitoring the performance levels of rehabilitation patients. As well in movement analysis [6, 7] developing a robot hand controlled by EMG signals or biomechanics [8, 9] trying to predict motor functions after neural injuries and in ergonomics as risk prevention [10]. A computational model that can be applied to these different disciplines needs to implement two important features. First, the ability to compute reliable predictions to group subjects and second, to analyze EMG signal similarities in a lower-dimensional and thus easy to visualize feature space.

Figure 1.1.: Example of a prosthetic hand [11]

For classification neural networks [12, 13, 14], linear discriminant analysis, kernel based methods [15], or support vector machines [16] have been proposed. For dimensionality reduction, principal components analysis (PCA) [17], non negative matrix factorization (NMF) [18] or wavelet transformation [19] were investigated. While probabilistic classifiers demonstrated to be robust in terms of signal noise [20], currently used feature representations can not reproduce the omnipresent EMG signal variation. Solely the mean of the EMGs signals is reproduced and a large quantity of the entropy is lost through averaging.
We propose a probabilistic EMG model that captures the mean and the covariance of multiple EMG channels. The model learns a distribution over the signals which can be used either directly in a naive Bayes classifier (we refer to this model as mixture model) or PCA and NMF are applied to classify EMG trials in a lower-dimensional feature space, see Figure 1.2. PCA and NMF are often used when working with EMG data. Either directly on the signals or after feature extraction like wavelet transformation. Here we apply the dimensionality transformation by PCA and NMF to probabilistic features, and as such we present 3 alternatives. We can show that the mixture model is the best choice and dimensionality reduction yields no further improvement on this simple data set. Together, the mixture model and the probabilistic feature space variants of PCA and NMF are the contributions of this work.

1.1 Related work

There have been many studies trying to utilize EMG data for various applications. Different methods have been used for classification or decomposition and dimensionality reduction.

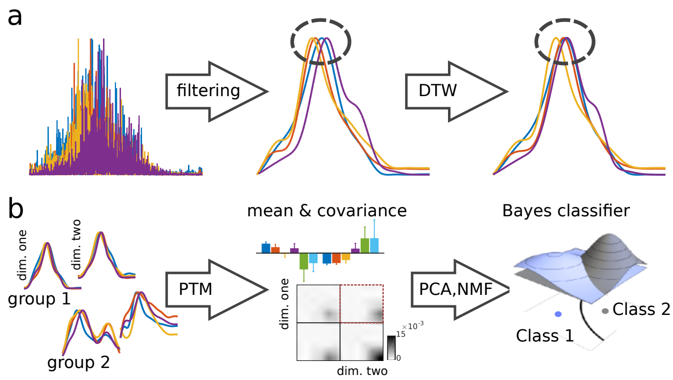


Figure 1.2.: Concept of the probabilistic EMG model: (a) EMG signals are rectified, low-pass filtered and optionally Dynamic Time Warping (DTM) is applied to correct for varying initial velocities in the trials. (b) For individual groups, the aligned EMG-channels (i.e., the first two EMG channels for subjects with an without knee abnormalities are illustrated) are mapped to the features space using a Probabilistic Trajectory Model (PTM). This model captures the mean and the covariance of the features. The statistics can be used directly in a naive Bayes Classifier (the mixture model approach) or dimensionality reduction techniques such as PCA or NMF are applied beforehand, i.e., the proposed probabilistic feature space variants of PCA and NMF.

1.1.1 Classification of EMG signals

Neural networks (NN) can learn the relationship between the EMG pattern and the actual finger movement [12]. This was demonstrated with recorded EMG data from hands of persons and the conclusion was that learning linear separation functions with perceptrons are poorly classifying EMG patterns when they are not linear separable. This could be improved by the use of NN due to good learnability, adaptability and non-linear separability. Recognition of EMG patterns of NN is superior then methods using linear separation functions.

The classification of EMG signals needs some basic steps of preprocessing. The signal is filtered, rectified, normalized and finally down sampled. The data is then normalized, for each muscle its maximum activity will be set to one and its minimum activity to zero. Such a normalization procedure can be adopted to give an equal importance to all the muscles and also to preserve information concerning differences in EMG amplitude between patients and control.

After collecting EMG recordings from the lower limbs of patients with arthritis several classifiers had been tested to discriminate them from healthy subjects [15]. Working with raw EMG data, a comparison of leastsquares kernel (LSK) algorithms, neural network algorithms like the Kohonen self organizing map, learning vector quantification, the multilayer perceptron and linear discriminant analysis (LDA) have been used. To identify the muscles that were critical for the classification, the classification rate was tested again after deleting one muscle at a time from each of the classes. The proposed LSK method had the best classification results and to ensure that the successful performance was not due to the sampling methods used, k-fold cross validation has been applied.

Another use of an artificial neural network (ANN) is the classification of Motor unit action potentials (MUAPs) [13]. The decomposition of raw EMG signals into their constituent MUAPs and their classification into groups of similar shapes is a typical case of an unsupervised learning pattern recognition problem. The number of MUAP classes composing the EMG signal, the number of MUAPs per class, their firing pattern, and the expected shape of the MUAP waveforms are unknown beforehand. Even worse to handle is the variability of the MUAP waveform, jitter of single fiber potentials, and MUAP superpositions, which is why we propose a probabilistic EMG model to capture the mean and the covariance.

The MUAPs clasification was also tested with a statistical pattern recognition technique based on the Euclidean distance, which we used as a deterministic classification approach opposing the probabilistic model.

1.1.2 Decomposition and dimensionality reduction of EMG signals

An algorithm for the decomposition of EMG signals should consists of four processing stages: segmentation, wavelet transformation, PCA, and clustering [17]. For clinical interests, the main feature of the EMG signal is the number of active motor unit (MUs) and the MUAP waveforms. In detail the decomposition of the EMG signal was done by wavelet spectrum matching and principle component analysis of wavelet coefficients. To classify the EMG signal, the Euclidean distance between the MUAP waveforms was used.

Before classifying various types of movements, further preprocessing can be applied. After recording, Surface EMG (SEMG) signals can be decomposed by wavelet packet transform (WPD). The resulting feature space obtained by the WPD decomposition has a relatively high dimensionality, hence why PCA is applied. The low dimensional features then form the input space for a neural network classifier, which can e.g. discriminate between four types of prosthesis movement [14].

Another approach is the use of discrete wavelet transform (DWT) where the EMG signals are decomposed into different frequency bands. Then statistical features are extracted from these subband decomposed EMG signals. The resulting characteristics of the EMG waveform can be used for diagnosing patients with neuromuscular disorders. A healthy patient can be classified against patients with various forms of the disorder like Myopathic (where the disorder is in the muscle cells) and Neurogenic (where the disorder comes from the nerve cells). Different classification methods have been tested, like k-nearest-neighbor, Radial basis function networks and Support vector machines (SVM). Optimizing the good results from the SVM, the use of Particle swarm optimization SVM (PSO-SVM) was proposed [16].

Other common practices after recording the signals are amplification, filtering (using 2nd order Butterworth filter), sampling and segmentation [21]. Instead of raw signals various feature extraction methods can be applied, such as time series analysis (AR, MA, ARMA), Wavelet Transform (WT), Discrete Wavelet Transform (DWT) Wavelet Packet Transform (WPT), Fast Fourier Transform (FFT), Discrete Fourier Transform (DFT).The resulting features can then be used to classify EMG signal patterns.

Wavelet transformation, which is a time-frequency transformation, as a preprocessing step can also be used to reduce noise or to evaluate the energy of the signal [22]. The recorded EMG signals of multiple muscles can be described in matrix form. PCA, NMF and Gain Shape k-Means can then be used to gain a low dimensional approximation of the input matrix. A direct influence on the EMG signals can be measured with stroke patients, because the loss of brain functionality leads to restricted muscle activity on the affected body side. Using EMG signals from the limbs, it is possible to identify the latent dimensionality of EMG data and derive a criterion for the health status of stroke patients. Since PCA and NMF are methods for data decompositions, they can be used to analyze synergies in EMG signals, when different kinds of muscles are activated at the same time to produce a movement. The transformation matrix gained from mentioned methods can be used as explanation and visualization. Each of the column vectors represents a synergy and each row refers to the corresponding muscle. The value indicates how much this muscle contributes to the signal for that synergy. A final comparison found different patterns in healthy and stroke patients and was based on the found synergies.

1.2 Outlook

In Section 2 we give a brief overview about the fundamentals of EMG signals and showing how challenging it is to work with them. Followed by an introduction into commonly used techniques for EMG analysis. Like non-negative matrix factorization and principal component analysis. We discuss their physiological interpretation and how they can be visualized. Section 3 introduces a probabilistic model for EMG signals and how a distribution over EMG signals can be modeled. The proposed probabilistic model is complemented by descriptions of the deterministic methods used for comparisons. Furthermore the used classifiers are presented, along with the quality measures. Finally the relationship to muscle synergies is discussed. The results are shown in Section 4 that range from the description of the data set and its preprocessing, to finding the optimal parameters for all used methods. The performance for all models is evaluated in two real-world data sets, where in contrast their robustness against noise is shown. In the end the principal components are shown. Concluding with Section 5 to discuss which parts of this thesis can be studied further in future work.

2 Background about EMG signals and EMG analysis

2.1 Electromyography signals

Electromyography (EMG) can detect and record electric signals resulting from the activation of muscle cells. EMG can describe the behavior of what muscles are doing, by measuring their muscle performance, help making decisions before and after surgeries or documentation of treatment and training improvements. [1]
EMG can measure the action potential of single motorneurons, which are nerve cells located in the spinal cord. Their axons form the efferent connection from the brain to the muscle fibers to transmit signals for the contraction or relaxing of the muscle, Figure 2.1.
A raw EMG signal can be obtained in two different ways, either as surface EMG above the muscle on the skin or by invasive methods directly from the muscle fiber membrane [1].
Because muscle activity consists not only of a single fiber the recorded EMG signal contains a superposition of magnitudes of involved fibers, also called Motor unit action potentials (MUAPs). Related work deals with decomposing those superposed signals into MUAPs (also Synergies) while others work directly on the EMG signals. Working with surface EMG has a few drawbacks since the signals goes a long way from the muscle membrane to the electrodes. External noise can alter the recordings as well as unwanted activity of neighboring muscles can change the result. [1]

2.2 What is the challenge working with EMG signals ?

Humans can move elegantly or gracefully through dynamic environments which requires the coordination of different body parts like limbs, muscles and neurons. But those are characteristics which can hardly be measured which is why its a challenge understanding how we move [23]. It becomes even more of a problem realizing that even by performing the same task over and over again, the motor activity will never be the exact same [24]. Yet a characteristic pattern can still arise even when a task is

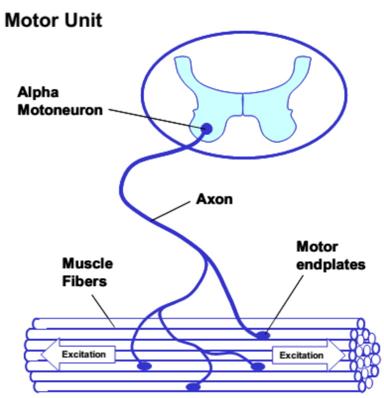


Figure 2.1.: Concept of a Motor Unit and the efferent path ways from the spinal cord to the muscles. From [1]

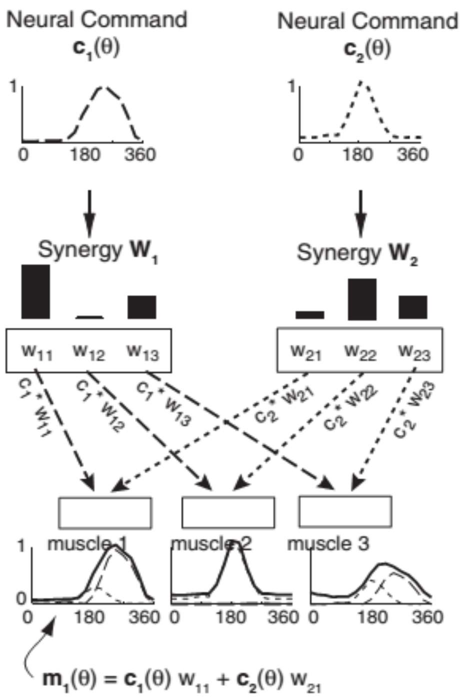


Figure 2.2.: Concept of Muscle Synergy models: the two synergies W_1 and W_2 and the weights c_1 and c_2 can reconstruct the activation patterns of all three given muscles after finding the synergies by decomposition. From [23]

performed by different sets of muscles, as in movements on varied surfaces. Still the performance of such a motor task can happen on a consistent basis even with huge amount of *variability* in the underlying body systems needed for that task [25] [26] [27] [28]. In other cases the synchrony observed in one movement might be canceled out in another. What may appear like a coordinated move, can be independent during another measurement [24] [29]. An explanation is that small changes in neural control have a noticeable effect on the muscles and the environment which is interacting with the body never stays exactly the same. Yet the coordination of motor activity is responsible for generating *predictable* biomechanical processes, like force generation and motion, allowing the reliable performance of motor tasks [24].

2.3 NMF and PCA for EMG signals

For the analysis and decomposition of EMG Signals in scientific fields such as motor control or neuroscience, two techniques have established themselves: principal components analysis (PCA) and non-negative matrix factorization (NMF) [23]. Other methods which could be used as decomposition techniques are independent components analysis (ICA) or k-means analysis [18].

2.3.1 What PCA and NMF have in common

PCA and NMF are both linear decomposition techniques which try to find linear combinations of basic elements in the measured data.
A observation $\boldsymbol{M_j(t)}$ could be represented as

$$\boldsymbol{M_j(t)} = c_{1j}\boldsymbol{W_1} + c_{2j}\boldsymbol{W_2} + \ldots + c_{nj}\boldsymbol{W_n} + error,$$

where $\boldsymbol{M_j}$ is a vector of measured activity of m muscles, the components $\boldsymbol{W}$, also of length m, is a vector representing invariant patterns of activity over different muscle recordings, which can be described by n

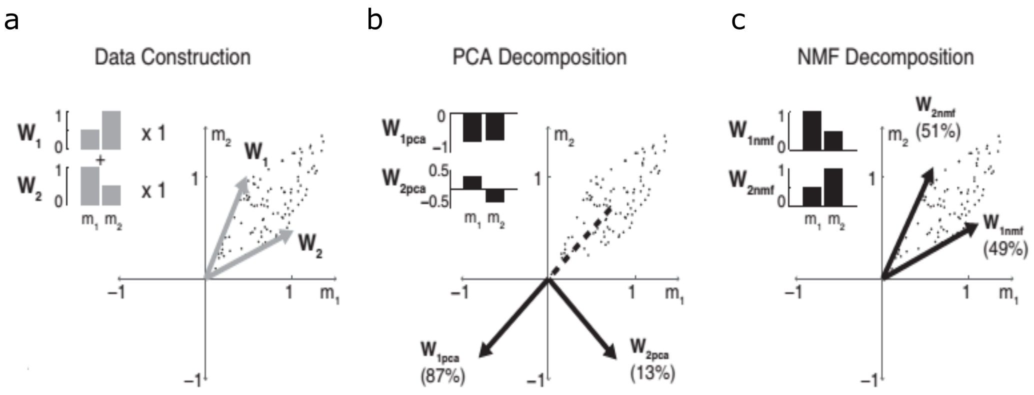


Figure 2.3.: Visualization of PCA and NMF (a) the orginal data (b) the describing PCA decomposition (c) the prescribing NMF decomposition. From [23]

scalar values c_{ij}, each specifying the weight of each component for the measured muscle activation pattern $\boldsymbol{M_j}$. The goal is to find a low dimensional representation where the m muscles can be expressed with $n < m$ components, where the basic elements $\boldsymbol{W_1}, \ldots, \boldsymbol{W_n}$ remain fixed. Only the factors c_{ij} can change and therefore absorb the variations of the data measured across different conditions [23]. The components $\boldsymbol{W_1}, \ldots, \boldsymbol{W_n}$ are often referred to as muscle synergies [30] [31] [32] [33] or M-modes [34] [35] [36]. Figure 2.2 shows the two synergies $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$ and the weights c_1 and c_2 which can reconstruct the activation patterns of all three given muscles after finding the synergies by decomposition.

2.3.2 Differences of PCA and NMF

The differences of principal components analysis (PCA) and non-negative matrix factorization (NMF) lie in the way each method decomposes the variability within a given data set. PCA is an analytical technique and requires the components to be orthogonal to each other, with a unique solution for any decomposition. While the resulting components for PCA are real numbers, the NMF features can only be positive. NMF is found by a search algorithm, starting from a set of random components and iteratively improving until the limit of iterations or a minimum error is reached. Repeated execution of the algorithm yields numerically different but still similar components. The factorization is a convex problem because NMF constrains the factors to be non negative. Components with such constraints can not be orthogonal but they must be independent [23].

2.3.3 Visualization of PCA and NMF: Describing vs. Prescribing

PCA *describes* the mean and the largest variance in each data set with its first principal component. Each following component describes the maximal possible variance in orthogonal direction. By allowing negative and positive values, two independent components can be scaled to reach any data point, even when facing in opposite direction of the data, see Figure 2.3 b). In NMF, the components *prescribe* a subspace. Due to the constraint of non-negativity, the scaling of two components is limited so that the data points can only reside inside of that subspace. Delimited by the components the data points are enclosed by a convex hull. Adding components increases the subspace, Figure 2.3 c) [23].

2.3.4 Physiological interpretation of PCA and NMF components

The PCA components $\boldsymbol{W}$, representing muscle activation patterns, and the weights c_{nj} can be positive and negative values. This representation contradicts an interpretation where the action potentials of

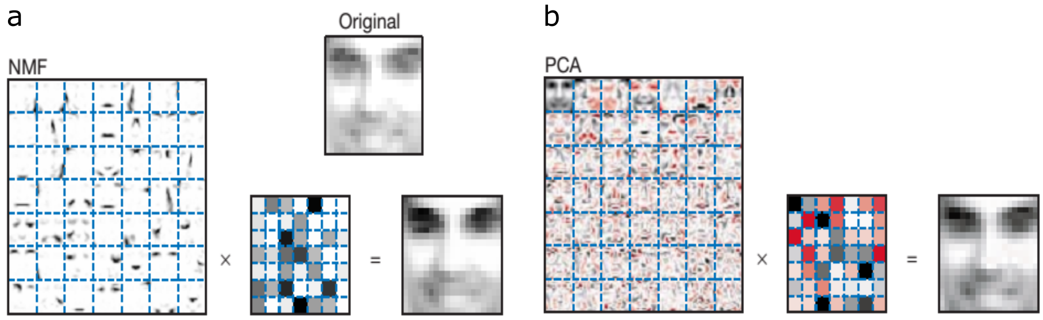


Figure 2.4.: Example for Parts-based vs. Holistic Decomposition From [37].

motorneurons control the muscle activity. Although motorneurons are capable of receiving inhibitory and excitatory neural signals, the inhibition can only be noticed with highly active muscles. Already inactive muscles can not process further inhibition and will remain in the same state. NMF is more physiological accurate due to the non-negativity constraint for its components. Neurons can fire action potentials which translates to a positive signal or remain dormant, waiting for further neural input [23].

2.3.5 Parts-based vs. holistic decomposition

From the non-negativity constraint in NMF comes another interesting topic. Each component of the decomposition equals only a specific part of the whole data. For a reconstruction all different parts must be summed. With PCA allowing negative values for its decomposition, a reconstruction would only be possible by addition and subtraction of different components. To illustrate the idea behind this, the example of decomposing faces is depicted in Figure 2.4. In a) the PCA components are all eigenfaces with different properties each. The reconstruction requires different characteristics to be added or subtracted from a mean face represented by the first principal component. In b) the NMF components are all parts of the face like eyes or mouth. Reconstructing a face would be done by selecting all required parts and scaling them to the proper size [23].

