

Muscle Categorization using Quantitative Needle Electromyography A 2-Stage Gaussian Mixture Model based Approach

Meena AbdelMaseeh*, Pascal Poupart[†], Benn Smith[‡] and Daniel Stashuk*
m2adly@uwaterloo.ca

**Systems Design Engineering*

†David R. Cheriton School of Computer Science

**,†University of Waterloo, ON, Canada*

‡Department of Neurology

Mayo Clinic, Scottsdale, AZ, USA

Abstract—Needle Electromyography, in combination with nerve conduction studies, is the gold standard methodology for assessing the neurophysiologic effects of neuromuscular diseases. Muscle categorization is typically based on visual and auditory assessment of the morphology and activation patterns of its constituent motor units. A procedure which is highly dependent on the skills and level of experience of the examiner. This motivates the development of automated or semi-automated categorization techniques.

This paper describes a 2-stage Gaussian mixture model based approach. In the first stage, a muscle is classified as neurogenic or myopathic. The second stage uses a classifier specific to each disease category to confirm or refute the disease involvement. A total of 2556 motor unit potentials sampled from 48 normal, 30 neurogenic and 20 myopathic tibialis anterior muscles were utilized for this study. The proposed approach showed an average accuracy of 91.25%, which is higher than the compared linear and non-linear multi-class schemas. In addition to improved accuracy, the 2-stage approach is more suitable for the muscle categorization, because it has a hierarchical decision structure similar to current clinical practice, and its output can be interpreted as a measure of confidence.

Keywords—Needle EMG, Decomposition based quantitative EMG, Muscle categorization, 2-stage approach, Gaussian Mixture Model.

I. INTRODUCTION

A motor unit (MU) comprises an alpha motor neuron and all the muscle fibres it drives. Fibres forming a MU intermingle with fibres of other MUs belonging to the same muscle. Therefore the activation of a single motor neuron will produce a distributed contraction across a muscle.

MUs need to be repeatedly activated to maintain or increase muscle force. The neural activation of each MU can be modelled as a train of impulses. To acquire needle electromyographic (EMG) signals, a concentric needle or a monopolar electrode is inserted through the skin into the subcutaneous tissue or superficial layers of a muscle. Even though the fibres of the same MU are activated almost simultaneously, their contributions to a detected motor unit potential (MUP) are not identical and do not coincide with each other. Hence, each MU will produce a unique motor

unit potential train (MUPT) which is representative of the activation and morphology of its fibres.

For a given acquisition protocol (needle type and level of muscle activation), MUPs belonging to the same MUPT are expected to have some inherent physiological shape variability but yet to have similar morphology compared to the MUPs created by other MUs. A detected needle EMG signal is the superposition of the MUPTs created by the MUs contributing to a muscle contraction.

Disorders affecting muscles can be broadly classified as myopathic or neurogenic. Both of these disorders affect the MUP morphology and activation patterns of the affected MUs. Myopathic disorders are characterized by loss of muscle fibres. Therefore more MUs are recruited and required to fire more often to produce a specific level of muscle force. On the other hand, neurogenic disorders lead to the loss of motor neurons. The denervated muscle fibres are reinnervated by axonal sprouts of other surviving MUs, resulting in MUs with more distributed and larger numbers of fibres. Fewer MUs are then recruited to reach a specific level of muscle force.

During the 1940s and 1950s, many studies like [5] correlated certain disorders to variations in the morphology and frequency of the resulting MUPs. In 1954, Buchthal et al [1] introduced the use of quantitative analysis of EMG signals or quantitative electromyography (QEMG). QEMG extracts morphological characteristics describing isolated MUPs. Muscle characterization is then based on comparison of the extracted features to reference data of matching age range, gender, electrode type and other physical parameters. QEMG is time consuming, when performed manually. Therefore, current clinical practice is still based on qualitative visual and auditory assessment. This assessment is highly subjective and dependent on experience and skill, which motivates the development of automated (or semi-automated) systems to support diagnosis which use QEMG data.

