Learning Probabilistic Features from EMG Data for Predicting Knee Abnormalities

Jan Kohlschuetter*, Jan Peters*,† and Elmar Rueckert*

Abstract—Identifying movement abnormalities from raw Electromyography (EMG) data requires three steps that are the data preprocessing, 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., principal components analysis, non negative matrix factorization, 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. In first tests, we found that these probabilistic extensions outperforms classical approaches in terms of the prediction of knee abnormalities from few samples with a performance of 86 percent of correctly classified abnormalities.

I. 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 (see, e.g., [1] for a recent tutorial). There is a wide spread use of EMG data in research disciplines such as muscle surgery [2], neurology [3], rehabilitation [4], [5], movement analysis [6], [7], biomechanics [8], [9], or 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.

For classification neural networks [11]–[13], linear discriminant analysis, kernel based methods [14], or support vector machines [15] have been proposed. For dimensionality reduction, principal components analysis (PCA) [16], non negative matrix factorization (NMF) [17] or wavelet transformation [18] were investigated. While probabilistic classifiers demonstrated to be robust in terms of signal noise [19], 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.

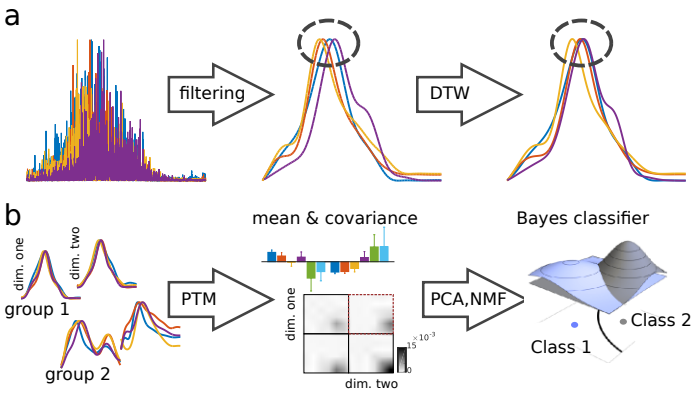

Fig. 1. **Concept of the probabilistic EMG model:** (a) EMG signals are rectified, low-pass filtered and optionally Dynamic Time Warping (DTM) is applied to correct for varying initial velocities in the trials. (b) For individual groups, the aligned EMG-channels (i.e., the first two EMG channels for 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.

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. 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.

*Intelligent Autonomous Systems Lab, Technische Universität Darmstadt, Hochschulstr. 10, 64289 Darmstadt, Germany {kohlschuetter, rueckert}@ias.tu-darmstadt.de

‡Robot Learning Group, Max-Planck Institute for Intelligent Systems, Tuebingen, Germany mail@jan-peters.net

II. A PROBABILISTIC MODEL OF EMG SIGNALS

A. Problem statement

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 $\boldsymbol{Y} = \langle \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T \rangle$ 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. Classical principal components analysis (PCA) [16] and non-negative matrix factorization (NMF) [17] 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., [12], [14], [20]. We follow here a different approach and model the data $\boldsymbol{Y}$ as a probability distribution.

B. Modeling a distribution over EMG signals

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

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

where the upper scrip ^{T} denotes the transpose operation and must not be confused with the number of discrete time steps T. As in [21] we use a Gaussian mixture model approach to represent the vector of concatenated EMG measurements $\boldsymbol{o}$ with

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

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 $\boldsymbol{\Omega} = \langle \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \ldots, \boldsymbol{\Psi}_T \rangle$. 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,

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

For each dimension denoted by i we use a vector of K scalar basis functions, i.e., $\boldsymbol{\phi}_{t,i}^T = \langle \phi_{t,1}, \phi_{t,2}, \ldots, \phi_{t,K} \rangle^T$. Different distributions are applicable for the model, depending on the task. A popular choice for rhythmic movements are Von-Mises basis functions [22], whereas for point to point movements Gaussian basis functions are widely used [21],

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

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 (1) is learned. This parameter vector is a concatenated vector of D feature vectors, one per dimension, where $\boldsymbol{w} = \left\langle \boldsymbol{w}_1^T, \boldsymbol{w}_2^T, \ldots, \boldsymbol{w}_D^T \right\rangle^T$. Note that we omitted the variances in this notation for the sake of brevity.