While the PCA components can change with different levels of muscle activity, NMF can identify components which remain stable over different conditions, while still allowing them to be combined in different ways, which can be seen as *robustness* [23]. This was shown during an experiment of postural body sway, where PCA identified components equivalent to the direction of center of pressure changes which is required to stabilize the body [38] [39]. In a different experiment concerning standing balance control, NMF components equal the direction of force applied to stabilize the body [32] [40]. With the addition of different conditions, the PCA components changed [35], while the NMF components remained consistent [40] [41].

Adding to the physiological interpretation, the NMF parts-based decomposition is similar to neural representations observed in the visual and other sensory encoding systems referring to it as *sparse-coding* [42]. The idea is that only a minimum of neurons are active in the particular encoding system using only parts of all available information to activate only certain parts of the body. While with PCA all the available information needs to be considered. The property of sparseness can also be applied to motor systems, reducing the energy consumption with less involved neurons or improving efficiency during motor adaptation [42] [43] [44].

2.3.6 Working with EMG signals

Only the coordination of muscles in various patterns can result in meaningful components. Hence why the number of recorded muscles should be high enough to gather different patterns of co-variation. It is also important to have sufficient varying conditions for the experiments, else some coordination patterns could be missed among the muscles. Independent muscle activation patterns will always be uncovered through component analysis.

The number of extractable components is limited by the number of recorded muscles. It is also critical if muscles are co-activated during certain conditions or not. If only a *few muscles are recorded*, it can result in two worst cases. They either form each a single synergy when all are independently activated, or they are always co-activated and form only one single muscle synergy.

Making sense of EMG analysis in a physiological context will always rely on the judgment and intuition of the researcher. The interpretation of gained components must depend on knowledge about physiological and bio-mechanical mechanisms. [23]. The component decompositions are not suitable in every study. But assuming similar basic principals apply, the comparison of complex muscle coordination with changing muscle activity over several task and trials fits the technique quite well, as seen in fast and slow walking [45], or one- and two-legged postural control [41]. Another possibility is to differentiate between patients with different EMG patterns having similar components with different activation and patients having additionally different compositions of components [33] [45].

2.3.7 PCA vs. NMF for EMG signals conclusion

Non-negative matrix factorization should be used when working with signals with neural pathways to induce muscle activity since those are non-negative by nature. Analyzing bio-mechanical mechanisms excluding muscle activity is the fitting application of principal component analysis due to positive and negative values [23].

3 A probabilistic model of EMG signals

Let $\boldsymbol{y}_t \in \mathbb{R}^D$ denote a D-dimensional column vector of, e.g., EMG measurements from D channels. The subscript t denotes a discrete time index. A sequence of T consecutive measurements is denoted by the matrix $\boldsymbol{Y} = \left[\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T\right]$ which is of the dimension $\mathbb{R}^{D\times T}$.
The goal of a computational model is to approximate the data $\boldsymbol{Y}$ through some function, i.e, $\tilde{\boldsymbol{Y}} = f(\boldsymbol{w})$. The vector $\boldsymbol{w}$ denotes a set of scalars that can be learned. For example classical principal components analysis (PCA) [17] and non-negative matrix factorization (NMF) [18] approximate $\boldsymbol{Y}$ through a reduced feature representation (encoded by $\boldsymbol{w}$) and classify unseen observations using a Mahalanobis distance measure on the reconstructed signals, see, e.g., [13, 46, 15]. We follow here a different approach and model the data $\boldsymbol{Y}$ as a probability distribution.

3.1 Modeling a distribution over EMG signals

We use a multinomial distribution to model the output vector of a function approximator

$$\boldsymbol{o} = \left[\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \ldots, \boldsymbol{y}_T^T\right]^T \quad \in \mathbb{R}^{D\cdot T\times 1} \quad ,$$

where the upper scrip ^{T} denotes the transpose operation and must not be confused with the number of discrete time steps T. As in [47] we use a Gaussian mixture model approach to represent the vector of concatenated EMG measurements $\boldsymbol{o}$ with

$$p(\boldsymbol{o}|\boldsymbol{w}) = \mathcal{N}(\boldsymbol{o}|\boldsymbol{\Omega}\,\boldsymbol{w}, \tilde{\Sigma}_y) = \prod_{t=1}^{T} \mathcal{N}(\boldsymbol{y}_t|\boldsymbol{\Psi}_t\,\boldsymbol{w}, \boldsymbol{\Sigma}_y) \quad . \tag{3.1}$$

The matrix $\boldsymbol{\Omega} \in \mathbb{R}^{T\cdot D\times D\cdot K}$ is a concatenation of T block diagonal matrices (K is the number of Gaussian basis functions introduced later), where $\boldsymbol{\Omega} = [\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \ldots, \boldsymbol{\Psi}_T]$. The block diagonal matrix $\boldsymbol{\Psi}_t \in \mathbb{R}^{D\times D\cdot K}$ is a clever arrangement of basis function vectors for multi-dimensional data,

$$\boldsymbol{\Psi}_t = \left\{\begin{matrix} \boldsymbol{\phi}_{t,1}^T & 0 & \ldots & 0 \\ 0 & \boldsymbol{\phi}_{t,2}^T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \boldsymbol{\phi}_{t,D}^T \end{matrix}\right\} \quad .$$

For each dimension denoted by i we use a vector of K scalar basis functions, i.e., $\boldsymbol{\phi}_{t,i}^T = \left[\phi_{t,1}, \phi_{t,2}, \ldots, \phi_{t,K}\right]^T$. A popular choice for rhythmic movements are Von-Mises basis functions [48], whereas for point to point movements Gaussian basis functions are widely used [47],

$$\phi_{t,k} = \frac{\exp\left(-0.5(t - c_k)^2\right)}{\sum_{k=1}^{K} \exp\left(-0.5(t - c_k)^2\right)} \quad .$$

Usually, the means (denoted by c_k) and the variances of the Gaussian features are kept fixed and only the parameter vector $\boldsymbol{w} \in \mathbb{R}^{D\cdot K\times 1}$ in (3.1) is learned. This parameter vector is a concatenated vector of D feature vectors, one per dimension, where $\boldsymbol{w} = \left[\boldsymbol{w}_1^T, \boldsymbol{w}_2^T, \ldots, \boldsymbol{w}_D^T\right]^T$. Note that we omitted the variances in this notation for the sake of brevity.
The covariance matrix $\boldsymbol{\Sigma}_y$ in (3.1) denotes the measurement noise. We assume Zero mean Gaussian noise where $\boldsymbol{y}_t = \boldsymbol{\Psi}_t\,\boldsymbol{w} + \boldsymbol{\epsilon}_y$ where $\boldsymbol{\epsilon}_y$ is sampled from $\boldsymbol{\epsilon}_y \sim \mathcal{N}(\boldsymbol{\epsilon}_y|\boldsymbol{0}, \boldsymbol{\Sigma}_y)$.

3.2 Learning of EMG models

In (3.1) we assumed that the parameter vector $\boldsymbol{w}$ is known. Now for learning the vector $\boldsymbol{w}$ we introduce a prior distribution $p(w)$. This prior is in the simplest case a Gaussian distribution,

$$p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}|\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w) \ , \tag{3.2}$$

where the generative probabilistic model can be computed in closed form, i.e.,

$$\begin{aligned} p(\boldsymbol{o}) &= \int p(\boldsymbol{o}|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w} \\ &= \mathcal{N}(\boldsymbol{o}|\boldsymbol{\Omega w}, \boldsymbol{\Omega}\boldsymbol{\Sigma}_w\boldsymbol{\Omega}^T + \tilde{\boldsymbol{\Sigma}}_y) \ . \end{aligned} \tag{3.3}$$

The prior is used to model a distribution over multiple recordings $\boldsymbol{o}^{[m]}$, where m denotes the m-th trial or sample of recorded multi-dimensional EMG signals. Usually the mean $\boldsymbol{\mu}_w$ and covariance matrix $\boldsymbol{\Sigma}_w$ are learned from the data by maximum likelihood with help of, e.g., the Expectation Maximization algorithm as in [49], which generalizes to more complex hierarchical prior distributions. For our Gaussian prior a much simpler approach based on least squares regression was proposed [47], i.e.,

$$\boldsymbol{w}^{[m]} = (\boldsymbol{\Omega}^T\boldsymbol{\Omega} + \lambda \boldsymbol{I})^{-1}\boldsymbol{\Omega}^T\boldsymbol{o}^{[m]} \ . \tag{3.4}$$

The scalar λ denotes a regularization term that is typically set to a small value (we used $1e-6$). The mean and the covariance of $p(\boldsymbol{w})$ can be estimated by the sample mean and sample covariance of the $\boldsymbol{w}^{[m]}$'s.

3.2.1 Deterministic models

The raw trajectories are given by the matrix $\boldsymbol{O} = \left[\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \dots, \boldsymbol{y}_T^T\right]^T$ and are of the dimension $\mathbb{R}^{D \cdot T \times M}$ where $m = 1, ..., M$ is the number of samples. This matrix is then used for all deterministic models.

Non Negative Matrix Factorization - NMF

The implementation of the NMF algorithm is based on the paper "Algorithms, Initializations, and Convergence for the Non-negative Matrix Factorization" [50]. Calculating the decomposition of the original data matrix $\boldsymbol{O}$ results in two matrices, a non-negative data matrix $\boldsymbol{V} \in \mathbb{R}^{D \cdot T \times C}$ and a weight matrix $\boldsymbol{H} \in \mathbb{R}^{C \cdot M}$, where C denotes the number of components for each model which needed to be manually set beforehand.

The iterative algorithm computes $\boldsymbol{V}$ and $\boldsymbol{H}$ in an alternating manner, starting usually with a random initialization of $\boldsymbol{V}$. For faster convergence the before mentioned paper proposes various other techniques. The chosen implementation uses the Random Acol Initialization, which averages 20 random columns of the original matrix $\boldsymbol{O}$ for each column of initial matrix $\boldsymbol{V}(0)$ used in the first iteration step. The idea is that this initialization is closer to a desired result than a complete random matrix. This implementation may also take advantage of regularization parameters, which can be used to control the sparsity of the matrices or even the sparsity in each column of the matrices. By setting the parameters to zero, the result matches the standard implementation of NMF (for detailed parameter settings, see Results section).

The model parameter is $\boldsymbol{V} \in \mathbb{R}^{D \cdot T \times C}$, where the reconstruction is given by $\tilde{\boldsymbol{O}} = \boldsymbol{V} \cdot H \in \mathbb{R}^{D \times T \times M}$.

Principal Component Analysis - PCA

Standard PCA is applied to the matrix $\boldsymbol{O}$ by subtracting the mean, dividing by the standard deviation, computing the eigenvectors and eigenvalues of the covariance matrix and selecting the chosen number of components. Then model matrix $\boldsymbol{V} \in \mathbb{R}^{M \times C}$ is saved and from the transformed data $\boldsymbol{Z} = \boldsymbol{O} \times \boldsymbol{V} \in \mathbb{R}^{D \cdot T \times C}$ the reconstructed matrix $\tilde{\boldsymbol{O}} = \boldsymbol{Z} \times \boldsymbol{V}^T \in \mathbb{R}^{D \cdot T \times M}$ can be computed, which gives a good approximation of the original raw data matrix $\boldsymbol{O}$, while gaining a reduction in dimensionality.

Wavelet Transformation

The raw trajectory matrix $\boldsymbol{O}$ is also used for standard Wavelet transformation. In this case a predefined level is chosen which, the higher the level, results in a decreasing number of components (see Result section for more details). The model matrix $\boldsymbol{V} \in \mathbb{R}^{D \cdot T \times C}$ and the reconstructed matrix $\tilde{\boldsymbol{O}} \in \mathbb{R}^{D \cdot T \times M}$ have been saved as model parameters.

3.2.2 Probabilistic models

For the probabilistic models, the matrix $\boldsymbol{W} \in \mathbb{R}^{D \cdot K \times M}$ is used. M denotes the number of trials where $\boldsymbol{W} = \left[w^{[1]}, w^{[2]}, \ldots, w^{[M]}\right]$. $N = DK$ denotes the number of features.

Mixture Model

Without dimensionality reduction the mean and the covariance of the prior distribution $p(\boldsymbol{w})$ can be estimated by computing the mean and the covariance of $\boldsymbol{W}$. We refer to this technique as *mixture model* in the results section.

Probabilistic non-negative matrix factorization - p-NMF

For the p-NMF model two approaches have been applied. First the decomposition of $\boldsymbol{W}$ into $\boldsymbol{V} \in \mathbb{R}^{D \cdot K \times C}$ and $\boldsymbol{H} \in \mathbb{R}^{C \cdot M}$ as described in the NMF section but with one difference. Due to the fact that the factors in the $\boldsymbol{W}$ matrix are not all non-negative, a standard implementation could not be used. Hence, we applied [50]. This proposition of the algorithm will set all negative values to zero in each iteration of each factor, which might seem odd but works very well in practice. In addition, instead of computing the prior statistics directly from $\boldsymbol{W}$, the mean and the covariance are computed from the approximation denoted by $\tilde{\boldsymbol{W}} = VH \in \mathbb{R}^{D \cdot K \times M}$.

Probabilistic principal components analysis - p-PCA

The p-PCA model follows the same principal as the p-NMF model. First the standard PCA is computed and thereby the desired number of components is chosen. Resulting in a low dimensional model of $\boldsymbol{W} \in \mathbb{R}^{D \cdot K \times M}$, given by $\boldsymbol{V} \in \mathbb{R}^{D \cdot K \times C}$, which can be used for the deterministic classifiers. While for the probabilistic classifiers, the mean and the covariance are computed from the approximation $\tilde{\boldsymbol{W}} \in \mathbb{R}^{D \cdot K \times M}$ as show before.

3.3 Classification with EMG models

Given either the data matrix $\boldsymbol{O}$ or $\boldsymbol{W}$, the goal in classification is assign a new observation to a discrete class label $\vec{c} \in \{0, 1\}$. In this binary representation $\vec{c} = 0$ represents class C_1 and $\vec{c} = 1$ class C_2. In our case this results in the classes healthy and ill. Deterministic as well as probabilistic classifiers were evaluated [51].

3.3.1 Deterministic

Five classifiers have been used to test the deterministic models. All are based on an euclidean distance measure. For NMF, the calculation is $\boldsymbol{h} = \boldsymbol{V}/\boldsymbol{y}$ where $\boldsymbol{V} \in \mathbb{R}^{D \cdot T \times C}$ is the model and $\boldsymbol{y} \in \mathbb{R}^{D \cdot T \times 1}$ is the test trajectory [46]. For PCA and Wavelet transformation $\boldsymbol{H} = \boldsymbol{y} * \boldsymbol{V}$. Here the test trajectory is $\boldsymbol{y} \in \mathbb{R}^{D \cdot T \times M}$ and the model is $\boldsymbol{V} \in \mathbb{R}^{M \times C}$.

The first classifier tries to exploit the concatenation of all channels choosing the class with the smaller distance calculated by the norm of $\boldsymbol{h}$ or $\boldsymbol{H}$ respectively. The classifiers two through five use each single channel for the calculation. For NMF $\boldsymbol{h} = \boldsymbol{V}/\boldsymbol{y}$ where $\boldsymbol{V} \in \mathbb{R}^{T \times C}$ and $\boldsymbol{y} \in \mathbb{R}^{T \times 1}$. And for PCA and Wavelet $\boldsymbol{H} = \boldsymbol{y} * \boldsymbol{V}$ where $\boldsymbol{y} \in \mathbb{R}^{T \times M}$ and $\boldsymbol{V} \in \mathbb{R}^{M \times C}$.

Due to an even number of channels for the EMG data set, a majority vote is not sufficient. That is why the second classifier finds the smaller distance and weights the found class by its distance to make a collective decision. The third classifier will calculate the mean of all distances and then find the smallest mean distance. The fourth classifier calculates the minimum and finds the smallest minimum distance. And the fifth classifier calculates the maximum to find the smallest maximum distance.

3.3.2 Probabilistic

The learned prior distribution over EMG recordings in (3.2) can be used in a naive Bayes classifier,

$$p(l|\boldsymbol{w}^*) = \frac{\mathcal{N}(\boldsymbol{w}^*|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\,\alpha_l}{\sum_{l'=1}^{L} \alpha'_l \mathcal{N}(\boldsymbol{w}^*|\boldsymbol{\mu}'_l, \boldsymbol{\Sigma}'_l)} \quad ,$$

where l denotes the cluster index and $\boldsymbol{w}^*$ is the feature vector under test which was obtained through applying (3.4) on a unseen test trail. The scalar α_l denotes the cluster prior weight. It can account for a different number samples per cluster. In our experiments, we used balanced training sets, where for both groups, subjects with and without knee abnormalities, an equal number of trials were selected ($\alpha_1 = \alpha_2 = 0.5$).

Due to the two approaches both the deterministic and the probabilistic classifiers could be used for the p-NMF and p-PCA models, since only the dimensions change from $D \cdot T$ to $D \cdot K$, while NMF, PCA and Wavelet could only use the deterministic and the MM could only use the probabilistic classifier.

3.4 Relationship to muscle synergy models

The generative probabilistic model in (3.3) can be related to time-invariant [52] and time-varying [53] muscle synergy models. Time-invariant synergies are represented as a set of shared synergy vectors v that is scaled by task dependent time-varying temporal profiles

$$\boldsymbol{y}_t^m = \sum_{k=1}^{K} \alpha_k^m(t)\, v_k \quad .$$

Time-varying muscle synergy models [53] generate EMG measurements as weighted sum over time-shifted synergy profiles

$$\boldsymbol{y}_t^m = \sum_{k=1}^{K} \alpha_k^m\, v_k(t - t_k^m) \quad ,$$

where the activity vector $v_k(t)$ is shared among m tasks. For simplicity we assumed here equal activations α_k^m and time shifts t_k^m for a D-dimensional vector v_k.
Both generative laws relate to a single time step prediction in (3.1), where the basis function matrix $\boldsymbol{\Psi}_t$ is shared among tasks (like v_k) and the learnable feature vector $\boldsymbol{w}$ becomes task dependent. Such task dependent feature weights were used in [49] for transfer learning. Note that time shifts and task dependent activity vectors v_k^m as used in temporal components can not be modeled in this formulation.

In summary, the proposed model provides a probabilistic formulation of well established muscle synergy models [53, 52]. However, it utilizes a linear basis function approach where the model parameters can be learned *in a single step* (through least squares regression) in contrast to the *iterative approaches* used in [53, 52].

3.5 Quality measures

All tests in the results section have been conducted with $N = 20$ fold cross validation by splitting the bimodal data set into a training and test set (by drawing samples without replacement). All methods were evaluated on the test set using three quality criteria: success rate of classification, f-score and explained variance.

3.5.1 Classification success

The classification measure is simply the ratio of correct classified to wrong classified test samples, i.e. $(\sum_{i=1}^{N} c_i)/N$, where the c_i are the correct assigned class labels and N is the number of folds.

3.5.2 F-score

The $F-score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$ with $Precision = TP/TP + FP$ and $Recall = TP/TP + FN$ where $TP = TruePositive$, $FP = FalsePositive$ and $FN = FalseNegative$ [54].