Suitable quality QEMG data can currently be automatically obtained using the decomposition based quantitative EMG

(DQEMG) technique described in [10]. However, it is still not straight forward how to qualitatively create a muscle characterization using the QEMG data provided, because of the number of different features generated, dependence among features, and the overlap of feature value probability distributions between categories. The focus of this study is to provide an accurate and consistent muscle characterization utilizing features extracted by DQEMG. This problem is challenging for the following reasons:

- The level of disease involvement is not uniform among all the MUs of the same muscle. For example, a neurogenic muscle may have slightly affected, severely affected and normal MUs. There is no clear cut rule when to consider a muscle to be affected. Clinicians normally will consider a muscle to be affected, when they observe "enough" significantly affected MUs. This demands that the characterization of each MU be in the form of a score reflecting disease involvement to allow aggregation of these scores into an overall muscle characterization.
- Most of the estimated features suffer from dependence on the acquisition protocol. The two main sources of variations are focus (i.e., adjusting the electrode position to acquire suitably sharp MUPs) and the level of contraction [9]. These undesired variations may obscure discriminability.
- The training dataset is annotated at the muscle level, consequently each MUP label represents the condition of the muscle to which the MU belongs rather than the actual condition of the MU. The dataset is also highly unbalanced due to the fact that normal muscles are more likely to happen and the number of sampled MUs may vary across different muscles.
- Biovariability, poor understanding of the underlying electrophysiological process, and variation in acquisition protocols lead to poor agreement among annotators. For instance, Kendall and Werner [4] showed that agreement among 31 examiners for determining the exact level of neuropathy was only 46.9%. The labelling of the training or testing data therefore can't be considered certain nor accurate in all cases.

II. LITERATURE REVIEW

In [6], the performance of an artificial neural network (ANN) trained using both supervised and unsupervised learning was evaluated. The performance of an ANN and the K-nearest neighbour (KNN) technique was also compared. The feature vector was composed of the standard deviation and mean of 7 time domain features. For supervised training, back propagation was utilized, while a self organizing feature map algorithm was used for the unsupervised paradigm. Both ANN paradigms showed comparable performance with diagnostic accuracy of 80%, which substantially exceeded the KNN.

In [7], linear discriminant analysis was used to map duration, area, turns, and central frequency features into 2 scores. The mapping was a linear function of the features and was inferred using the training data, so as to maximize the separability among scores for different classes. The utilized algorithm also excluded features that didn't improve discrimination. The centroid of each categorization was then evaluated. During classification, the posterior probability was computed using Bayes' rule. The likelihood probabilities were estimated using the Euclidean distance from class centroids assuming features were normally distributed with equal covariance for all classes. For the first MUP, the prior probabilities of the classification were based on epidemiological data. The posterior probability of the previous MUP was then used as the prior probability for subsequent MUPs. Experimentation showed diagnostic probabilities above 0.95 in 91% of 223 biceps brachii muscles from 80 patients.

During training in [8], repeatable patterns were discovered from quantized feature vectors. A weight of evidence (WOE) measure was then estimated to reflect the support or refutation of each of the patterns to one of the classes. During classification, the summations of the WOE measures for all patterns for the three classes were then used to evaluate the posterior probabilities. These MUP characterizations were then aggregated using Baye's rule or averaged to produce a muscle characterization. Experimentation based on both clinical and simulated data using only 4 time domain features showed a comparable performance to the method described in [7]. The WOE measure can also be interpreted as a level of disease involvement [8].

The method described in [3] is different than the previous 3 approaches, because it uses the raw data of isolated MUPs. A feature vector constructed from the time samples of 25 msec of raw EMG data around the peak of a MUP was used to train 3 support vector machines in a one-against-one setting. No attempt to categorize the muscles was made. Instead, an expert was asked to label MUP templates estimated by a previous stage of the algorithm.