The covariance matrix $\boldsymbol{\Sigma}_y$ in (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 $\boldsymbol{\epsilon}_y \sim \mathcal{N}(\boldsymbol{\epsilon}_y|\boldsymbol{0}, \boldsymbol{\Sigma}_y)$.

C. Learning

In (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,

$$p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}|\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w) \quad , \tag{2}$$

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

$$\begin{aligned} p(\boldsymbol{o}) &= \int p(\boldsymbol{o}|\boldsymbol{w})\, p(\boldsymbol{w}) d\boldsymbol{w} \\ &= \mathcal{N}(\boldsymbol{o}|\boldsymbol{\Omega}\boldsymbol{w},\, \boldsymbol{\Omega}\,\boldsymbol{\Sigma}_w\,\boldsymbol{\Omega}^T + \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{y}}) \quad . \end{aligned} \tag{3}$$

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 [23], which generalizes to more complex hierarchical prior distributions. For our Gaussian prior a much simpler approach based on least squares regression was proposed [21], i.e.,

$$\boldsymbol{w}^{[m]} = (\boldsymbol{\Omega}^T\,\boldsymbol{\Omega} + \lambda\,\boldsymbol{I})^{-1}\,\boldsymbol{\Omega}^T \boldsymbol{o}^{[m]} \quad . \tag{4}$$

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'.

D. Dimensionality reduction

Let matrix $\boldsymbol{W} \in \mathbb{R}^{D \cdot K \times M}$ denote the collection of $m = 1, ..., M$ trials with $\boldsymbol{W} = \left\langle w^{[1]}, w^{[2]}, \ldots, w^{[M]} \right\rangle$. Without dimensionality reduction we argued in the previous subsection that 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.

In addition, we apply the matrix factorization techniques principal components analysis (PCA) [16] and non-negative matrix factorization (NMF) [17] to $\boldsymbol{W}$. Instead of computing the prior statistics directly from $\boldsymbol{W}$, the mean and the covariance are computed from an approximation denoted by $\tilde{\boldsymbol{W}}$.

In NMF, this approximation is given by $\tilde{\boldsymbol{W}} = V\,H$, where $V \in \mathbb{R}^{N \times R}$ is a non-negative data matrix and $H \in \mathbb{R}^{R \times M}$ denotes a weight matrix. The dimension R is chosen such that $\tilde{\boldsymbol{W}}$ is a compressed version of $\boldsymbol{W}$. We refer to this application of NMF to the probabilistic feature space as *p-NMF*. The application of PCA is straight forward and we refer to it as *p-PCA* in the results section. Note that $N = D\,K$ denotes the number of features and M is the number of samples or trials in our notation.

E. Classification

The learned prior distribution over EMG recordings in (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 (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 ($\alpha_1 = \alpha_2 = 0.5$).

F. Relationship to muscle synergy models

The generative probabilistic model in (3) can be related to time-invariant [24] and time-varying [25] muscle synergy models. Time-invariant synergies are represented as a set of shared synergy vectors $\boldsymbol{v}$ that is scaled by task dependent time-varying temporal profiles

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

Time-varying muscle synergy models [25] generate EMG measurements as weighted sum over time-shifted synergy profiles

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

where the activity vector $\boldsymbol{v}_k(t)$ 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 $\boldsymbol{v}_k$.

Both generative laws relate to a single time step prediction in (1), where the basis function matrix $\boldsymbol{\Psi}_t$ is shared among tasks (like $\boldsymbol{v}_k$) and the learnable feature vector $\boldsymbol{w}$ becomes task dependent. Such task dependent feature weights were used in [23] for transfer learning. Note that time shifts and task dependent activity vectors $\boldsymbol{v}_k^m$ 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 [24], [25]. 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 [24], [25].

III. Results

A. Data and Preprocessing

We evaluated the proposed EMG models on a clinical lower limb data set [26], 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.