3.5.3 Explained Variance

For the PCA and p-PCA models, given the matrix of eigenvalues $\boldsymbol{e}$, the explained variance is $\frac{cumsum(\boldsymbol{eVal})}{sum(\boldsymbol{eVal})}$.

The NMF and the Wavelet transformation models used the variance accounted for (VAF). The measure is defined by $VAF = 1 - \frac{\boldsymbol{O} - \tilde{\boldsymbol{O}}}{\boldsymbol{O}}$, where $\boldsymbol{O}$ are the raw trajectories and $\tilde{\boldsymbol{O}}$ denotes the approximation matrix.

For p-NMF and the MM model, the definition of the measure changes to $VAF = 1 - \frac{\boldsymbol{W} - \tilde{\boldsymbol{W}}}{\boldsymbol{W}}$, where $\boldsymbol{W}$ are the learned parameters and $\tilde{\boldsymbol{W}}$ denotes the approximation matrix of the learned parameters [55].

4 Performance of the probabilistic model of EMG signals

4.1 Data and preprocessing

We evaluated the proposed EMG models on a clinical lower limb data set [56], where 22 subjects had to perform two exercises. In the first, the subjects were instructed to fully flex their knee while sitting. In the second set of continuously recorded repetitions, the subjects had to stand up from the sitting position. Prior to the exercises, a professional diagnosed for 11 subjects some form of knee abnormalities.
For each subject the knee angle and four EMG-channels (rectus femoris, biceps femoris, vastus internus, and semitendinosus) were recorded in two to six repetitions. Excluding the first and last trial, we could manually extracted about two to three trajectories per subject, which resulted in 30 samples for each of the two exercises. The EMG data was rectified and low pass filtered with a 4 pass Butterworth filter with sampling frequency of 1000 Hz and a cutoff frequency of 2/1000 Hz and eventually normalized. The normalization was also applied to the knee flexion and consisted of two steps, first removing the baseline by subtraction of the minimum value of each channel and then divide through the maximum activation value of each channel.

4.1.1 Dynamic Time Warping

Due to motor variability among the patients and over several trials, all the gathered trajectories of the knee angles and the muscles signals had various lengths. The manual on- and offset determination resulted in slightly different samples. For a better comparison an alignment over all repetitions and over all patients was needed. For that the Dynamic Time Warping algorithm (DTW) [57] was used on the knee flexion. First the trajectory closest to the average length was determined and used as reference trajectory. Then the data was scaled down from about $3000 - 11000$ steps to $T = 300$ time steps. DTW then aligned the trajectories of all repetitions from all patients to the chosen reference trajectory. The resulting time mapping from original to aligned trajectory was then applied to the EMG signals of the corresponding repetition for each patient, see Figure 4.1 and 4.2.
Four samples were used for testing in cross validation with 20 sets.

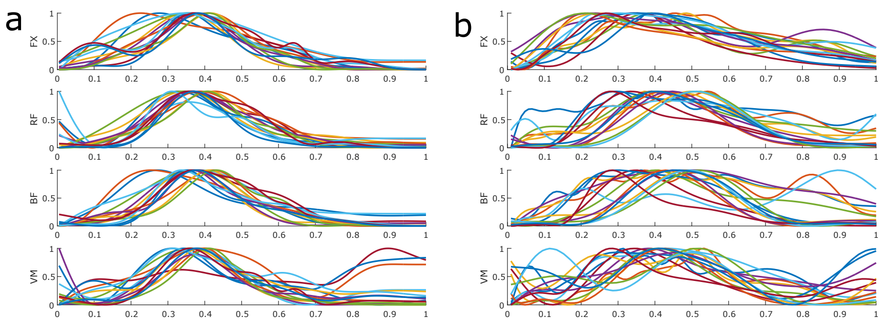


Figure 4.1.: Aligned Trajectories for the knee flexion while sitting exercise (a) Healthy and (b) Ill for all muscle recodrings and the knee angle.

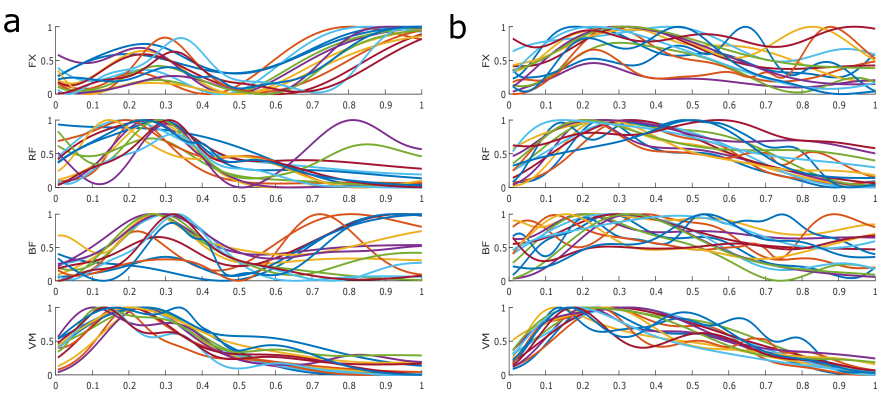


Figure 4.2.: Aligned Trajectories for the standing up exercise (a) Healthy and (b) Ill for all muscle recodrings and the knee angle.

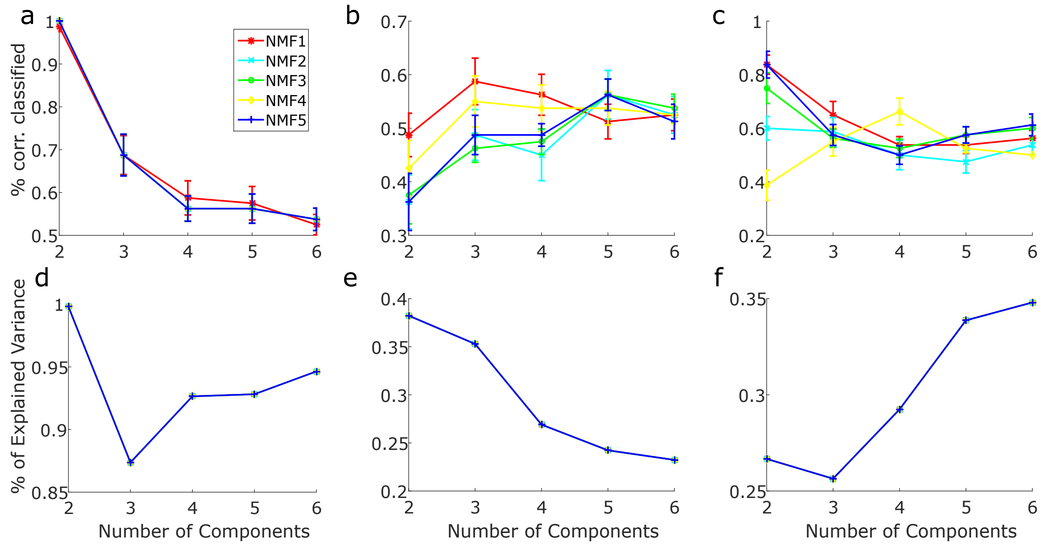


Figure 4.3.: Optimal number of components for model NMF on data set for the knee flexion while sitting exercise NMF1-5 show the different classifier (a) Knee angle only. (b) All four muscle recordings. (c) Knee angle and the four muscles. (d-f) corresponding explained variance of the data sets in (a-c).

4.2 Finding the optimal number of components and classifier

Before comparing the models to each other a test was conducted to identify the optimal number of components and model parameters.

4.2.1 Non-negative matrix factorization - NMF

The optimal number of components was tested first.

Optimal number of components

For the data set with one channel of the knee degree of the knee flexion while sitting exercise show in Figure 4.3 a) all non concatenated classifiers have the same performance, while two components are

	Components	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	2	NMF2-5	1.0000	1.0000	0.9981
EMG	3	NMF1	0.5875	0.5600	0.3529
Knee Degree + EMG	2	NMF1,5	0.8375	0.8434	0.2667

Table 4.1.: Comparison of NMF components and Classifiers on data set for the knee flexion while sitting exercise, where NMF1 is concatenated while the rest tests each dimension separately to combine the result as weighted (NMF2), mean (NMF3), minimum (NMF4) and maximum (NMF5)

	Components	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	3	NMF1	0.8625	0.8736	0.7793
EMG	3	NMF2	0.5500	0.5610	0.6574
Knee Degree + EMG	3	NMF4	0.6875	0.6032	0.5874

Table 4.2.: Comparison of NMF components and Classifiers on data set for the standing up exercise, where NMF1 is concatenated while the rest tests each dimension separately to combine the result as weighted (NMF2), mean (NMF3), minimum (NMF4) and maximum (NMF5)

the optimal choice. On the standing up exercise data set, Figure B.1 a), the concatenated has the best performance with three components (for one dimension the performance should be the same between concatenated and non concatenated but due to separate training of the models NMF can yield slightly different results, all non concatenated classifiers share the same model and yield the same performance and both the concatenated and non concatenated have the same percentage of explained variance).
Using the EMG data with four channels of the knee flexion while sitting exercise in Figure 4.3 b) the performance of all classifiers regardless the number of components seems to be around 50 Percent with the concatenated scoring the best rate with almost 60 Percent. The standing up exercise data set, Figure B.1 b) has a similar poor performance, but an overall higher explained variance (avg. of 63 vs. 30 Percent, Figure 4.3-B.1 e)).
The combination of both sets to a data set with five channels for the knee flexion while sitting exercise, Figure 4.3 c), shows a combination of the single results as the knee degree seems to strengthen the overall prediction. The standing up exercise data, Figure B.1 c) confirms that observation since the knee degree once again improves the combined result. For both data sets it can be noted that, a higher number of channels, decreases the explained variance, Figure 4.3-B.1 f).

Optimal number of parameters

The implementation of the NMF algorithm has the two parameters, the $lambda-W$ and $thelambda-H \in \{0,1\}$ which increase the sparsity of both decomposition matrices with increasing lambda values. Additional the $alpha-W$ and $alpha-H \in \{0,1\}$ parameters control the sparsity in each column of the decomposition matrices.
Clearly visible in Figure 4.4 (8) is that the last parameter set, Set seven from Table 4.3, can not produce any results better than chance because both classes are not distinguishable anymore. That is why this set was not used in the following tests. Also visible is that Figure 4.4 b)(7) produce instabilities and simply collapses, yielding a prediction of 50 Percent.

The data set of one channel for the knee flexion while sitting exercise had a perfect result of 100 percent correctly classified tests with the standard implementation of the NMF algorithm and two components, Figure 4.3 a). Changes of the parameters resulted in a similar classification performance except for parameter set six which can not discriminate between both classes anymore, see Figure 4.5 a). For the standing up exercise data set the standard NMF with three components had a good classification rate,

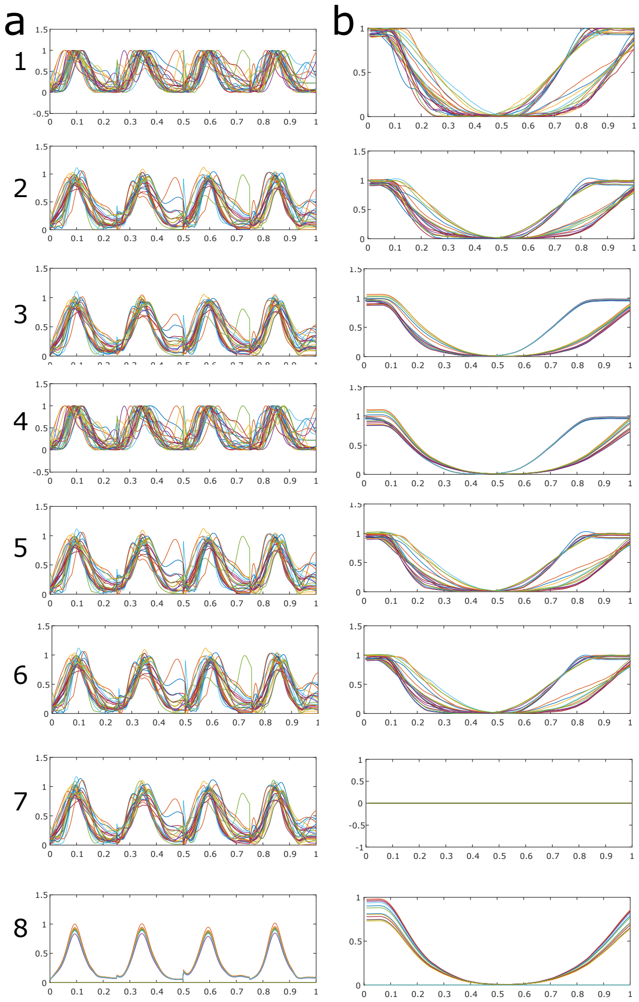


Figure 4.4.: Visualization of the orginal trajectories (1) and the reconstructions for parameter sets 1-7 depicted as 2-8 (a) EMG data concatenated (b) Knee Angle

	Lambda-W	Lambda-H	Alpha-W	Alpha-H
Set 1	0	0	0	0
Set 2	0.5	0.5	0	0
Set 3	1	1	0	0
Set 4	0	0	0.5	0.5
Set 5	0	0	1	1
Set 6	0.5	0.5	0.5	0.5
Set 7	1	1	1	1

Table 4.3.: All sets of tested parameter settings

which could be tweaked slightly by using parameter set four (plus 2.5 percent from Figure B.1 a) to B.2 a)).

The set of four channels for the knee flexion while sitting exercise performs very poorly with the standard NMF, but adjusting the parameters yields a perfect classification with parameter set two and six, which is an improvement of more than 40 percent, Figure 4.5 b). The same can be seen with the standing up exercise data set, Figure B.2 b), while the standard NMF performance is not much better than chance, parameter set 2 improves the results by almost 70 percent (a plus of 14 percent.

In the data set with all five channels for the knee flexion while sitting exercise once again the strong results from the one dimensional knee angle data set are reflected in the results of the combination of both sets for two components. All tested parameters maintain a high classification rate with the optimal set number three achieving 90 percent correctly classified (a plus of 8.75 percent, see Figure 4.3 c) and 4.5 c)). The standing up exercise data set can not profit from the parameter optimization and stays about the same level with all different sets of parameters, see Figure B.2 c).

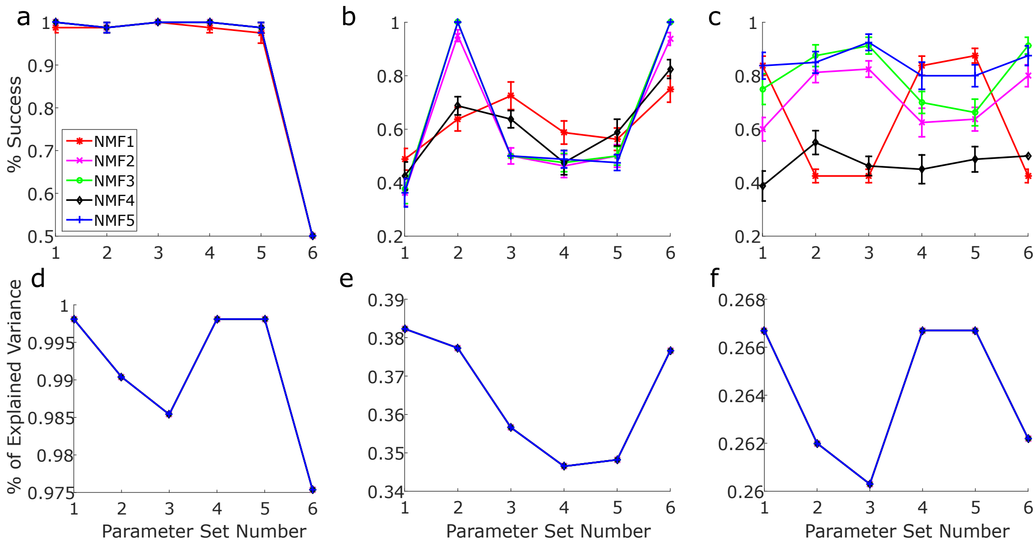


Figure 4.5.: Optimal parameter settings with best determined number of components for model NMF on data set for the knee flexion while sitting exercise NMF1-5 show the different classifier (a) Knee Angle only with two components. (b) All four muscle recordings with three components. (c) Knee Angle and the four muscles with two components. (d-f) corresponding explained variance of the parameter sets in (a-c).

	Components	Best Parameter Set	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	2	Set 3	NMF1-5	1	1	0.9854
EMG	3	Set 2,6	NMF3,5	1.0000	1.0000	0.37
Knee Degree + EMG	2	Set 3	NMF5	0.9250	0.9189	0.2603

Table 4.4.: Comparison of NMF parameters sets on data set for the knee flexion while sitting exercise, where NMF1 is concatenated while the rest tests each dimension separately to combine the result as weighted (NMF2), mean (NMF3), minimum (NMF4) and maximum (NMF5)

	Components	Best Parameter Set	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	3	Set 4	NMF2-5	0.8875	0.8861	0.8082
EMG	3	Set 2	NMF5	0.6875	0.5455	0.7022
Knee Degree + EMG	3	Set 2	NMF5	0.6750	0.5185	0.5957

Table 4.5.: Comparison of NMF parameter Sets on data set for the standing up exercise, where NMF1 is concatenated while the rest tests each dimension separately to combine the result as weighted (NMF2), mean (NMF3), minimum (NMF4) and maximum (NMF5)

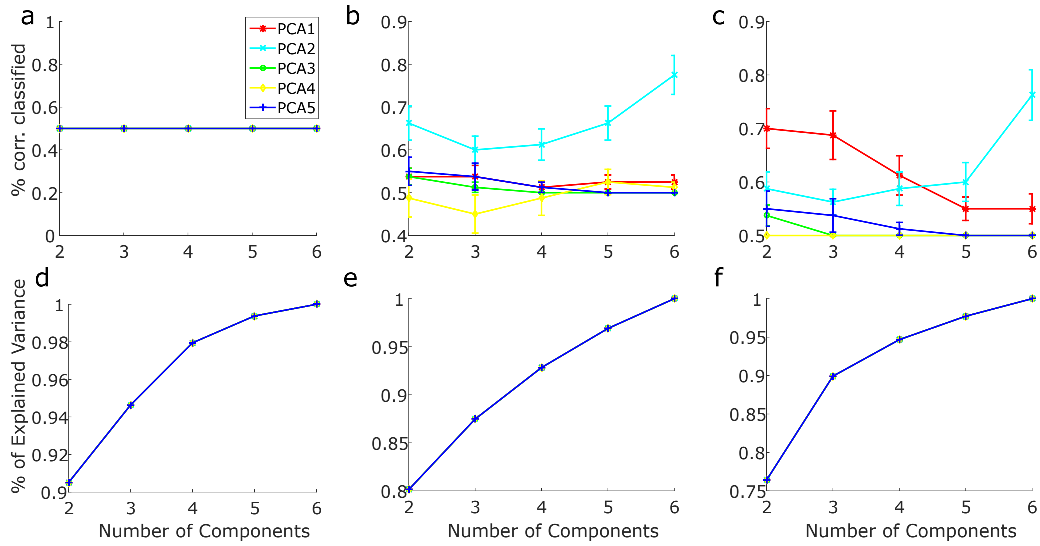


Figure 4.6.: Optimal number of components for model PCA on data set for the knee flexion while sitting exercise PCA1-5 show the different classifier (a) Knee Angle only. (b) All four muscle recordings. (c) Knee Angle and the four muscles. (d-f) corresponding explained variance of the data sets in (a-c).