The work presented here is based on a 2-stage approach. In the first stage, a classifier decides whether a muscle is neurogenic or myopathic. The level of disease involvement (i.e., whether the muscle is in fact diseased or normal) is investigated in the second stage using a different classifier for each disease category. The reasoning behind this approach is that for most of the QEMG morphological features, normal MUP values tend to lie between the neurogenic and myopathic values and therefore the neurogenic and myopathic classes are well separated. This work also compares the proposed approach to conventional multi-class linear and non-linear schemas.

III. METHODOLOGY

A. Overview

The approach is summarized in Figure 1. During decomposition, DQEMG [10] segments the MUPs. It then uses a variation of k-means clustering to estimate the number of MUs contributing significantly to the acquired EMG signal (i.e., the number of MUPTs expected to account for the majority of the energy of the acquired signal). It then employs supervised classification techniques to determine the membership of each MUP to one of the expected MUPTs.

Later, each MUPT is represented using a morphological template and quantitatively described using a feature vector X_{MUP} . In the first stage, a subvector $X_{MUP_i}^*$ is extracted and used to characterize the MUP with a Gaussian mixture model (GMM) to evaluate $P(Y|X_{MUP_i}^*)$, which is the likelihood that a given MUP was sampled from a myopathic or neurogenic muscle (i.e., $Y \in \{myo, neuro\}$). The characterizations of the N MUPs are then aggregated to estimate an overall muscle characterization $P(Y|X_{MUP_1}^* \dots X_{MUP_N}^*)$. If $P(Y = neuro|X_{MUP_1}^* \dots X_{MUP_N}^*)$ is higher than $P(Y = myo|X_{MUP_1}^* \dots X_{MUP_N}^*)$ in the first stage, a possibly different subvector $X_{MUP_i}^{\prime}$ is then extracted to characterize the MUP in the second stage with a GMM to evaluate $P(Y_n|X_{MUP_i}^{\prime})$ where $Y_n \in \{neuro, normal\}$. The MUP characterizations are then aggregated to estimate muscle characterization $P(Y_n|X_{MUP_1}^{\prime} \dots X_{MUP_N}^{\prime})$. The muscle is categorized as belonging to the muscle category Y_n with the higher muscle characterization. The same steps are applied when $P(Y = myo|X_{MUP_1}^* \dots X_{MUP_N}^*)$ is higher, but using a GMM to evaluate $P(Y_m|X_{MUP_i}^{\prime})$ where $Y_m \in \{myo, normal\}$, and possibly utilizing a different subvector $X_{MUP_i}^{\prime}$.

B. Data Acquisition

The needle EMG data was obtained using a disposable concentric needle electrode. Subjects maintained approximately 15 seconds of isometric voluntary contraction at a level that created a detected EMG signal with between 40 to 60 MUPs per second. Data was obtained from 48 normal, 30 neurogenic and 20 myopathic tibialis anterior muscles annotated by an expert neurophysiologist.

MUPs are therefore annotated with the label of the muscle in which they were detected and regardless of their actual condition. This resulted in 1342 MUPs detected in normal muscles and 659 and 555 MUPs detected in neurogenic and myopathic muscles, respectively. All data was acquired under institutional review board approval and sanitized of any personal identifying information.

C. QEMG Features

Fifteen features were used to describe each MUPT. They can be broadly classified into 3 categories:

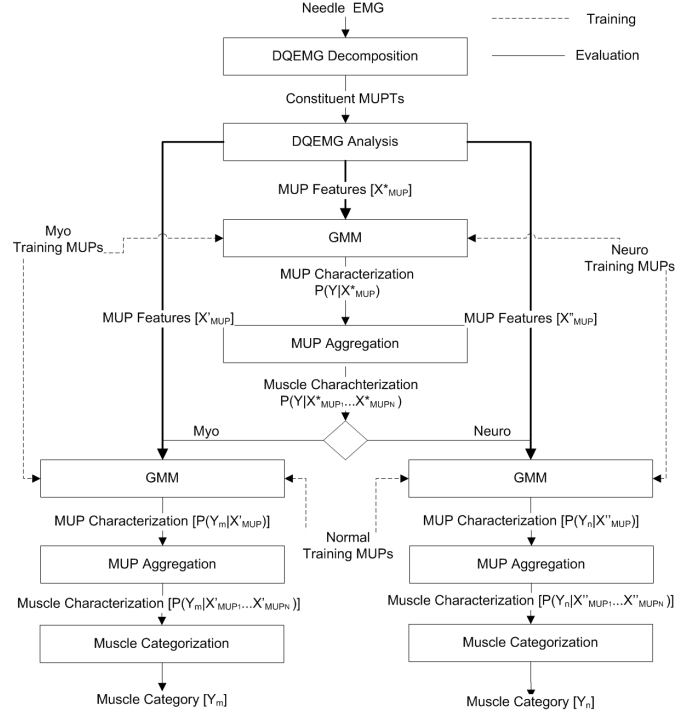


Figure 1. 2-stage GMM based approach block diagram.

- **MUP Template morphology features:** Which can be further classified as size features (e.g. area, amplitude, size index and duration) or complexity features (e.g. length, phases and turns). Size features mostly capture global aspects of the MUP template, while complexity features provide local information.
- **Stability features:** QEMG features (e.g., jiggle and shimmer covariance) represent the MUP shape consistency across the constituent MUPs of the same MUPT. For example, the normalized standard deviation of the distances of the MUPs of a train to its MUP template.
- **Background signal features:** These features describe the background activity from which MUPs are detected. For instance, acceleration ensemble RMS is the root mean square of a fixed length window before a MUP.

Other features are derived by mapping and/or normalization of the above features like area to amplitude ratio and size index. The resulting features were shown to be immune to acquisition variations or more discriminative [10].

D. MUP Characterization using GMM

MUP characterization is the evaluation of the probability that the MUP under investigation was sampled from a muscle belonging to a specific class $P(Y_k|X_{MUP})$, where Y_k is the class label. In the multi-class schema, $Y_k \in \{normal, neurogenic, myopathic\}$. In the 2-stage approach, the domain of Y_k depends on what the GMM models. For example, $Y_k \in \{neurogenic, myopathic\}$ in

the first stage.

X_{MUP} refers to the feature vector evaluated using DQEMG corresponding to a given MUPT or a subvector including features that showed most relevance to the model. The GMM is a generative probabilistic model. The main assumption is

$$P(X_{MUP}|Y_k) = N(X_{MUP}|\mu_k, \Psi) \quad (1)$$

where $N(X_{MUP}|\mu, \Psi)$ is a multivariate Gaussian distribution. The posterior probability is then calculated using Bayes' rule:

$$P(Y_k|X_{MUP}) = \frac{P(X_{MUP}|Y_k)P(Y_k)}{\sum_k P(X_{MUP}|Y_k)P(Y_k)} \quad (2)$$

During estimation, maximum likelihood is used to estimate the mean feature vector μ_k for each class Y_k , and the initial probability $P(Y_k)$ for each class. The averaged covariance matrix Ψ is estimated using:

$$\Psi = \sum_k \Psi_k P(Y_k) \quad (3)$$

E. MUP Characterization using Artificial neural networks

A feed forward ANN is a parametric non-linear mapping of a feature vector. The model parameters are the weights connecting the nodes. During training, the back propagation algorithm was used. The selected transfer function for each node was a tan-sigmoid and the hidden layer included 18 nodes. The output of each output node $z_k(X_{MUP})$ corresponds to a different category.

F. Muscle Characterization

A muscle characterization is the probability that the muscle belongs to a specific class Y_K given the set of MUPS detected from it. In this work, it was calculated as the mean of the MUP characterizations.