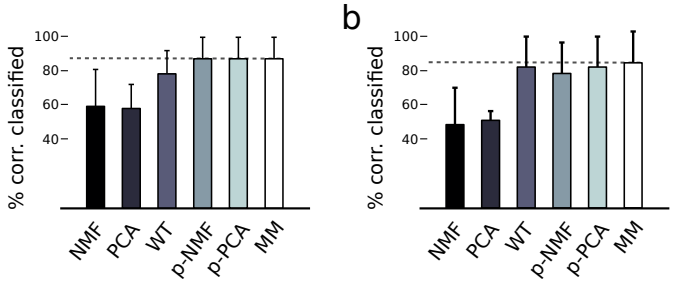

Fig. 2. **Classification performance of deterministic and probabilistic approaches:** Comparison of the deterministic methods PCA, NMF and WT to the proposed probabilistic variants p-PCA, p-NMF and MM. (a) Correctly classified subjects for the knee flexion while sitting exercise. (b) Classification success rate in percent on data recorded in the standing up exercise.

TABLE I
Classification results for an increasing number of components. Shown is the classification performance and the F-score in parentheses. For WT the number of components were 10, 13, 19, and 32.

#comp.:	**2**	**3**	**4**	**5**
PCA	0.55 (0.67)	0.56 (0.68)	0.56 (0.69)	**0.60 (0.7)**
NMF	0.53 (0.3)	**0.62 (0.59)**	0.55 (0.57)	0.57 (0.65)
WT	0.77 (0.75)	**0.81 (0.8)**	0.80 (0.77)	0.81 (0.79)
p-PCA	**0.86 (0.86)**	0.86 (0.86)	0.86 (0.86)	0.86 (0.86)
p-NMF	**0.86 (0.86)**	0.86 (0.86)	0.86 (0.86)	0.86 (0.86)

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 with a sample rate of 1000 Hz. For more details we refer to [26]. 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. Four samples were used for testing in cross validation with 20 sets. The knee angle was only used to align the trajectories using dynamic time warping [27] and the models were trained with the four EMG signals scaled to $T = 300$ time steps.

B. Probabilistic models outperform PCA, NMF and WT

We compared the prediction performance of the proposed features space models to standard principal components analysis (PCA) [16], classical non-negative matrix factorization (NMF) [17] and wavelet transformation (WT) [18]. The investigated probabilistic approaches are a Gaussian mixture model denoted by MM (without dimensionality reduction) and two extensions where we applied PCA and NMF in the Gaussian features space (denoted as p-PCA and p-NMF).

For both exercises, we found that the three probabilistic approaches outperform standard PCA, NMF and WT, see Figure 2. For all approaches the optimal number of (principal) components were determined, i.e., PCA (5), NMF (3), WT (13 = level 5), p-PCA (2), p-NMF(2). For the probabilistic trajectory model 30 Gaussians were sufficient to model the EMG data. Details are provided in Table I. Note that for the limited data set more than two components did not change the classification strategy for the p-PCA and the p-NMF method on the same data set.

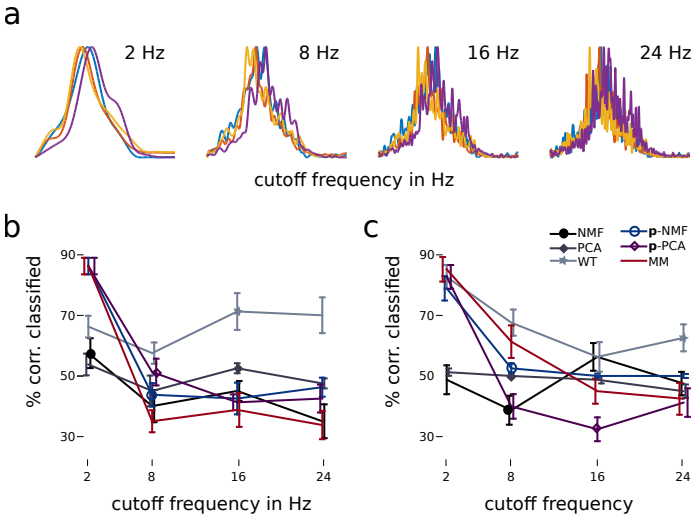

Fig. 3. **Investigation of the robustness to Gaussian noise:** (a) From left to right, the raw EMG signals were low-pass filtered with an increasing cutoff frequency to simulate additive measurement noise. (b) Classification performance for an increasing cutoff frequency on the knee flexion exercise. (c) Results for the standing up exercise.