	Components	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	6	PCA1-5	0.5000	0.6667	1
EMG	6	PCA2	0.7750	0.8085	1
Knee Degree + EMG	6	PCA2	0.7625	0.8000	1

Table 4.6.: Comparison of PCA components and Classifiers on data set for the knee flexion while sitting exercise, where PCA1 is concatenated while the rest tests each dimension separately to combine the result as weighted (PCA2), mean (PCA3), minimum (PCA4) and maximum (PCA5)

4.2.2 Principal component analysis - PCA

The PCA model had difficulties to discriminate both classes in both data sets independent of the number of channels. The only exception is the set with four channels of the knee flexion while sitting exercise, Figure 4.6 b) where one out of five classifiers can produce good results with 77.5 percent correctly classified. For five channels in Figure 4.6 c), a performance of 76.25 percent was achieved.

4.2.3 Wavelet transformation

Using only one channel for the the knee flexion while sitting exercise data set, the performance is optimal at wavelet level six with 97.5 percent classification rate for all classifiers, see Figure 4.7 a). For the

	Components	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	3	PCA1-5	0.5375	0.6667	0.8894
EMG	3	PCA4	0.5250	0.5250	0.7906
Knee Degree + EMG	3	PCA4	0.5375	0.6667	0.8176

Table 4.7.: Comparison of PCA components and Classifiers on data set for the standing up exercise, where PCA1 is concatenated while the rest tests each dimension separately to combine the result as weighted (PCA2), mean (PCA3), minimum (PCA4) and maximum (PCA5)

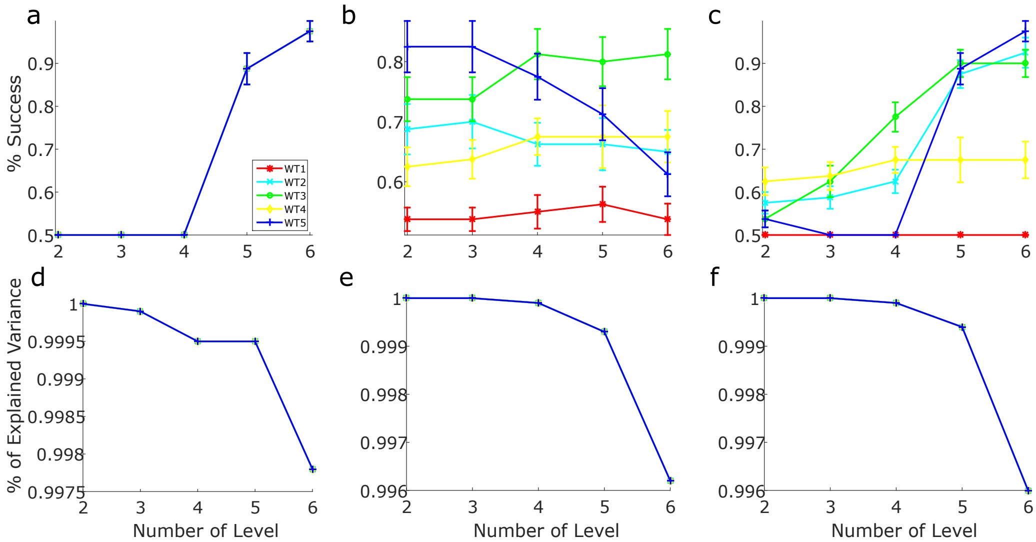


Figure 4.7.: Optimal number of components for model WT on data set for the knee flexion while sitting exercise WT1-5 show the different classifier (a) Knee Angle only. (b) All four muscle recordings. (c) Knee Angle and the four muscles. (d-f) corresponding explained variance of the data sets in (a-c).

	Level 2	Level 3	Level 4	Level 5	Level 6
Knee Degree	57	32	19	13	10
EMG	215	111	59	33	20
Knee Degree + EMG	267	137	72	39	23

Table 4.8.: Number of components for each tested Wavelet Level for the knee flexion while sitting exercise

	Level	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	6	WT1-5	0.9750	0.9744	0.9978
EMG	2	WT5	0.8250	0.8250	1.0000
Knee Degree + EMG	6	WT5	0.9750	0.9744	0.9960

Table 4.9.: Comparison of WT components and Classifiers on data set for the knee flexion while sitting exercise, where WT1 is concatenated while the rest tests each dimension separately to combine the result as weighted (WT2), mean (WT3), minimum (WT4) and maximum (WT5)

	Level	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	2	WT1-5	0.9625	0.9639	0.9997
EMG	4	WT4	0.8875	0.8861	0.9997
Knee Degree + EMG	4	WT4	0.9000	0.9000	0.9987

Table 4.10.: Comparison of WT components and Classifiers on data set for the standing up exercise, where WT1 is concatenated while the rest tests each dimension separately to combine the result as weighted (WT2), mean (WT3), minimum (WT4) and maximum (WT5)

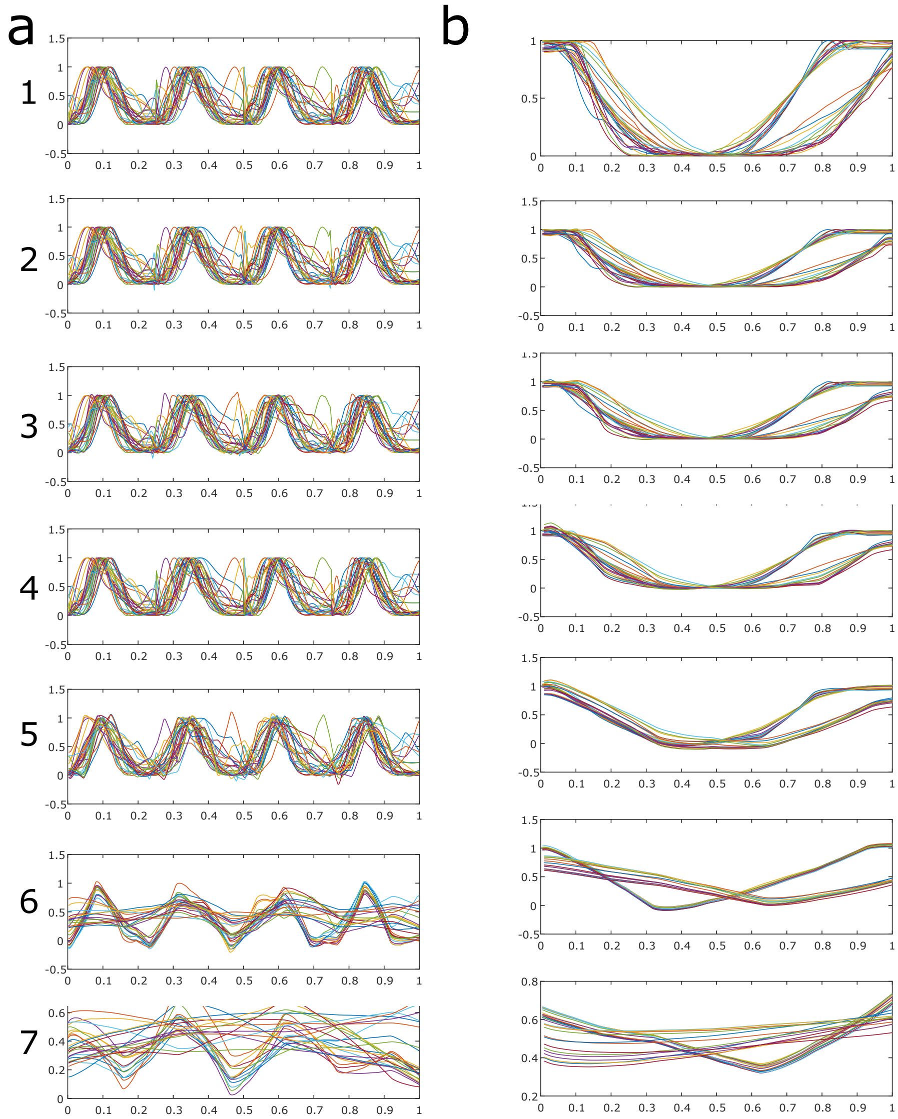


Figure 4.8.: Visualization of wavelet levels two until seven (2-7) compared to original trajectories on the top (1) (a) All four muscle recordings (b) Knee Angle only

standing up exercise data set the performance is mirrored, having its worst result with level six while all other levels have at least 95 percent, Figure B.4 a).

With four dimensions the knee flexion while sitting exercise data set in Figure 4.7 b) has the fifth classifier (WT5 - maximum) with the best result of 82.5 percent but with less components a declining classification rate, while the third classifier (WT3 - mean) has a worse rate with a high number of components but gets better with few components. The standing up exercise has the fourth classifier (WT4 - minimum) on top of all others with almost 90 percent success of classification, see Figure B.4 b).

Combining both previous tests to a five dimensional data set, the knee flexion while sitting exercise performs better with decreasing number of components coming to a final result of 97.5 percent at level six, Figure 4.7 c). The standing up exercise data set performs very similar to the four dimensions having the best result on the mid level four with exactly 90 percent, Figure B.4 c).

By using higher wavelet levels the performance could not be improved significantly and with the number of components falling under 20, see Table 4.8, a proper trajectory reconstruction seems to get lost and under ten the classification rate began to decrease rapidly, which can be seen in Figure 4.8.

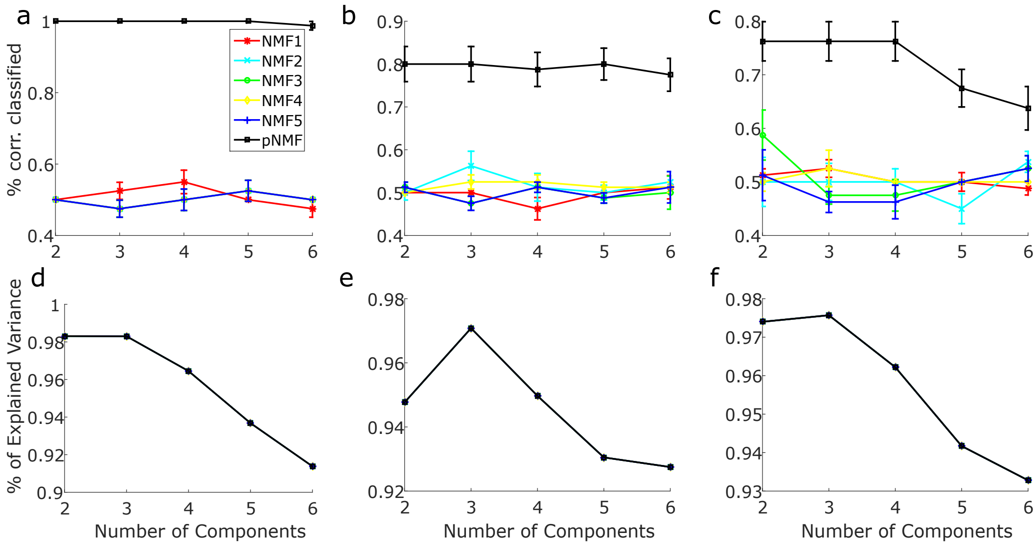


Figure 4.9.: Optimal number of components for model pNMF on data set for the knee flexion while sitting exercise NMF1-5 and pNMF show the different classifier (a) Knee Angle only. (b) All four muscle recordings. (c) Knee Angle and the four muscles. (d-f) corresponding explained variance of the data sets in (a-c).

	Dimension	Components	Best Classifier	Classification Rate	F-Score	Explained Variance
deterministic	Knee Degree	4	NMF1	0.5500	0.3077	0.9645
probabilistic	Knee Degree	2	Naive Bayes	**1.0000**	**1.0000**	0.9831
deterministic	EMG	3	NMF2	0.5625	0.5783	0.9708
probabilistic	EMG	2	Naive Bayes	**0.8000**	**0.8222**	0.9477
deterministic	Knee Degree + EMG	2	NMF3	0.5875	0.5217	0.9740
probabilistic	Knee Degree + EMG	2	Naive Bayes	**0.7625**	**0.7912**	0.9740

Table 4.11.: Comparison of deterministic vs. probabilistic classifiers Classifiers and different components for the pNMF model on the data set for the knee flexion while sitting exercise, where NMF1 is concatenated while the rest tests each dimension separately to combine the result as weighted (NMF2), mean (NMF3), minimum (NMF4) and maximum (NMF5) and pNMF is Naive Bayes

4.2.4 Probabilistic non-negative matrix factorisation - p-NMF

Using the pNMF model in parameter space, the Naive Bayes classifier takes advantage of the learned variance and outperforms all deterministic classifiers by far.

Optimal number of components

For most cases the Naive Bayes classifier could be used with the lowest dimensional model for best performance while the deterministic classifiers needed more components. Compared to the raw trajectory space, the explained variance remains high for all channels on both data sets.

Optimal number of parameters

The different parameter sets do not matter in parameter space. The classification rate stays on the same high level for all channels on both data sets. The only noticeable effect can be seen with parameter set

	Dimension	Components	Best Classifier	Classification Rate	F-Score	Explained Variance
deterministic	Knee Degree	3	NMF2-5	0.5000	0.3333	0.9798
probabilistic	Knee Degree	2	Naive Bayes	**1.0000**	**1.0000**	0.9995
deterministic	EMG	3	NMF5	0.5500	0.5000	0.9863
probabilistic	EMG	3	Naive Bayes	**0.9125**	**0.9157**	0.9863
deterministic	Knee Degree + EMG	3	NMF2	0.5750	0.5750	0.9896
probabilistic	Knee Degree + EMG	2	Naive Bayes	**1.0000**	**1.0000**	0.9943

Table 4.12.: Comparison of deterministic vs. probabilistic classifiers Classifiers and different components for the pNMF model on the data set for the standing up exercise, where NMF1 is concatenated while the rest tests each dimension separately to combine the result as weighted (NMF2), mean (NMF3), minimum (NMF4) and maximum (NMF5) and pNMF is Naive Bayes

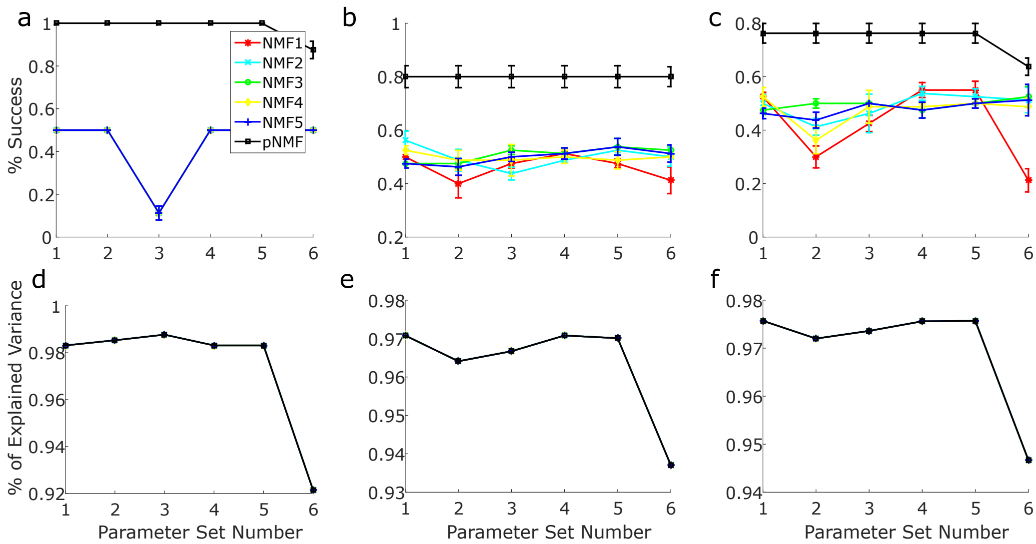


Figure 4.10.: Optimal parameter settings with best determined number of components for model pNMF on data set for the knee flexion while sitting exercise NMF1-5 and pNMF show the different classifier (a) Knee Angle only with two components. (b) All four muscle recordings with three components. (c) Knee Angle and the four muscles with two components. (d-f) corresponding explained variance of the parameter sets in (a-c).

	Components	Best Parameter Set	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	2	Set 1	Naive Bayes	1.0000	1.0000	0.9831
EMG	3	Set 1	Naive Bayes	0.8000	0.8222	0.9708
Knee Degree + EMG	3	Set 1	Naive Bayes	0.7625	0.7912	0.9757

Table 4.13.: Comparison of pNMF parameter Sets on data set for the knee flexion while sitting exercise, where NMF1 is concatenated while the rest tests each dimension separately to combine the result as weighted (NMF2), mean (NMF3), minimum (NMF4) and maximum (NMF5) and pNMF is Naive Bayes

	Components	Best Parameter Set	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	2	Set 1	Naive Bayes	1.0000	1.0000	0.9798
EMG	3	Set 1	Naive Bayes	0.9125	0.9157	0.9863
Knee Degree + EMG	3	Set 1	Naive Bayes	1.0000	1.0000	0.9896

Table 4.14.: Comparison of pNMF parameter Sets on data set for the standing up exercise, where NMF1 is concatenated while the rest tests each dimension separately to combine the result as weighted (NMF2), mean (NMF3), minimum (NMF4) and maximum (NMF5) and pNMF is Naive Bayes

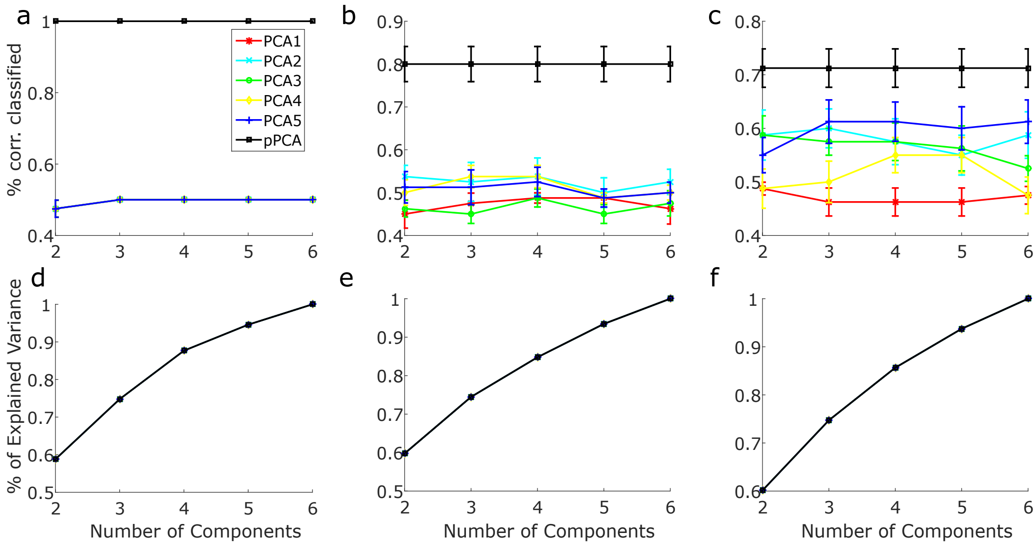


Figure 4.11.: Optimal number of components for model pPCA on data set for the knee flexion while sitting exercise PCA1-5 and pPCA show the different classifier (a) Knee Angle only. (b) All four muscle recordings. (c) Knee Angle and the four muscles. (d-f) corresponding explained variance of the data sets in (a-c).