Assuming a specific muscle sampling resulted in N MUPS, the muscle characterization can be estimated as:

$$P(Y_k|X_{MUP_1}, \dots, X_{MUP_N}) = \frac{\sum_{i=1}^N P(Y_k|X_{MUP_i})}{N} \quad (4)$$

The same is valid for the ANN, but $P(Y_k|X_{MUP_i})$ is replaced by the node output $z_k(X_{MUP_i})$.

G. Validation and performance metrics

Leave-one-out cross-validation was used. At each iteration, the feature vectors of the MUPS of a given muscle were used for testing, and the feature vectors of the remaining MUPS were used for training. For the ANN case, randomly selected balanced sets for each category were used for the training to avoid bias.

H. Feature Selection

A variation of backward feature selection described in Algorithm 1 was used to select the most relevant features. The objective function was chosen to be the average misclassification rate. The feature selection for the 2-stage approach was performed per classifier in sequence.

Algorithm 1 Backward Feature Selection

Require: Initial feature set Λ

$N \leftarrow |\Lambda|$

$\lambda_N \leftarrow \Lambda$

for $n = N - 1$ **downto** 1 **do**

 // Form all subsets of length n

$A = \{S | S \subseteq \lambda_{n+1} \wedge |S| = n\}$

$A_{best} = \{S^* | S^* = \operatorname{argmin}_{S \in A} \operatorname{obj}(S)\}$

 // $\operatorname{obj}(\cdot)$ is the objective function.

if $|A_{best}| = 1$ **then**

$\lambda_n \leftarrow S^*$ such that $A_{best} = S^*$

end if

end for

Return $\operatorname{arg min}_{\lambda_n} \operatorname{obj}(\lambda_n)$

IV. RESULTS

Tables I, II, and III show the confusion matrices for the discussed methods before and after the feature selection. Using algorithm 1, three features were selected for the first stage of the 2-stage approach, 5 features for the classifier responsible for the myopathic candidates and 6 features for the classifier responsible for the neurogenic candidate. The multi-class GMM utilized 8 features and the multi-class ANN utilized 9 features.

V. DISCUSSION

The 2-stage approach was compared to GMM and ANN conventional multi-class schema to evaluate its utility. These two methods were selected from a large of spectrum of machine learning methods, because they span different classification paradigms. The GMM method attempts to infer its hypothesis as a regularized linear combination of fixed mappings of the feature vector. This is true when the same covariance matrix is used for all conditional probabilities. On the other hand, an ANN uses a nonlinear parametrized mapping.

The results show that the 2-stage approach performed better than the multi-class GMM. These results are reasonable because cascading linear classifiers results in a more flexible non-linear separation hyper-plan. Results also show that the 2-stage approach and the multiclass GMM both performed better than the multi-class ANN, which might be attributed to the fact that an ANN is more likely to suffer from over fitting.

Figure 2 shows that all misclassifications occur at relatively low values of posterior probability. This suggests the utility of the estimated posterior probability as a measure of confidence. Clinicians can be alerted to do more careful analysis when a categorization is not based on a high posterior probability. This quantitative analysis and automated characterization is not designed to replace current clinical practice, because diagnosis is also based on accumulation

Table I
2-STAGE GMM CONFUSION MATRIX

	No Feature Selection				After Feature Selection		
	Classified as				Classified as		
	Normal	Myopathic	Neurogenic		Normal	Myopathic	Neurogenic
Normal	89.58%	0%	10.42%	Normal	93.75%	0	6.25%
Myopathic	10%	90%	0	Myopathic	10%	90%	0
Neurogenic	16.67%	0	83.33%	Neurogenic	10%	0	90%
Average Accuracy			87.63%	Average Accuracy			91.25%