C. Feature models are less sensitive to noise

We investigated the effect of Gaussian noise on the EMG signal 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 3(a) for sample recordings of the vastus internus.

The methods were optimized for a cutoff frequency of 2 Hz, see Table I. With an increasing cutoff frequency the performance decreased for both exercises. Note that cutoff frequencies smaller than 2 Hz were not applicable due to numerical instabilities and poor classification results. For the knee flexion while sitting exercise the performance values are shown in Figure 3(b) and for the standing up exercise the results are illustrated in Figure 3(c). For both exercises, the NMF and the PCA method tend to have inferior performance compared to the four feature based methods (WT, p-NMF, p-PCA, and MM). However, for significant results investigations on larger data sets are necessary.

IV. Discussion

In this work we introduced three new probabilistic models of EMG signals and evaluated them on two limited data sets with 22 subjects performing three to five motions. The good results in spite of working with the limited data sets only showed how much potential our proposed models have. However, for meaningful conclusions when comparing to traditional approaches larger data sets need to be evaluated. Not only the data sets themselves but also the number of recorded muscles is a crucial factor that needs to be investigated further. We belief that models of more than four EMG channels could greatly improve the classification rates. In particular, the number of extracted components is limited be the number of recorded muscles and the question of independently activated versus co-activated muscle patterns remains open. (Note that the shapes of all four EMG signals are very similar in our data set). Furthermore, the tests and the applied optimizations in our robustness against noise study should be extended.

V. Conclusion

Electromyography (EMG) signals in, e.g., prosthetic and rehabilitation tasks [4], [9], [28], [29] are typically corrupted by sensor noise, the surface electrodes' position might change, and even for the same executed movement different EMG patterns are observed (known as motor variability). 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 [30] and was previously used as part of a movement representation in robotics [21]. It can be trained through least squares regression or variational inference and scales to more complex hierarchical Bayesian models [23] relevant for EMG applications.

In this paper, we extended the model by applying the dimensionality reduction techniques principal components analysis (PCA) [16] and non-negative matrix factorization (NMF) [17] to the learned Gaussian features. We evaluated the resulting approaches in a clinical lower limb data set [26] with the task of predicting knee abnormalities. We found that the proposed models outperform standard PCA, NMF and wavelet transformation [18] in terms of classification performance. First tests on the limited data set indicate that the feature based models are less sensitive to noise.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreements No. 610967 (TACMAN) and No. 600716 (CoDyCo). The authors would like to thank Guilherme Maeda and Rudolf Lioutikov for comments that greatly improved the manuscript.

References

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

[2] John E Duggan and Gordon B Drummond. Abdominal muscle activity and intraabdominal pressure after upper abdominal surgery. *Anesthesia & Analgesia*, 69(5):598–603, 1989.

[3] Nihal Fatma Güler and Sabri Koçer. Use of support vector machines and neural network in diagnosis of neuromuscular disorders. *Journal of medical systems*, 29(3):271–284, 2005.

[4] Claudio Castellini and Patrick van der Smagt. Surface emg in advanced hand prosthetics. *Biological cybernetics*, 100(1):35–47, 2009.

[5] D Matheson Rittenhouse, Hussein A Abdullah, R John Runciman, and Otman Basir. A neural network model for reconstructing emg signals from eight shoulder muscles: Consequences for rehabilitation robotics and biofeedback. *Journal of biomechanics*, 39(10):1924–1932, 2006.

[6] Da-peng Yang, Jing-dong Zhao, Yi-kun Gu, Xin-qing Wang, Nan Li, Li Jiang, Hong Liu, Hai Huang, and Da-wei Zhao. An anthropomorphic robot hand developed based on underactuated mechanism and controlled by emg signals. *Journal of Bionic Engineering*, 6(3):255–263, 2009.