	Dimension	Components	Best Classifier	Classification Rate	F-Score	Explained Variance
deterministic	Knee Degree	3	PCA1-5	0.5000	0.6667	0.7478
probabilistic	Knee Degree	2	Naive Bayes	**1.0000**	**1.0000**	0.5879
deterministic	EMG	3	PCA4	0.5375	0.5934	0.7447
probabilistic	EMG	2	Naive Bayes	**0.8000**	**0.8222**	0.5978
deterministic	Knee Degree + EMG	3	PCA5	0.6125	0.6804	0.7472
probabilistic	Knee Degree + EMG	2	Naive Bayes	**0.7125**	**0.7356**	0.6014

Table 4.15.: Comparison of deterministic vs. probabilistic classifiers Classifiers and different components for the pPCA model on the data set for the knee flexion while sitting exercise, where PCA1 is concatenated while the rest tests each dimension separately to combine the result as weighted (PCA2), mean (PCA3), minimum (PCA4) and maximum (PCA5) and pPCA is Naive Bayes

six, see Table 4.3, where for almost all cases a significant drop off in performance and explained variance occurs.

4.2.5 Probabilistic principal component analysis - p-PCA

In parameter space the pPCA model performed good with Naive Bayes classifier while the deterministic classifiers produced inferior performance. The only exception is the concatenated (PCA1) classifier for the standing up exercise data set which reaches over 70 percent classification rate over four channels, Figure B.7 b) and over 80 percent in the combined five channels, Figure B.7 c). As was the case with the pNMF model, the deterministic classifiers all need more components for the optimum performance while the probabilistic classifier can perform equally well on all complexity levels, with more or less explained variance. Especially the standing up exercise data set showed a huge difference in components where the deterministic classifiers needed between two and four more components, Table 4.16, while on the knee flexion while sitting exercise data set only one more component was needed, see Table 4.15.

	Dimension	Components	Best Classifier	Classification Rate	F-Score	Explained Variance
deterministic	Knee Degree	4	PCA1-5	0.5750	0.7018	0.8427
probabilistic	Knee Degree	2	Naive Bayes	**1.0000**	**1.0000**	0.5472
deterministic	EMG	4	PCA1	0.7250	0.7843	0.8282
probabilistic	EMG	2	Naive Bayes	**0.8875**	**0.8889**	0.5631
deterministic	Knee Degree + EMG	6	PCA1	0.8375	0.8602	1
probabilistic	Knee Degree + EMG	2	Naive Bayes	**1.0000**	**1.0000**	0.5515

Table 4.16.: Comparison of deterministic vs. probabilistic classifiers Classifiers and different components for the pPCA model on the data set for the standing up exercise, where PCA1 is concatenated while the rest tests each dimension separately to combine the result as weighted (PCA2), mean (PCA3), minimum (PCA4) and maximum (PCA5) and pPCA is Naive Bayes

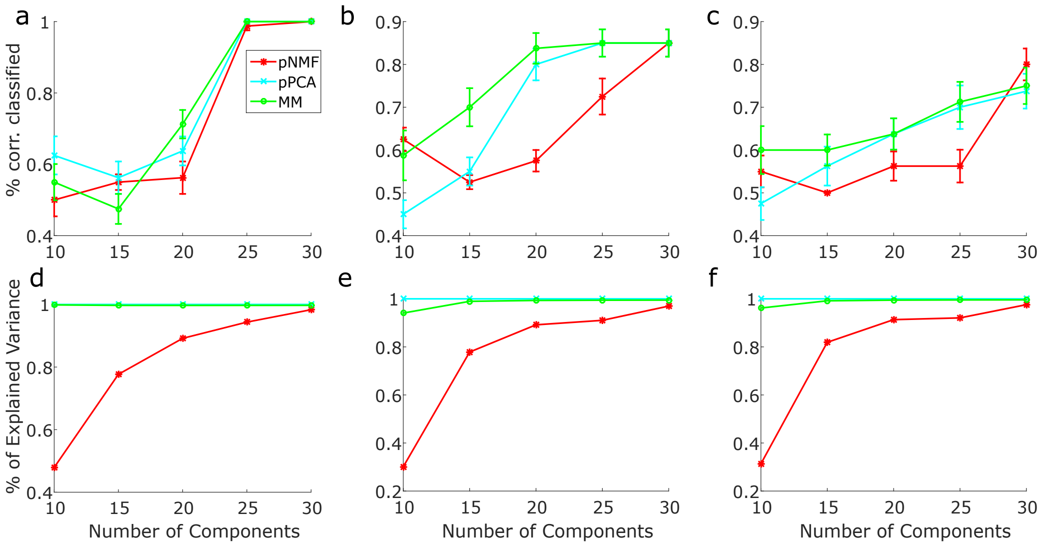


Figure 4.12.: Optimal number of Gaussians for model GMM on data set for the knee flexion while sitting exercise pNMF, pPCA and GMM show the different classifier (a) Knee Angle only. (b) All four muscle recordings. (c) Knee Angle and the four muscles. (d-f) corresponding explained variance of the data sets in (a-c).

4.2.6 Gaussian mixture model - GMM

For the mixture model a different number of Gaussians were used. We compared it also to the other probabilistic models p-NMF and p-PCA with the optimal settings and components from the previous subsections.

For one channel, the knee flexion while sitting exercise data set, Figure 4.12 a), all classifiers make a huge leap from 20 to 25 Gaussians, while in the standing up exercise data set the same leap happens between 10 and 20 Gaussians, Figure B.8 a).

With four EMG channels the data set of the knee flexion while sitting exercise and the standing up exercise are getting the first good results starting at 20 Gaussians, with 80 percent classification rate, see Figure 4.12-B.8 b).

In the combined five channel data set for the knee flexion while sitting exercise, the curve is not that steep and the combined result is worse than both previous ones, Figure 4.12 c). But for the standing up exercise data set the combination produces a perfect classification starting at 20 Gaussians.

For both data sets the mixture model produces better results with less Gaussians than pNMF and pPCA, leading to the optimum of 30 Gaussians for all channels in both data sets. Raising the number of Gaussians to 35 or above yields no further improvements as the classification rate remains constant.

	Gaussians	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	30	all	1.0000	1.0000	1.0000
EMG	30	all	0.8500	0.8605	1.0000
Knee Degree + EMG	30	pNMF	0.8000	0.8182	0.9757

Table 4.17.: Comparison of GMM number of Gaussians and probabilistic Classifiers on data set for the knee flexion while sitting exercise, where pNMF, pPCA and GMM are all Naive Bayes

	Gaussians	Best Classifier	Classification Rate	F-Score	Explained Variance
Knee Degree	20	GMM	1.0000	1.0000	0.9977
EMG	30	pNMF	0.9125	0.9157	0.9844
Knee Degree + EMG	30	pNMF	1.0000	1.0000	0.9878

Table 4.18.: Comparison of GMM number of Gaussians and probabilistic Classifiers on data set for the standing up exercise, where pNMF, pPCA and GMM are all Naive Bayes

Comparing Figure B.9 a) the first model with 10 Gaussians, from a single four channel data set, with the classification rate in Figure 4.12 b) at 10 Gaussians, it becomes clear why the prediction has to fail, since there is not much separation between both class models, which grows significant larger with a higher number of Gaussians until it delivers a good classification performance. Figure B.9 b) shows the same for a one channel case where 10 and 15 Gaussians are not enough to separate the classes far enough for a high correct classified percentage, as the good results start with 20 Gaussians for the MM, see 4.12 a).

4.3 Performance of probabilistic models vs. deterministic models

In the standing up exercise the three probabilistic approaches outperform standard PCA, NMF and WT, see Figure 4.14, in one, four and the combined five channels. While in the knee flexion while sitting exercise the standart NMF would still be outperformed, see Figure 4.5 b) parameter set one, it shows that in this limited data set and with parameter tuning of the non standard NMF very good results can be produced, Figure 4.13. A drawback is the tuning of components and parameters. The improvement of increasing the number of components in deterministic models is not as high as increasing the number of Gaussians in probabilistic methods. In contrast, the tuning of meta parameters can result in high improvements for the NMF model, while for the GMM the improvements are smaller. This is when the optimal number of Gaussians already reached a high level.

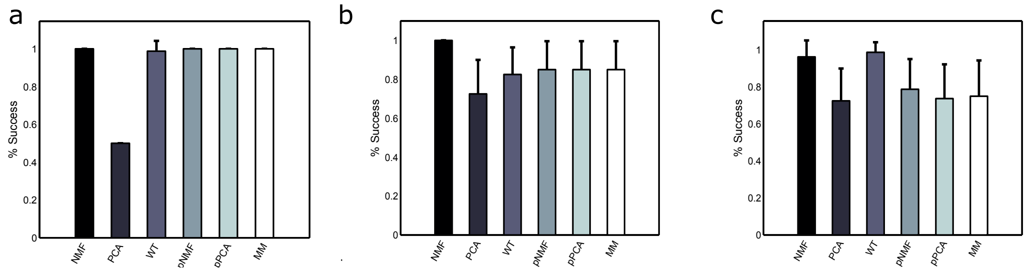


Figure 4.13.: Classification performance of deterministic and probabilistic approaches: Comparison of the deterministic methods PCA, NMF and WT to the proposed probabilistic variants p-PCA, p-NMF and MM. Correctly classified subjects for the knee flexion while sitting exercise for: (a) Knee Angle only (b) All four muscle recordings (c) Knee Angle and the four muscles combined

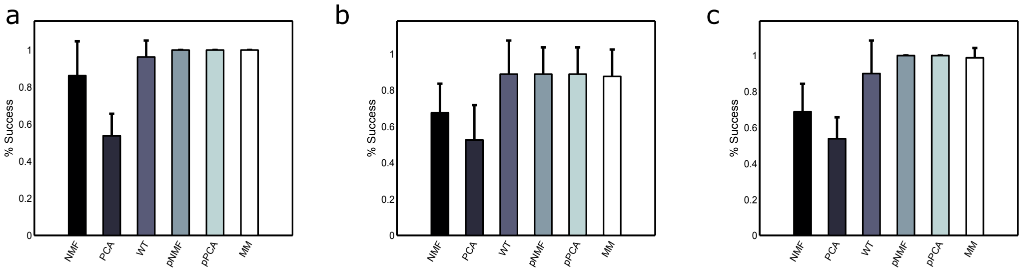


Figure 4.14.: Classification performance of deterministic and probabilistic approaches: Comparison of the deterministic methods PCA, NMF and WT to the proposed probabilistic variants p-PCA, p-NMF and MM. Classification success rate in percent on data recorded in the standing up exercise: (a) Knee Angle only (b) All four muscle recordings (c) Knee Angle and the four muscles combined

4.4 Feature models are less sensitive to measurement noise

We investigated the effect of EMG signal noise on the classification performance. Noise was simulated through evaluating all six methods with EMG data filtered with an increasing cutoff frequency in the low-pass filter, see Figure 4.15(a) for sample recordings of the vastus internus.

The methods were optimized for a cutoff frequency of 2 Hz, see results from previous sections. To ensure the performance of the probabilistic models the number of Gaussians had to be increased with higher cutoff frequency.

For the cutoff frequency of 8 Hz at the standing up exercise data set the optimal number of Gaussians is 40, see Figure 4.16 a). While for 16 Hz 60 was sufficient since the results didn't improve after that, Figure 4.16 b) . For the highest tested cutoff frequency of 24 Hz 100 Gaussians had to be used, Figure 4.16 c). In the knee flexion while sitting exercise data set the number of Gaussians for the cutoff frequencies have been 60, 45 and 40 respectively. All tests with 2 Hz have been kept at the before mentioned 30 Gaussians.

With increasing number of Gaussians the variance of the MM decreases, which can be seen in Figure 4.17 from a) to c).

In the standing up exercise, Figure 4.15 c) the MM is not only superior to the deterministic models in the strongest filtered case, see Figure 4.14, but also outperforms the other three feature based methods (WT, p-NMF, p-PCA) with increasing cutoff frequency.

The knee flexion while sitting exercise data set shows in Figure 4.15 b), what can already be seen in the other data set, that the performance of the NMF model remains constant through all tested frequencies, even without optimizing it for the higher frequencies. Remember that the standard NMF algorithm delivers classification results hardly better than chance, Figure 4.3 b), and with a constant performance for all cutoff frequencies it would be inferior to all other models and finding the optimal settings needs by far more computational time.

The vastly different outcomes on this limited data sets show that for significant results investigations on larger data sets are necessary.

4.5 Principle components of the models

It is interesting to plot the first two principal components in a $2D$ plot. With that we can identify the potential of a seperating hyperplane for classification. All visualizations shown have been conducted with the knee flexion while sitting exercise data set. Analysis on the standing up exercise revealed no further insight and are not shown.

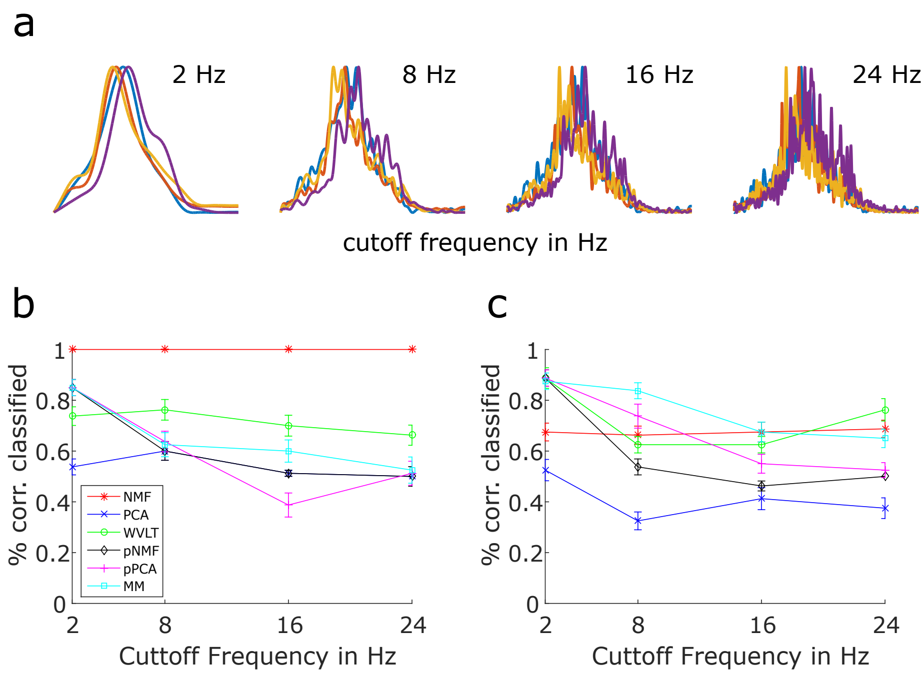


Figure 4.15.: Investigation of the robustness to EMG signal noise: (a) From left to right, the raw EMG signals were low-pass filtered with an increasing cutoff frequency to simulate additive measurement noise. (b) Classification performance for an increasing cutoff frequency on the knee flexion while sitting exercise. (c) Results for the standing up exercise.

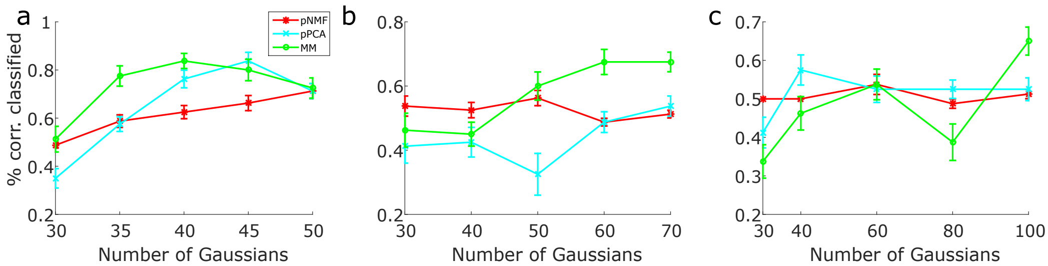


Figure 4.16.: Increasing number of Gaussians and their classification performance for the probabilistic models at the standing up exercise data set with a cutoff frequency of: (a) 8 Hz (b) 16 Hz (c) 24 Hz

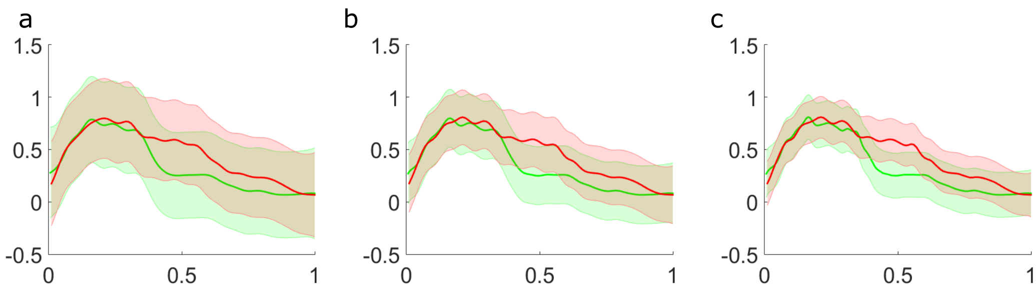


Figure 4.17.: MM prior plots of Healthy and Ill model trajectories from the standing up exercise data set and a cuttoff frequency of 8 Hz from the semitendinosus muscle with (a) 30 Gaussians (b) 40 Gaussians (c) 80 Gaussians

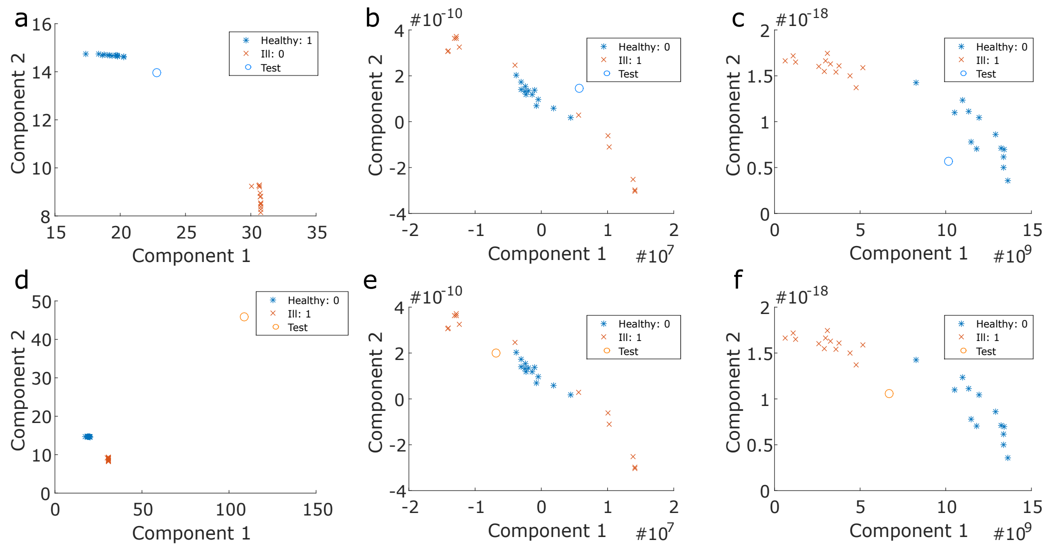


Figure 4.18.: Visualization of Components for the knee flexion while sitting exercise data set with the NMF model The first two components are show for: (a) Knee Angle only (b) All four muscle recordings (c) A single recoding from the rectus femoris muscle where the testdata is from healthy patients and d)-f) the testdata is from ill patients.

4.5.1 Principal components of the deterministic models

To extract the principal components for the deterministic models, the matrix of the raw trajectories $\boldsymbol{O} \in \mathbb{R}^{D \cdot T \times M}$ have been used, where M is the number of samples, D the number of channels and T the length of the trajectories.

Non negative matrix factorization - NMF

While in the data set of one channel the prediction works perfectly, see legend in Figure 4.18 a) and d), it can not discriminate between the classes in the muscle recordings with the NMF model, Figure 4.18 b) and e). However, in theory a separation should be possible with a linear classifier, Figure 4.18 c) and f).