Table II
MULTI-CLASS GMM CONFUSION MATRIX

	No Feature Selection				After Feature Selection		
	Classified as				Classified as		
	Normal	Myopathic	Neurogenic		Normal	Myopathic	Neurogenic
Normal	85.42%	4.16%	10.42%	Normal	89.58%	8.33%	2.08%
Myopathic	15%	85%	0	Myopathic	10%	90%	0
Neurogenic	16.67%	0	83.33%	Neurogenic	6.67%	0	93.33%
Average Accuracy			84.58%	Average Accuracy			90.97%

Table III
MULTI-CLASS ANN CONFUSION MATRIX

	No Feature Selection				After Feature Selection		
	Classified as				Classified as		
	Normal	Myopathic	Neurogenic		Normal	Myopathic	Neurogenic
Normal	90.94%	0.52%	8.54%	Normal	86.25%	2.92%	10.83%
Myopathic	12%	88%	0	Myopathic	13%	87%	0
Neurogenic	17.67%	0.33%	82.00%	Neurogenic	8.67%	0	91.33%
Average Accuracy			86.98%	Average Accuracy			88.19%

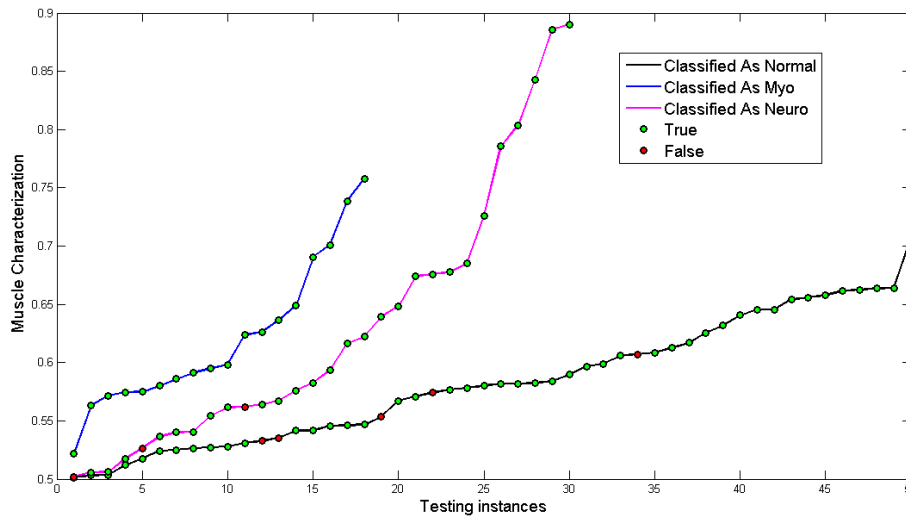


Figure 2. The output of the 2-stage approach after feature selection. The muscle characterizations of each muscle category are sorted in ascending order and shown in a different trace. The green dots represent correctly categorized instances, while the red dots represent misclassified instances.

of evidences from other sources including patient history, clinical and physical examinations.

Notwithstanding the only marginal, and probably not clinically important, improvement in accuracy of the 2-stage approach, we believe that the 2-stage approach is more ap-

propriate for muscle categorization, because of the following reasons:

- The categorization of the second stage of the 2-stage approach can be interpreted as the likelihood of being diseased as compared to normal. Moreover, the

estimated posterior probabilities can be utilized as a measure of confidence.

- The 2-stage approach searches a less complicated hypothesis space than the multi-class GMM and ANN. Therefore, confidence in the empirical risk is higher for a given sample size.
- GMM parameter estimation is much faster than the training of an ANN. This is crucial for this application, because the system user may wish to add to or tune the labelled examples.
- After feature selection, the 2-stage GMM resulted in better performance using fewer features. This could be attributed to the fact that the feature selection works better for each classifier independently.
- The 2-stage approach matches more closely the decision logic currently used by clinicians.

The choice of a GMM to model the conditional probabilities in the two stages is justifiable, because the marginal conditional probabilities of each feature is either normal or log normal and therefore this normality is expected to hold for their joint distribution. It should also be emphasized that other methods of voting and aggregation of MUP scores are possible including the Bayesian schema [2].