[7] Sebastian Bitzer and Patrick 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*, pages 2819–2823. IEEE, 2006.

[8] J Lucas McKay and Lena H Ting. Optimization of muscle activity for task-level goals predicts complex changes in limb forces across biomechanical contexts. *PLoS Comput. Biol*, 8(4):e1002465, 2012.

[9] Zeeshan O Khokhar, Zhen G Xiao, Carlo Menon, et al. Surface emg pattern recognition for real-time control of a wrist exoskeleton. *Biomedical engineering online*, 9(1):41, 2010.

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

[11] Akira Hiraiwa, Katsunori Shimohara, and Yukio Tokunaga. Emg pattern analysis and classification by neural network. In *Systems, Man and Cybernetics, 1989. Conference Proceedings., IEEE International Conference on*, pages 1113–1115. IEEE, 1989.

[12] Christodoulos Christodoulou, Constantinos S Pattichis, et al. Unsupervised pattern recognition for the classification of emg signals. *Biomedical Engineering, IEEE Transactions on*, 46(2):169–178, 1999.

[13] Gang Wang, Zhizhong Wang, Weiting Chen, and Jun Zhuang. Classification of surface emg signals using optimal wavelet packet method based on davies-bouldin criterion. *Medical and Biological Engineering and Computing*, 44(10):865–872, 2006.

[14] Sumitra S Nair, Robert M French, Davy Laroche, and Elizabeth Thomas. The application of machine learning algorithms to the analysis of electromyographic patterns from arthritic patients. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 18(2):174–184, 2010.

[15] Abdulhamit Subasi. Classification of emg signals using pso optimized svm for diagnosis of neuromuscular disorders. *Computers in biology and medicine*, 43(5):576–586, 2013.

[16] MBI Reaz, MS Hussain, and Faisal Mohd-Yasin. Techniques of emg signal analysis: detection, processing, classification and applications. *Biological procedures online*, 8(1):11–35, 2006.

[17] Matthew C Tresch, Vincent CK Cheung, and Andrea d'Avella. Matrix factorization algorithms for the identification of muscle synergies: evaluation on simulated and experimental data sets. *Journal of Neurophysiology*, 95(4):2199–2212, 2006.

[18] Ingrid Daubechies et al. *Ten lectures on wavelets*, volume 61. SIAM, 1992.

[19] Adrian DC Chan and Kevin B Englehart. Continuous myoelectric control for powered prostheses using hidden markov models. *Biomedical Engineering, IEEE Transactions on*, 52(1):121–124, 2005.

[20] Emmanouil Benetos, Margarita Kotti, and Constantine 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.

[21] 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.

[22] Elmar Rueckert and Andrea d'Avella. Learned parametrized dynamic movement primitives with shared synergies for controlling robotic and musculoskeletal systems. *Frontiers in computational neuroscience*, 7, 2013.

[23] 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.

[24] Yuri P Ivanenko, Richard E Poppele, and Francesco Lacquaniti. Five basic muscle activation patterns account for muscle activity during human locomotion. *The Journal of physiology*, 556(1):267–282, 2004.

[25] Andrea d'Avella, Philippe Saltiel, and Emilio Bizzi. Combinations of muscle synergies in the construction of a natural motor behavior. *Nature neuroscience*, 6(3):300–308, 2003.

[26] M. Lichman. UCI machine learning repository. http://archive.ics.uci.edu/ml/datasets/EMG+dataset+in+Lower+Limb, 2013.

[27] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.

[28] He Huang, Todd Kuiken, Robert D Lipschutz, et al. A strategy for identifying locomotion modes using surface electromyography. *Biomedical Engineering, IEEE Transactions on*, 56(1):65–73, 2009.

[29] Ali H Al-Timemy, Guido Bugmann, Javier Escudero, and Nicholas Outram. Classification of finger movements for the dexterous hand prosthesis control with surface electromyography. *Biomedical and Health Informatics, IEEE Journal of*, 17(3):608–618, 2013.

[30] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.