Principal component analysis - PCA

In opposite to the NMF example the predictions fail in the case of one channel, Figure 4.19 a) and d) but succeeds for four channels 4.19 b) and e), which is also shown with a single recording4.19 c) and f).

4.5.2 Principal components of the probabilistic Models

For the extraction of the principal components of the probabilistic methods the matrix of the learned parameters $\boldsymbol{W} \in \mathbb{R}^{D \cdot K \times M}$ was used. M denotes the number of trials and $N = DK$ denotes the number of features.

Probabilistic non-negative matrix factorization - p-NMF

In parameter space the NMF model fails to predict all shown examples. While with one channel Figure 4.20 a) and d) the data is clearly separated and the test data fits right in those two classes and could be predicted correctly with a different classifier like linear discriminant analysis or support vector machines,

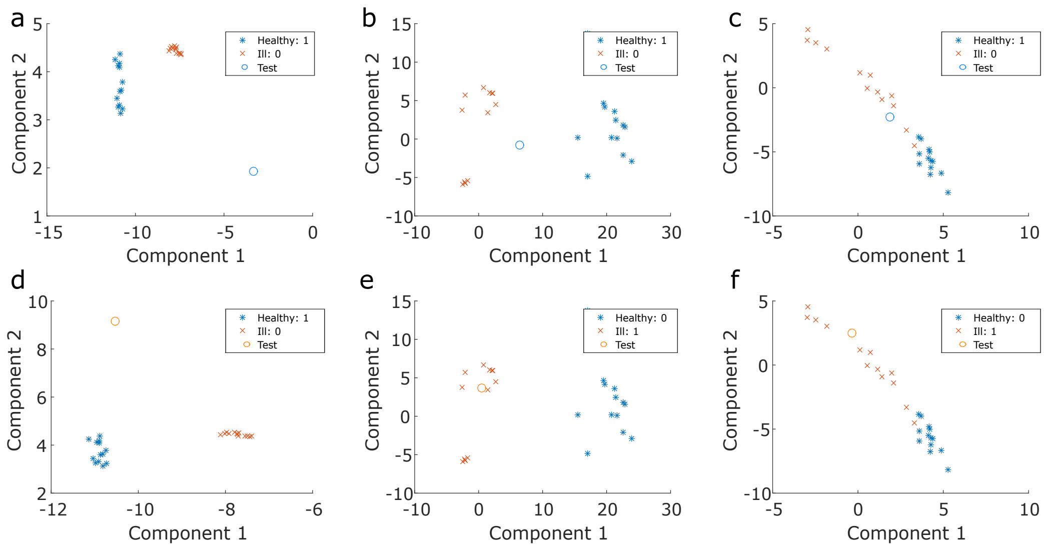


Figure 4.19.: Visualization of Components for the knee flexion while sitting exercise data set with the PCA model The first two components are show for: (a) Knee Angle only (b) All four muscle recordings (c) A single recoding from the biceps femoris muscle where the testdata is from healthy patients and d)-f) the testdata is from ill patients.

it becomes clear that in both four channel cases, Figure 4.20 b)-c) and e)-f) the two classes are not separated and can hardly be predicted properly.

Probabilistic principal component analysis - p-PCA

The p-PCA model in parameter space shows a better example than the p-NMF model. Here both, the one and four channel cases get predicted correctly, 4.21 a) and b) being healthy and are predicted healthy, while d) and e) being ill are predicted ill. The single recording in c) and f) shows that one of the four channels can not be distinguishable while the prediction still works overall when all four muscle recordings form a prediction together.

Visualisation of the Gaussian mixture model

In both data sets, Figure 4.22 a) and b) can hardly be assigned to a class in the first half of the model trajectories, but then the healthy trajectories subside much slower than the ill, which separates both classes well.

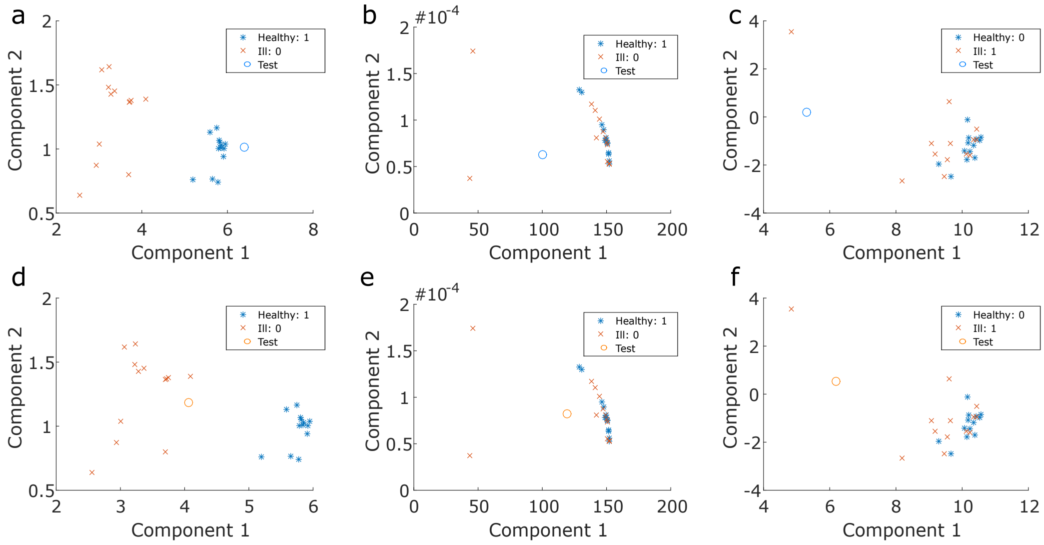


Figure 4.20.: Visualization of Components for the knee flexion while sitting exercise data set with the pNMF model The first two components are show for: (a) Knee Angle only (b) All four muscle recordings (c) A single recoding from the rectus femoris muscle where the testdata is from healthy patients and d)-f) the testdata is from ill patients.

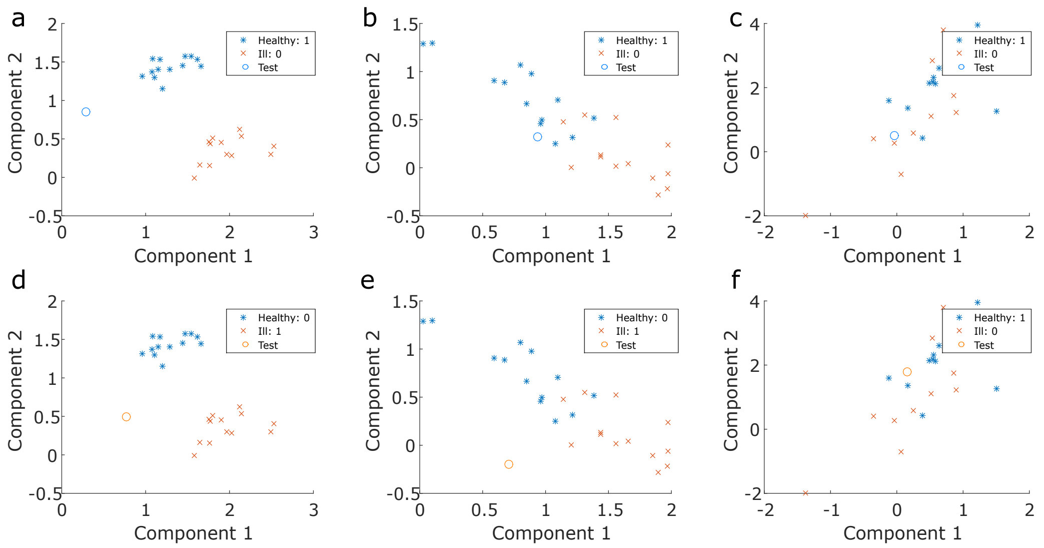


Figure 4.21.: Visualization of Components for the knee flexion while sitting exercise data set with the pPCA model The first two components are show for: (a) Knee Angle only (b) All four muscle recordings (c) A single recoding from the biceps vastus internus where the testdata is from healthy patients and d)-f) the testdata is from ill patients.

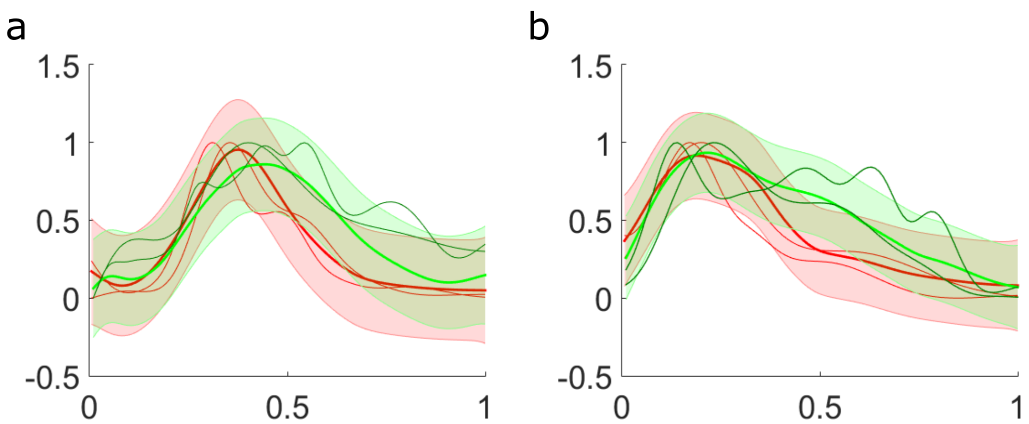


Figure 4.22.: Visualization of the MM with healthy and ill model trajectories and their test trajectories at once for: (a) the knee flexion while sitting exercise data set of the biceps femoris muscle (b) the standing up exercise data set of the semitendinosus muscle

5 Conclusion & Future Work

5.1 Conclusion

Electromyography (EMG) signals in, for example prosthetic and rehabilitation tasks [4, 58, 9, 59] are typically corrupted by sensor noise, the surface electrodes position might change, and even for the same executed movement different EMG patterns are observed (this is known as motor variability), see Section 2.2. While the first issue can be circumvented through averaging, the other two require EMG models that represent the variance of the data.

We presented a probabilistic model that maps EMG signals to a feature space using Gaussian basis functions. The Gaussian means and the variances are fixed while the amplitudes are scaled by learnable features. The probabilistic model implements Bayesian linear regression in fixed basis functions [51] and was previously used as part of a movement representation in robotics [47]. It can be trained through least squares regression or variational inference and scales to more complex hierarchical Bayesian models [49] relevant for EMG applications.

Most works concerning EMG signals use wavelet transformation as a first step of dealing with the measured data, followed or sometimes preceded by dimensionality reduction techniques. Instead of wavelet transformation we used learned Gaussian features. We extended the model by applying principal components analysis (PCA) [17] and the non-negative matrix factorization (NMF) [18]. In our results we show that those commenly used techniques give no further advantage to our proposed probabilistic model, see Subsection 2.3.

We evaluated the resulting approaches in a clinical lower limb data set [56] with the task of predicting knee abnormalities. Both used exercises, flexing the knee from a sitting position and standing up from a sitting position, can be seen as coordinated motor activity causing *predictable* biomechanical motion, Subsection 2.2.

We found that the proposed models outperform standard PCA, NMF and wavelet transformation [19] in terms of classification performance in the standing up exercise, see Figure 4.14, since they can captures the correlation over multiple channels (space) and time. Drawbacks of these approaches are the increased run-time of the learning of the model and additional the tuning beforehand.

On the other side stands the supreme classification performance of the NMF model in the knee flexion while sitting exercise. By using an improved implementation of NMF on EMG Signals where the sparsity of both factors could be controlled by several parameters, a connection could be drawn to the sparse coding found in sensory systems. This relationship is discussed in Subsubsection 2.3.5.

Our proposed Gaussian Mixture model is less sensitive to measurement noise. It automatically filters the data through the Gaussian representation, Figure 4.15 b) and c). NMF shows its ability to identify components which are stable across different conditions, Section 2.3.5.

5.2 Future work

The good results in spite of working with a limited data set only show how much potential our proposed methods can have when analyzing EMG signals. Therefore we can discuss several extensions.

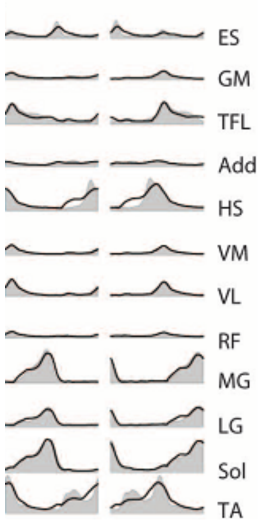


Figure 5.1.: Example for a recording of 12 muscles with varying shapes [60]

5.2.1 Increasing the dimension

As pointed out in Section 2.3.6, the recording of more than four muscles can help improving the results significantly. The number of extracted components suffers under few recorded muscles and the question of independently activated versus always coactivated muscles remains as well. The improvement for deterministic models would be by finding better fitting synergies and for probabilistic models a higher correlation over space and time.

5.2.2 Different shapes of EMG

Furthermore the shape of our tested EMG signals over all four muscles is very similar, see Figure 4.1. In the standing up exercise data set the shapes are less similar, which results in an improved classification performance from 85 percent, Figure 4.13 to 91.25 percent, Figure 4.14. Having a data set with different shapes of muscles like in Figure 5.1, should further improve classifications. This becomes even more critical for patients with various disease where one specific muscle has no activity at all, compared to high activity in healthy patients, and all other muscles are relatively the same. In such cases discrimination becomes easier in that single dimension and can influence the overall prediction dramatically.

5.2.3 Additional features

During the creation of the data set, its possible to integrate different kinds of information, similar to data stored on health insurance cards, e.g., the eGK (Elektronische Gesundheitskarte) [61]. Knowing little things like the age of the patient might aid the classification. Is a person just old and has less strength and weaker signals, or is the person young and ill? Had the patient any pain before? Like a knee injury or some muscle damage, or other seemingly unrelated injuries? There may be different kinds of patients who may be clustered differently even inside an existing class.
Additionally, EMG model parameters could be an extension of health insurance cards.

5.2.4 Testing more classifiers and methods

As seen in Section 4.5 on the visualization of the models, some results could be improved by the use of different classifiers like linear discriminant analysis or support vector machines [51]. It would also

Figure 5.2.: eGK (Elektronische Gesundheitskarte) [62]

be possible to further extending the proposed model in the same way as p-NMF and p-PCA, applying wavelet transformation in Gaussian feature space to either use the wavelet features or working with the wavelet reconstruction of the parameters.

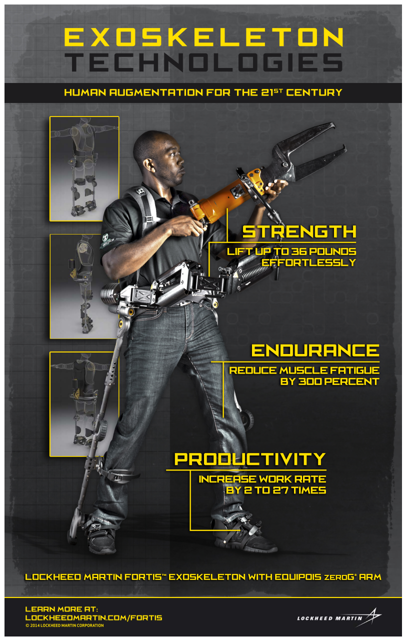
5.2.5 Latent variables

Based on the model in [49], the proposed Gaussian mixture model could be exploited even further by utilizing the latent variables. This would enable to not only classify a patient with a disease but also to determine the severity or progress of an illness. It is also easy to imagine to find variations of diseases inside a certain class of a medical condition.

5.2.6 Discrimination of different movements

Introducing a new way of EMG analysis could also help the classification of various movements. This would be applicable for support of an exoskeleton which could relieve a worker from heavy lifting or enhancing strength and endurance during walking [63] [64]. Or more general the amplification of motions to compensate for a disease like amyotrophic lateral sclerosis (ALS).

(a) Example from [65]



(b) Example from [66]

Figure 5.3.: Exoskeleton examples

Bibliography

[1] P. Konrad, "The abc of emg," *A practical introduction to kinesiological electromyography*, vol. 1, 2005.

[2] J. E. Duggan and G. B. Drummond, "Abdominal muscle activity and intraabdominal pressure after upper abdominal surgery.," *Anesthesia & Analgesia*, vol. 69, no. 5, pp. 598–603, 1989.

[3] N. F. Güler and S. Koçer, "Use of support vector machines and neural network in diagnosis of neuromuscular disorders," *Journal of medical systems*, vol. 29, no. 3, pp. 271–284, 2005.

[4] C. Castellini and P. van der Smagt, "Surface emg in advanced hand prosthetics," *Biological cybernetics*, vol. 100, no. 1, pp. 35–47, 2009.

[5] D. M. Rittenhouse, H. A. Abdullah, R. J. Runciman, and O. Basir, "A neural network model for reconstructing emg signals from eight shoulder muscles: Consequences for rehabilitation robotics and biofeedback," *Journal of biomechanics*, vol. 39, no. 10, pp. 1924–1932, 2006.

[6] D.-p. Yang, J.-d. Zhao, Y.-k. Gu, X.-q. Wang, N. Li, L. Jiang, H. Liu, H. Huang, and D.-w. Zhao, "An anthropomorphic robot hand developed based on underactuated mechanism and controlled by emg signals," *Journal of Bionic Engineering*, vol. 6, no. 3, pp. 255–263, 2009.

[7] S. Bitzer and P. van der Smagt, "Learning emg control of a robotic hand: towards active prostheses," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pp. 2819–2823, IEEE, 2006.

[8] J. L. McKay and L. H. Ting, "Optimization of muscle activity for task-level goals predicts complex changes in limb forces across biomechanical contexts," *PLoS Comput. Biol*, vol. 8, no. 4, p. e1002465, 2012.

[9] Z. O. Khokhar, Z. G. Xiao, C. Menon, *et al.*, "Surface emg pattern recognition for real-time control of a wrist exoskeleton," *Biomedical engineering online*, vol. 9, no. 1, p. 41, 2010.

[10] M. Bland *et al.*, *An introduction to medical statistics. Ed. 3*. Oxford University Press, 2000.

[11] D. Farrell, "Bebionic 3 prosthetic hand." `http://vr-zone.com/articles/bebionic-3-prosthetic-hand-looks-like-something-from-the-terminator/17769.html`, 2012. [Online; accessed 2015-12-10].

[12] A. Hiraiwa, K. Shimohara, and Y. Tokunaga, "Emg pattern analysis and classification by neural network," in *Systems, Man and Cybernetics, 1989. Conference Proceedings., IEEE International Conference on*, pp. 1113–1115, IEEE, 1989.

[13] C. Christodoulou, C. S. Pattichis, *et al.*, "Unsupervised pattern recognition for the classification of emg signals," *Biomedical Engineering, IEEE Transactions on*, vol. 46, no. 2, pp. 169–178, 1999.

[14] G. Wang, Z. Wang, W. Chen, and J. Zhuang, "Classification of surface emg signals using optimal wavelet packet method based on davies-bouldin criterion," *Medical and Biological Engineering and Computing*, vol. 44, no. 10, pp. 865–872, 2006.

[15] S. S. Nair, R. M. French, D. Laroche, and E. Thomas, "The application of machine learning algorithms to the analysis of electromyographic patterns from arthritic patients," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 18, no. 2, pp. 174–184, 2010.

[16] A. Subasi, "Classification of emg signals using pso optimized svm for diagnosis of neuromuscular disorders," *Computers in biology and medicine*, vol. 43, no. 5, pp. 576–586, 2013.

[17] M. Reaz, M. 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.