The GMM multi-class, GMM 2-Stage and ANN multi-class classifiers had better performance than the pattern discovery classifiers discussed in [8]. This could be attributed to the fact that all these methods search a larger hypothesis space and do not require quantization of feature values. The main advantage of the pattern discovery approach however, is transparency. This means that the end users are capable of interpreting the resulting categorization based on the MUP feature values.

The 2-stage approach is linear, has low dimensionality and a clinically relevant hierarchical decision process. However, the transparency of the 2-stage approach is still an open question.

It is not straight forward to conclude that these methodologies are superior to those described in [3], [6], [7]. A fair comparison would be based on EMG data acquired from the same or similar sets of muscles, preferably using the same acquisition protocol. Another reason that make such comparison difficult is that the decomposition and the feature estimations are achieved using different manual, semi-automatic and automatic techniques across the reported methods.

The proposed feature selection procedure is a variation of the backward selection wrapper method. The main difference is that it doesn't discard a feature unless there is a significant advantage. It revealed results that are consistent with the underlying physiological phenomena. For example, both neurogenic and myopathic disorders lead to instability across the ensemble of MUPs of a given MUPT. Therefore, no stability feature showed relevance in the first stage. Also,

it is worth noting that at least one feature of each category (size, complexity, stability and background noise) showed relevance.

VI. CONCLUSION

A 2-stage approach is proposed for muscle categorization based on decomposition based quantitative needle electromyography. In the first stage, a muscle is categorized to be either a neurogenic or myopathic candidate, while the actual disease involvement is confirmed or refuted in the second stage. The proposed algorithm showed an average accuracy of 91.25%. Its performance exceeded linear and non-linear multi-class classification schemas. The proposed approach has other advantages including linearity, lower dimensionality, the possibility of tuning each stage independently, the capability of reporting confidence measures and a hierarchical decision structure that matches current clinical practice.

REFERENCES

- [1] Fritz Buchthal and Poul Rosenfalck. Action potential parameters in different human muscles. *Acta Psychiatrica Scandinavica*, 30(1-2):125–131, 1955.
- [2] C. Farkas, A. Hamilton-Wright, H. Parsaei, and D. W. Stashuk. A review of clinical quantitative electromyography. *Crit Rev Biomed Eng*, 38:467–485, 2010.
- [3] C. D. Katsis, Y. Goletsis, A. Likas, D. I. Fotiadis, and I. Sarmas. A novel method for automated emg decomposition and muap classification. *Artificial Intelligence in Medicine*, 37(1):55 – 64, 2006.
- [4] R. Kendall and R. A. Werner. Interrater reliability of the needle examination in lumbosacral radiculopathy. *Muscle Nerve*, 34(2):238–241, Aug 2006.
- [5] E. KUGELBERG. Electromyography in muscular dystrophies; differentiation between dystrophies and chronic lower motor neurone lesions. *J. Neurol. Neurosurg. Psychiatr.*, 12:129–136, May 1949.
- [6] C. S. Pattichis, C. N. Schizas, and L. T. Middleton. Neural network models in emg diagnosis. *Biomedical Engineering, IEEE Transactions on*, 42(5):486–496, may 1995.
- [7] G. Pfeiffer. The diagnostic power of motor unit potential analysis: An objective bayesian approach. *Muscle and Nerve*, 22(5), 1999.
- [8] L.J. Pino and D. W. Stashuk. Using motor unit potential characterizations to estimate neuromuscular disorder level of involvement. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 4138–4141, aug. 2008.
- [9] Masahiro Sonoo. New attempts to quantify concentric needle electromyography. *Muscle Nerve*, 25(S11):S98–S102, 2002.
- [10] D. W. Stashuk. Emg signal decomposition: how can it be accomplished and used? *Journal of Electromyography and Kinesiology*, 11(3):151 – 173, 2001.