[18] M. C. Tresch, V. C. Cheung, and A. d'Avella, "Matrix factorization algorithms for the identification of muscle synergies: evaluation on simulated and experimental data sets," *Journal of Neurophysiology*, vol. 95, no. 4, pp. 2199–2212, 2006.

[19] I. Daubechies *et al.*, *Ten lectures on wavelets*, vol. 61. SIAM, http://www.siam.org/journals/ojsa.php, 1992.

[20] A. D. Chan and K. B. Englehart, "Continuous myoelectric control for powered prostheses using hidden markov models," *Biomedical Engineering, IEEE Transactions on*, vol. 52, no. 1, pp. 121–124, 2005.

[21] B. Karlık, "Machine learning algorithms for characterization of emg signals," *ijiee.org*, 2014.

[22] M. Künemund, "Machine learning methods for emg-based stroke patient movement analysis," *Master Thesis*, 2014.

[23] L. H. Ting and S. A. Chvatal, "Decomposing muscle activity in motor tasks," *Motor Control Theories, Experiments and Applications. Oxf. Univ. Press, New York*, pp. 102v–138, 2010.

[24] N. A. Bernstein, "The co-ordination and regulation of movements," *Pergamon Press Ltd.*, 1967.

[25] K. Newell and L. G. Carlton, "Force variability in isometric responses.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 14, no. 1, p. 37, 1988.

[26] M. L. Latash, J. P. Scholz, and G. Schöner, "Motor control strategies revealed in the structure of motor variability," *Exercise and sport sciences reviews*, vol. 30, no. 1, pp. 26–31, 2002.

[27] Y.-G. Ko, J. H. Challis, J. P. Stitt, and K. M. Newell, "Organization of compensatory postural coordination patterns," *Journal of motor behavior*, vol. 35, no. 4, pp. 325–342, 2003.

[28] D. S. Reisman and J. P. Scholz, "Workspace location influences joint coordination during reaching in post-stroke hemiparesis," *Experimental brain research*, vol. 170, no. 2, pp. 265–276, 2006.

[29] J. M. Macpherson, "How flexible are muscle synergetics?," *Motor control: concepts and issues*, 1991.

[30] M. C. Tresch, P. Saltiel, and E. Bizzi, "The construction of movement by the spinal cord," *Nature neuroscience*, vol. 2, no. 2, pp. 162–167, 1999.

[31] V. C. Cheung, A. d'Avella, M. C. Tresch, and E. Bizzi, "Central and sensory contributions to the activation and organization of muscle synergies during natural motor behaviors," *The Journal of neuroscience*, vol. 25, no. 27, pp. 6419–6434, 2005.

[32] L. H. Ting and J. M. Macpherson, "A limited set of muscle synergies for force control during a postural task," *Journal of neurophysiology*, vol. 93, no. 1, pp. 609–613, 2005.

[33] G. Torres-Oviedo and L. H. Ting, "Muscle synergies characterizing human postural responses," *Journal of neurophysiology*, vol. 98, no. 4, pp. 2144–2156, 2007.

[34] F. Danion, G. Schöner, M. L. Latash, S. Li, J. P. Scholz, and V. M. Zatsiorsky, "A mode hypothesis for finger interaction during multi-finger force-production tasks," *Biological cybernetics*, vol. 88, no. 2, pp. 91–98, 2003.

[35] V. Krishnamoorthy, M. L. Latash, J. P. Scholz, and V. M. Zatsiorsky, "Muscle modes during shifts of the center of pressure by standing persons: effect of instability and additional support," *Experimental brain research*, vol. 157, no. 1, pp. 18–31, 2004.

[36] M. L. Latash, J. P. Scholz, and G. Schoner, "Toward a new theory of motor synergies," *MOTOR CONTROL-CHAMPAIGN-*, vol. 11, no. 3, p. 276, 2007.

[37] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[38] A. S. Aruin, W. R. Forrest, and M. L. Latash, "Anticipatory postural adjustments in conditions of postural instability," *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, vol. 109, no. 4, pp. 350–359, 1998.

[39] V. Krishnamoorthy, M. L. Latash, J. P. Scholz, and V. M. Zatsiorsky, "Muscle synergies during shifts of the center of pressure by standing persons," *Experimental brain research*, vol. 152, no. 3, pp. 281–292, 2003.

[40] G. Torres-Oviedo, J. M. Macpherson, and L. H. Ting, "Muscle synergy organization is robust across a variety of postural perturbations," *Journal of neurophysiology*, vol. 96, no. 3, pp. 1530–1546, 2006.

[41] G. Torres-Oviedo and L. H. Ting, "Subject-specific muscle synergies in human balance control are consistent across different biomechanical contexts," *Journal of neurophysiology*, vol. 103, no. 6, pp. 3084–3098, 2010.

[42] B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs," *Current opinion in neurobiology*, vol. 14, no. 4, pp. 481–487, 2004.

[43] I. R. Fiete, M. S. Fee, and H. S. Seung, "Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances," *Journal of neurophysiology*, vol. 98, no. 4, pp. 2038–2057, 2007.

[44] L. H. Ting and J. L. McKay, "Neuromechanics of muscle synergies for posture and movement," *Current opinion in neurobiology*, vol. 17, no. 6, pp. 622–628, 2007.

[45] D. J. Clark, L. H. Ting, F. E. Zajac, R. R. Neptune, and S. A. Kautz, "Merging of healthy motor modules predicts reduced locomotor performance and muscle coordination complexity post-stroke," *Journal of neurophysiology*, vol. 103, no. 2, pp. 844–857, 2010.

[46] E. Benetos, M. Kotti, and C. Kotropoulos, "Musical instrument classification using non-negative matrix factorization algorithms," in *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, IEEE, 2006.

[47] A. Paraschos, C. Daniel, J. Peters, and G. Neumann, "Probabilistic movement primitives," in *Advances in Neural Information Processing Systems (NIPS), Cambridge, MA: MIT Press.*, 2013.

[48] E. Rueckert and A. d'Avella, "Learned parametrized dynamic movement primitives with shared synergies for controlling robotic and musculoskeletal systems," *Frontiers in computational neuroscience*, vol. 7, 2013.

[49] E. Rueckert, J. Mundo, A. Paraschos, J. Peters, and G. Neumann, "Extracting low-dimensional control variables for movement primitives," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2015.

[50] R. Albright, J. Cox, D. Duling, A. N. Langville, and C. Meyer, "Algorithms, initializations, and convergence for the nonnegative matrix factorization," tech. rep., Tech. rep. 919. NCSU Technical Report Math 81706., 2006.

[51] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[52] Y. P. Ivanenko, R. E. Poppele, and F. Lacquaniti, "Five basic muscle activation patterns account for muscle activity during human locomotion," *The Journal of physiology*, vol. 556, no. 1, pp. 267–282, 2004.

[53] A. d'Avella, P. Saltiel, and E. Bizzi, "Combinations of muscle synergies in the construction of a natural motor behavior," *Nature neuroscience*, vol. 6, no. 3, pp. 300–308, 2003.

[54] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *Bioinfo*, 2011.

[55] J. Frère and F. Hug, "Between-subject variability of muscle synergies during a complex motor skill," *Frontiers in computational neuroscience*, vol. 6, 2012.

[56] M. Lichman, "UCI machine learning repository." `http://archive.ics.uci.edu/ml`, 2013. [Online; accessed 2015-12-10].

[57] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series.," in *KDD workshop*, vol. 10, pp. 359–370, Seattle, WA, 1994.

[58] H. Huang, T. Kuiken, R. D. Lipschutz, *et al.*, "A strategy for identifying locomotion modes using surface electromyography," *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 1, pp. 65–73, 2009.

[59] A. H. Al-Timemy, G. Bugmann, J. Escudero, and N. Outram, "Classification of finger movements for the dexterous hand prosthesis control with surface electromyography," *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, no. 3, pp. 608–618, 2013.

[60] N. Dominici, Y. P. Ivanenko, G. Cappellini, A. d'Avella, V. Mondì, M. Cicchese, A. Fabiano, T. Silei, A. Di Paolo, C. Giannini, *et al.*, "Locomotor primitives in newborn babies and their development," *Science*, vol. 334, no. 6058, pp. 997–999, 2011.

[61] J. Caumanns, H. Weber, A. Fellien, H. Kurrek, O. Boehm, J. Neuhaus, J. Kunsmann, and B. Struif, "Die egk-lösungsarchitektur architektur zur unterstützung der anwendungen der elektronischen gesundheitskarte," *Informatik-Spektrum*, vol. 29, no. 5, pp. 341–348, 2006.

[62] Wikipedia, "Elektronische gesundheitskarte." `https://upload.wikimedia.org/wikipedia/commons/2/29/Elektronische_Gesundheitskarte_2011.svg`, 2011. [Online; accessed 2015-12-10].

[63] J. E. Pratt, B. T. Krupp, C. J. Morse, and S. H. Collins, "The roboknee: an exoskeleton for enhancing strength and endurance during walking," in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, vol. 3, pp. 2430–2435, IEEE, 2004.

[64] H.-C. Hsieh and C.-C. Lan, "A lightweight gravity-balanced exoskeleton for home rehabilitation of upper limbs," in *Automation Science and Engineering (CASE), 2014 IEEE International Conference on*, pp. 972–977, IEEE, 2014.

[65] S. Yang, "Uc berkeley researchers developing robotic exoskeleton that can enhance human strength and endurance," *University of California, Berkeley News*, 2004.

[66] L. Martin, "Fortis exoskeleton." `http://www.lockheedmartin.com/us/products/exoskeleton/FORTIS.html`, 2014. [Online; accessed 2015-12-10].

[67] R. Davis, "Diagnostic reasoning based on structure and behavior," *Artificial intelligence*, vol. 24, no. 1, pp. 347–410, 1984.

[68] P. Zhou, X. Li, F. Jahanmiri-Nezhad, W. Z. Rymer, and P. E. Barkhaus, "Duration of observation required in detecting fasciculation potentials in amyotrophic lateral sclerosis using high-density surface emg," *J Neuroeng Rehabil*, vol. 9, no. 78, pp. 0003–9, 2012.

[69] G. R. Naik and H. T. Nguyen, "Nonnegative matrix factorization for the identification of emg finger movements: Evaluation using matrix analysis," *Biomedical and Health Informatics, IEEE Journal of*, vol. 19, no. 2, pp. 478–485, 2015.

[70] M. A. Giese, A. Mukovskiy, A.-N. Park, L. Omlor, and J.-J. E. Slotine, "Real-time synthesis of body movements based on learned primitives," in *Statistical and Geometrical Approaches to Visual Motion Analysis*, pp. 107–127, Springer, 2009.

[71] D. J. Berger, R. Gentner, T. Edmunds, D. K. Pai, and A. d'Avella, "Differences in adaptation rates after virtual surgeries provide direct evidence for modularity," *The Journal of Neuroscience*, vol. 33, no. 30, pp. 12384–12394, 2013.

[72] S. H. Scott and J. F. Kalaska, "Reaching movements with similar hand paths but different arm orientations. i. activity of individual cells in motor cortex," *Journal of neurophysiology*, vol. 77, no. 2, pp. 826–852, 1997.

[73] M. S. Graziano and T. N. Aflalo, "Rethinking cortical organization moving away from discrete areas arranged in hierarchies," *The Neuroscientist*, vol. 13, no. 2, pp. 138–147, 2007.

[74] B. Schepens and T. Drew, "Independent and convergent signals from the pontomedullary reticular formation contribute to the control of posture and movement during reaching in the cat," *Journal of neurophysiology*, vol. 92, no. 4, pp. 2217–2238, 2004.

[75] B. Schepens and T. Drew, "Descending signals from the pontomedullary reticular formation are bilateral, asymmetric, and gated during reaching movements in the cat," *Journal of neurophysiology*, vol. 96, no. 5, pp. 2229–2252, 2006.

[76] B. Schepens, P. Stapley, and T. Drew, "Neurons in the pontomedullary reticular formation signal posture and movement both as an integrated behavior and independently," *Journal of neurophysiology*, vol. 100, no. 4, pp. 2235–2253, 2008.

[77] P. J. Stapley and T. Drew, "The pontomedullary reticular formation contributes to the compensatory postural responses observed following removal of the support surface in the standing cat," *Journal of neurophysiology*, vol. 101, no. 3, pp. 1334–1350, 2009.

[78] J. Quevedo, B. Fedirchuk, S. Gosgnach, and D. McCrea, "Group i disynaptic excitation of cat hindlimb flexor and bifunctional motoneurones during fictive locomotion," *The Journal of Physiology*, vol. 525, no. 2, pp. 549–564, 2000.

[79] D. A. McCrea, "Spinal circuitry of sensorimotor control of locomotion," *The Journal of physiology*, vol. 533, no. 1, pp. 41–50, 2001.

[80] T. Drew, J. Kalaska, and N. Krouchev, "Muscle synergies during locomotion in the cat: a model for motor cortex control," *The Journal of physiology*, vol. 586, no. 5, pp. 1239–1245, 2008.

A List of publications

Rueckert,E.;Kohlschuetter,J.;Peters,J. (2015). Learning Probabilistic Features from EMG Data for Predicting Knee Abnormalities. Accepted for MEDICON 2016.

A.1 Comments and contributions to publications

The paper *Learning Probabilistic Features from EMG Data for Predicting Knee Abnormalities* was written by Elmar Rueckert and myself. This paper builds the foundation for Section 3.

B Supplementary figures

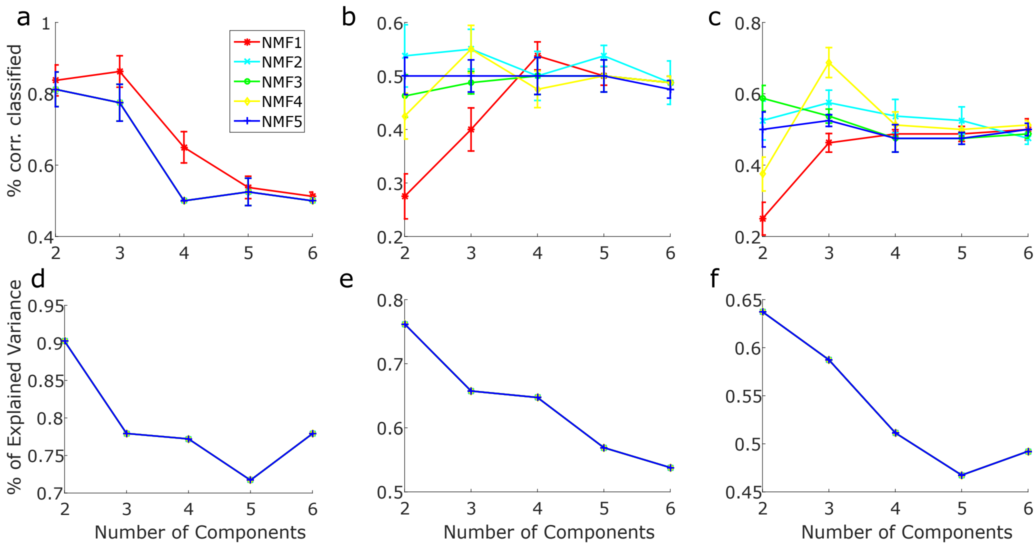


Figure B.1.: Optimal number of components for model NMF on data set for the standing up exercise NMF1-5 show the different classifier (a) Knee angle only. (b) All four muscle recordings. (c) Knee angle and the four muscles. (d-f) corresponding explained variance of the datasets in (a-c).

	Level 2	Level 3	Level 4	Level 5	Level 6
Knee Degree	46	26	16	11	9
EMG	171	89	48	27	17
Knee Degree + EMG	212	109	58	32	19

Table B.1.: Number of components for each tested Wavelet Level for the standing up exercise

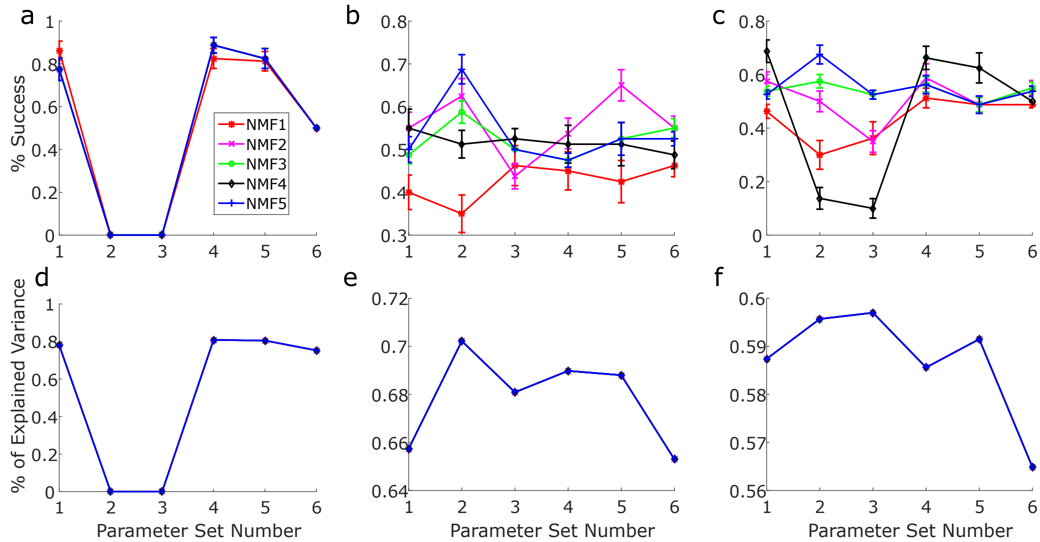


Figure B.2.: Optimal parameter settings with best determined number of components for model NMF on data set for the standing up exercise NMF1-5 show the different classifier (a) Knee Angle only with three components. (b) All four muscle recordings with three components. (c) Knee Angle and the four muscles with three components. (d-f) corresponding explained variance of the parameter sets in (a-c).

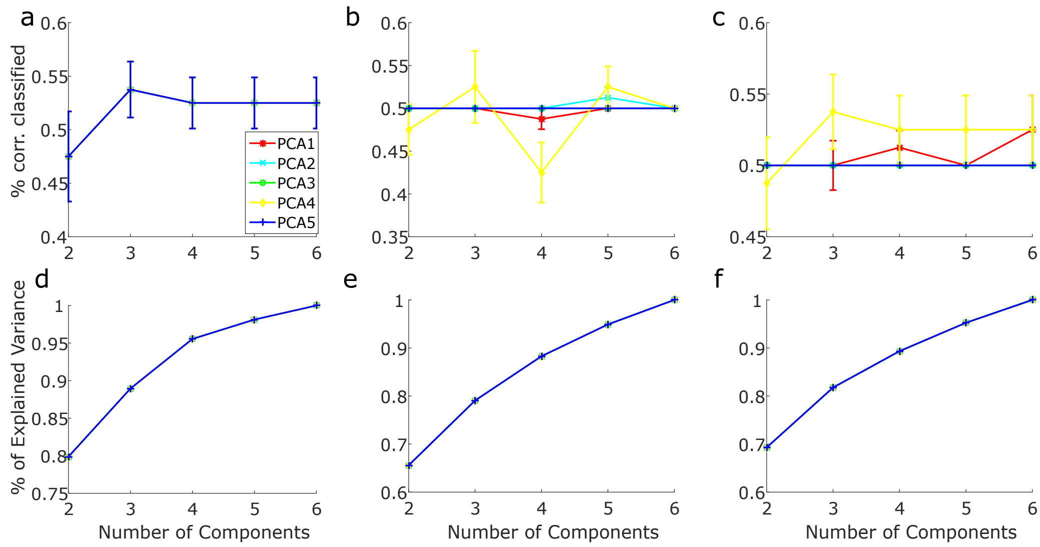


Figure B.3.: Optimal number of components for model PCA on data set for the standing up exercise PCA1-5 show the different classifier (a) Knee Angle only. (b) All four muscle recordings. (c) Knee Angle and the four muscles. (d-f) corresponding explained variance of the data sets in (a-c).

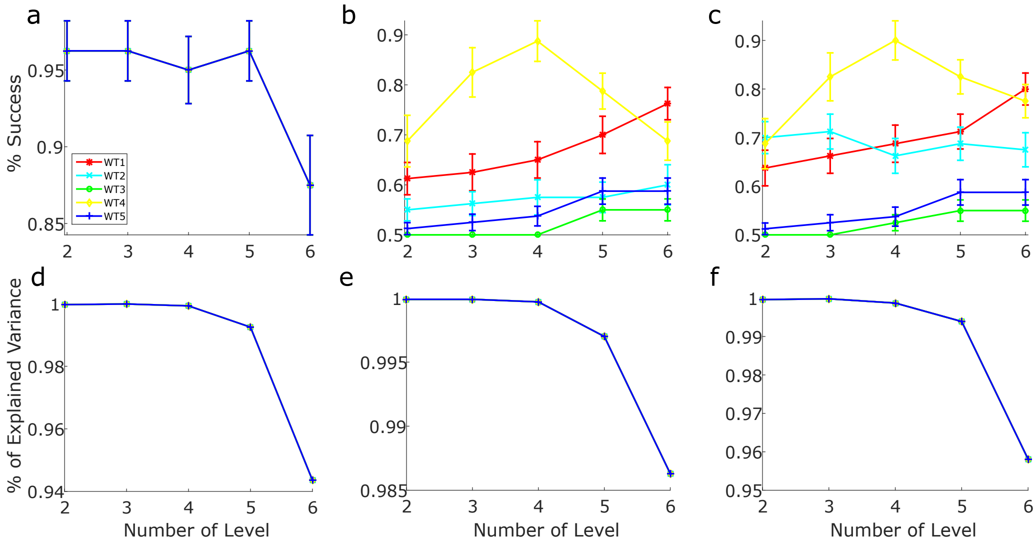


Figure B.4.: Optimal number of components for model WT on data set for the standing up exercise WT1-5 show the different classifier (a) Knee Angle only. (b) All four muscle recordings. (c) Knee Angle and the four muscles. (d-f) corresponding explained variance of the data sets in (a-c).

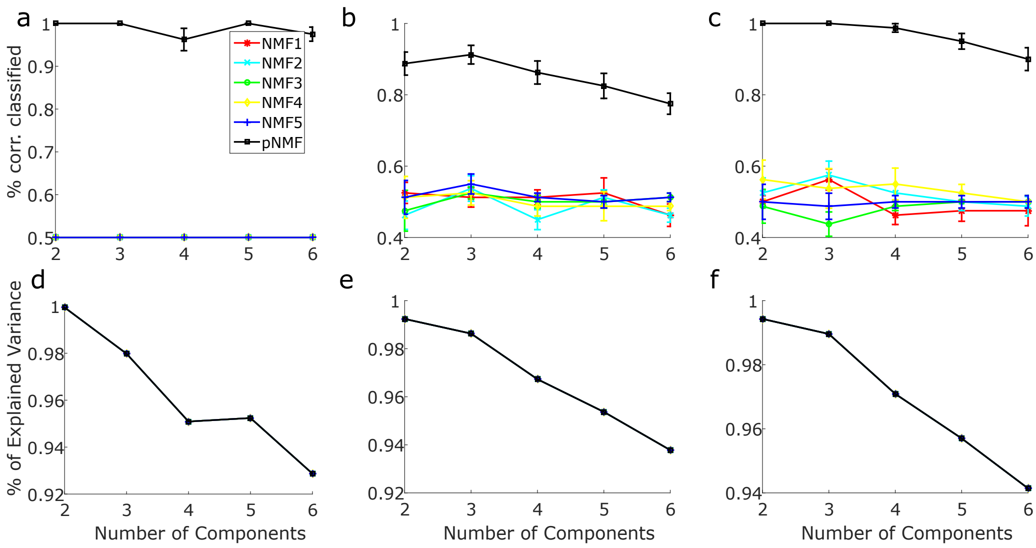


Figure B.5.: Optimal number of components for model pNMF on data set for the standing up exercise NMF1-5 and pNMF show the different classifier (a) Knee Angle only. (b) All four muscle recordings. (c) Knee Angle and the four muscles. (d-f) corresponding explained variance of the data sets in (a-c).

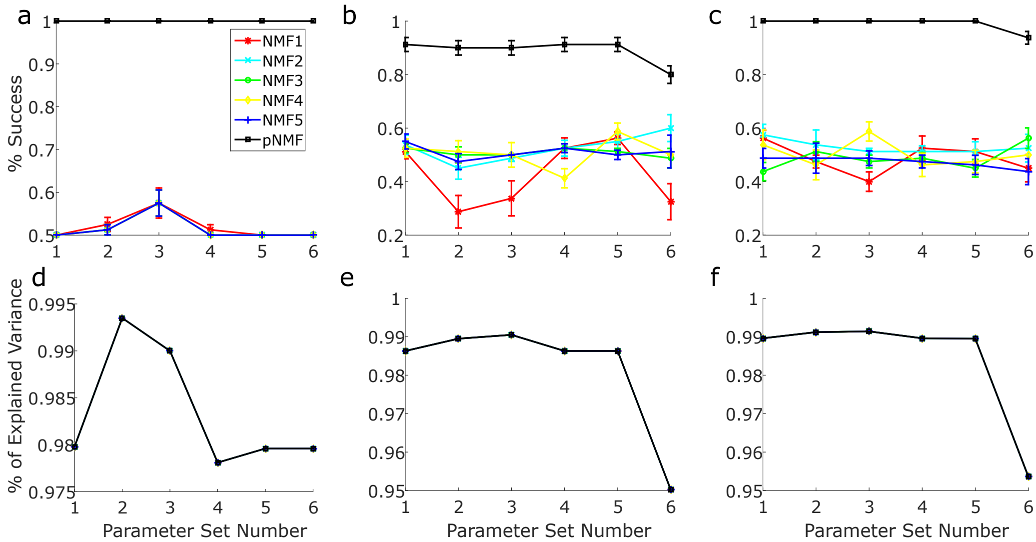


Figure B.6.: Optimal parameter settings with best determined number of components for model pNMF on data set for the standing up exercise NMF1-5 and pNMF show the different classifier (a) Knee Angle only with three components. (b) All four muscle recordings with three components. (c) Knee Angle and the four muscles with three components. (d-f) corresponding explained variance of the parameter sets in (a-c).

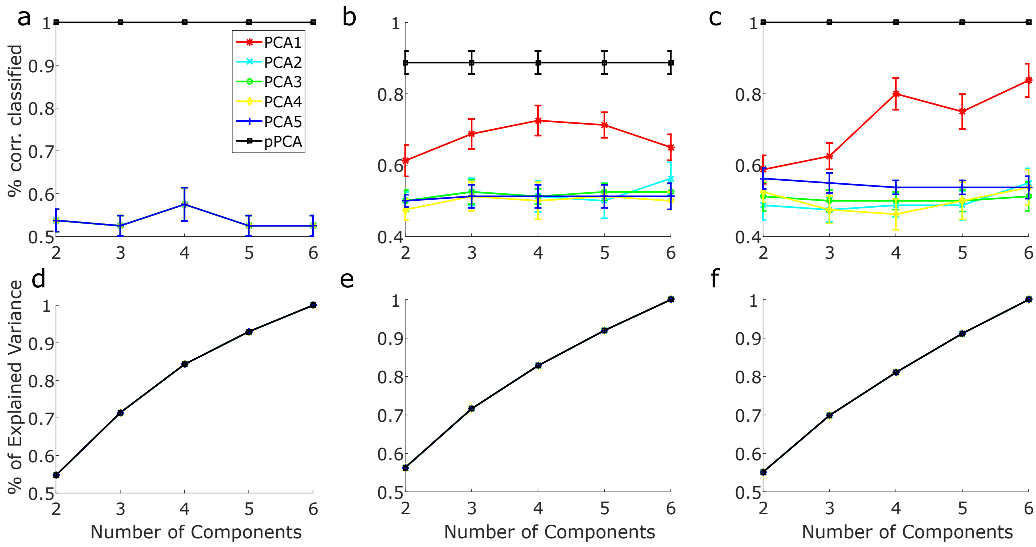


Figure B.7.: Optimal number of components for model pPCA on data set for the standing up exercise PCA1-5 and pPCA show the different classifier (a) Knee Angle only. (b) All four muscle recordings. (c) Knee Angle and the four muscles. (d-f) corresponding explained variance of the data sets in (a-c).

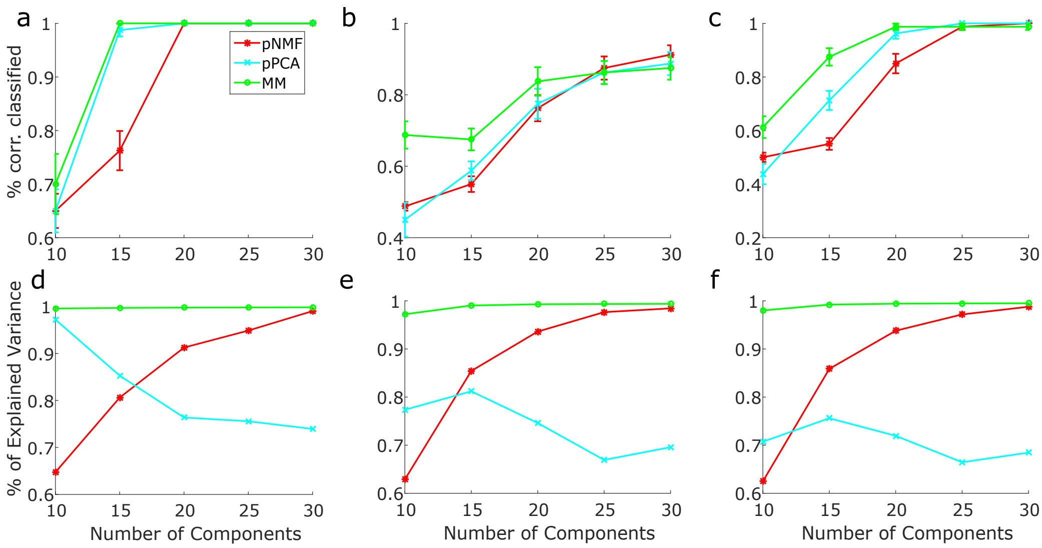


Figure B.8.: Optimal number of gaussians for model MM on data set for the standing up exercise pNMF, pPCA and MM show the different classifier (a) Knee Angle only. (b) All four muscle recordings. (c) Knee Angle and the four muscles. (d-f) corresponding explained variance of the data sets in (a-c).

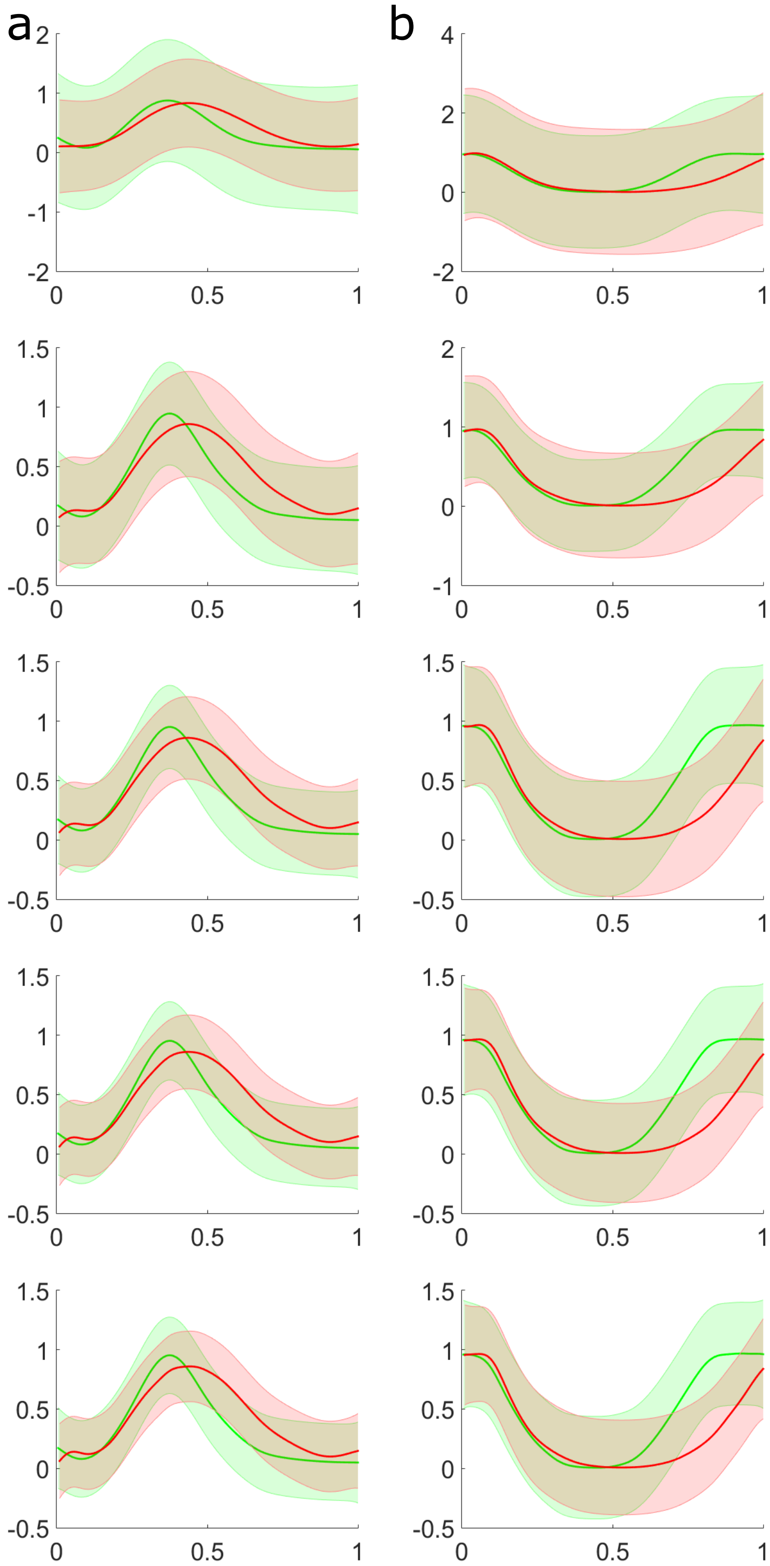


Figure B.9.: Visualisation of the MM with healthy and ill trajectories and their variance for $10, 15, 20, 25$ and 30 gaussians from top to bottom. All from the knee flexion while sitting exercise data set with: (a) single muscle recording from the biceps femoris muscle (b) knee angle recording

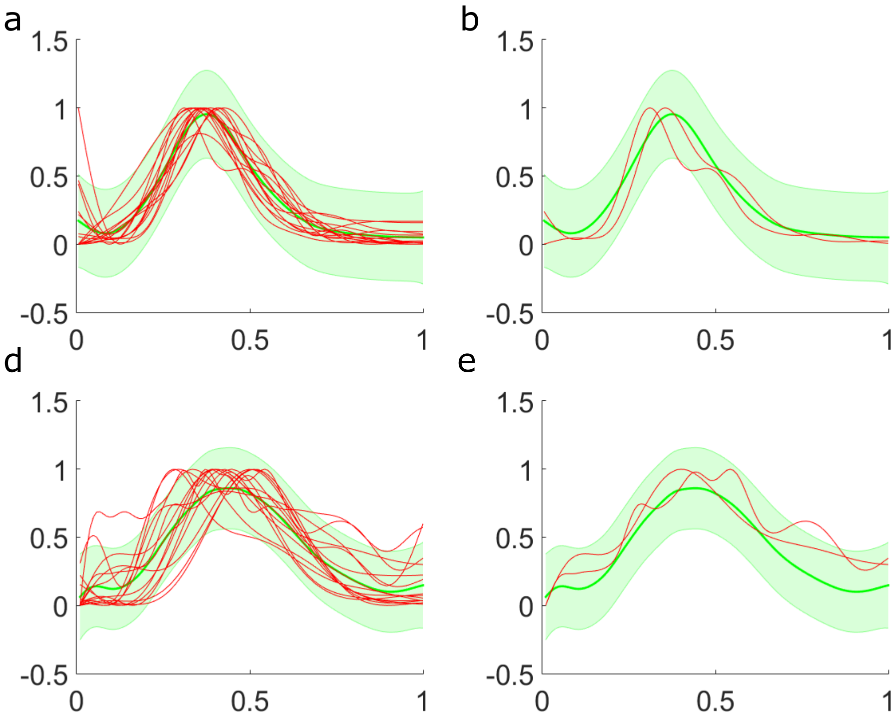


Figure B.10.: Visualization of the MM and the split of model data vs test data for the knee flexion while sitting exercise data set of the biceps femoris muscle: (a) Healthy model and model data (b) Healthy model and ill test trajectories (c) Ill model and model data (b) Ill model and ill test trajectories

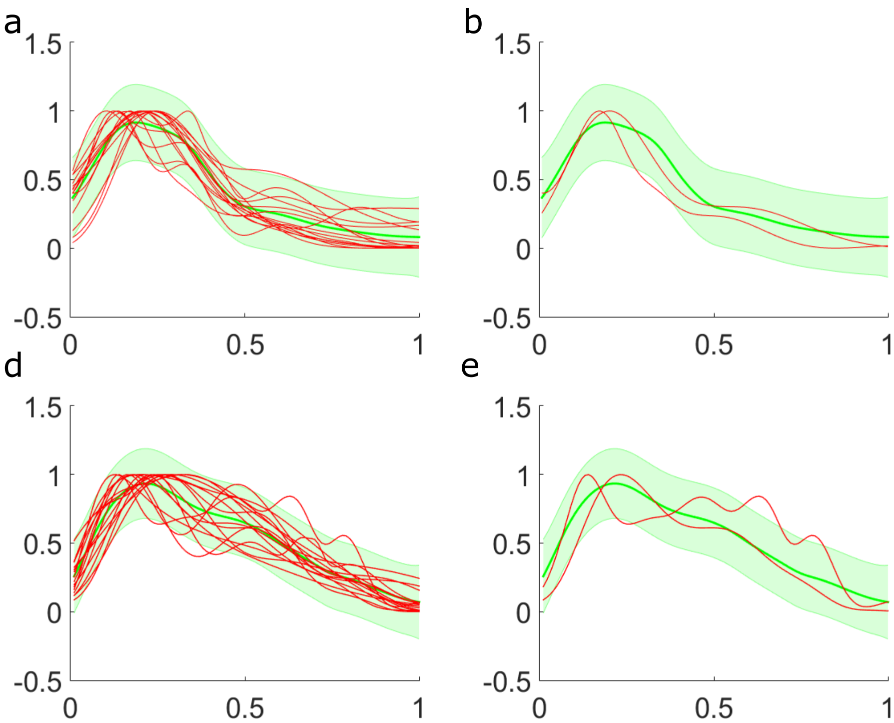


Figure B.11.: Visualization of the MM and the split of model data vs test data for the standing up exercise data set of the semitendinosus muscle: (a) Healthy model and model data (b) Healthy model and ill test trajectories (c) Ill model and model data (b) Ill model and ill test trajectories