B. Bue, E. Merényi and J. Killian "Classification and Diagnosis of Myopathy from EMG Signals," 2nd Workshop on Data Mining for Medicine and Healthcare, in conjunction with the 13th SIAM International Conference on Data Mining (SDM-DMMH), Austin, TX, May 2013.

Classification and Diagnosis of Myopathy from EMG Signals*

Brian D. Bue† Erzsébet Merényi‡† James M. Killian§

Abstract

We present a methodology to predict the presence of myopathy (muscle disease) from intramuscular electromyography (EMG) signals. By evaluating the shape and frequency of electrical action potentials produced by muscular fibers and captured in EMG measurements, a physician can often detect both the presence and the severity of such disorders. However, EMG measurements can vary significantly across different subjects, different muscles, and according to session-specific characteristics such as muscle fatigue and degree of contraction. By considering fixed-duration (0.5-2 sec) frequency-domain samples of diagnostic regions in EMG signals measured at full muscle contraction, we can automatically detect the presence of myopathies across different subjects and muscles with ~90% accuracy. We argue that our methodology is more generally applicable than existing methods that depend upon accurate segmentation of individual motor unit action potential (MUAP) waveforms. We present a rigorous evaluation of our technique across several different subjects and muscles.

Keywords

EMG, myopathy, classification, diagnosis, FFT, frequency domain

1 Automated Diagnosis of Myopathy from EMG Signals

Myopathy (muscle disease) is a form of neuromuscular disorder that results in muscle weakness due to dysfunctioning skeletal muscle fibers [2]. A wide variety of both acquired and hereditary myopathies have been identified, many of which are serious and often debilitating conditions that are difficult to accurately diagnose and treat [3]. Early detection of these diseases by clinical examination and laboratory tests can greatly reduce patient suffering and medical costs. Moreover, data gathered during such examinations may lead to an improved understanding of the nature and treatment of such diseases, and allow development of automated systems that assist diagnosis.

In clinical practice, intramuscular EMG is a standard method used to assess neurophysiologic characteristics of skeletal muscles to diagnose neuromuscular diseases. EMG records electrical action potentials generated by groups of muscle fibers controlled by the same motor nerve, called a motor unit. These motor units are the basic functional units of the muscle that can be voluntarily activated. The shape of individual motor unit action potential waveforms reflect the status and structure of a given motor unit. EMG measurements from patients with myopathy differ from healthy subjects in that their recruited MUAPs usually have shorter duration, lower amplitude, and increased polyphasicity. Figure 1 illustrates the difference between 0.5 second samples of EMG traces from the deltoid muscle of a healthy subject vs. the deltoid of a subject with myopathy. These, and many additional subtleties characterize differences between healthy and abnormal subjects, depending on the nature and severity of pathology and are extensively discussed in the literature.

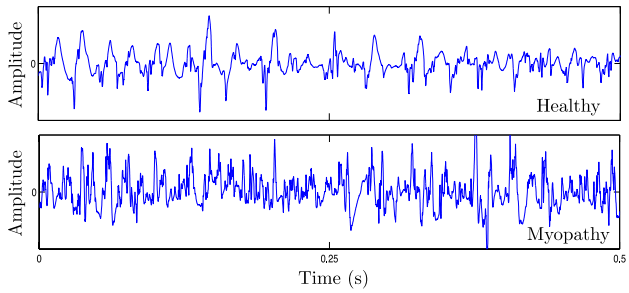

Figure 1: 0.5 second samples of EMG traces from the deltoid muscle from a healthy subject (top) vs. a subject with myopathy (bottom).

In recent years, a number of techniques have been proposed to classify EMG signals for medical diagnosis. Several authors (e.g., [8, 9, 4, 5, 7, 10]) propose segmenting the EMG data in the temporal domain into individual MUAP waveforms, which are then labeled and classified based upon (features derived from) the segmented waveforms. However, such techniques are limited in that they assume that individual MUAPs can be extracted from data in a consistent and reliable manner. Extracting individual MUAPs may be difficult or impossible since MUAPs at high muscle contraction are often in superposition, while pathologies of interest may not be observable at low muscle contraction. Most previous works analyze data obtained with low (less than full) muscle contraction. Moreover, the ratio of recruited MUAPs is another indicator of presence or absence of myopathy, and should also be taken into account, which is often not the case with previous work.

Given the issues with time-domain MUAP segmentation, classifying EMG data in the frequency domain may be a more robust approach. Some recent work has shown good results in classifying neuromuscular disease from EMG data in the frequency domain. For instance, [6] demonstrated 85% overall accuracy in classifying EMG signals of 59 subjects in the frequency domain into Normal, Myopathy and Neuropathy classes. In this work, we present a novel methodology for

* This work was partially supported by the Wheeler fund from the Baylor College of Medicine.

† Department of Electrical and Computer Engineering, Rice University, Houston, TX. (bbue@rice.edu)

‡ Department of Statistics, Rice University, Houston, TX. (erzsebet@rice.edu)

§ Department of Neurology, Baylor College of Medicine, Houston, TX. (jkillian@bcm.edu)

B. Bue, E. Merényi and J. Killian "Classification and Diagnosis of Myopathy from EMG Signals," 2nd Workshop on Data Mining for Medicine and Healthcare, in conjunction with the 13th SIAM International Conference on Data Mining (SDM-DMMH), Austin, TX, May 2013.

Muscle	# sec	# traces	Nor	Myo1	Myo2	Myo3	Myo4	Myo*
Biceps	51.00 (4)	4	0.00 (0)	7.50 (1)	0.00 (0)	35.00 (2)	8.50 (1)	51.00 (4)
Deltoid	53.00 (6)	6	26.00 (3)	8.50 (1)	0.00 (0)	18.50 (2)	0.00 (0)	53.00 (6)
Triceps	18.50 (2)	2	0.00 (0)	0.00 (0)	10.00 (1)	0.00 (0)	8.50 (1)	18.50 (2)
VL	51.50 (3)	3	0.00 (0)	0.00 (0)	21.50 (1)	12.00 (1)	18.00 (1)	51.50 (3)
Total	174.00 (8)	15	26.00 (3)	16.00 (2)	31.50 (2)	65.50 (5)	35.00 (3)	148.00 (6)
Percent	100.00	100.00	14.94	9.20	18.10	37.64	20.11	85.06

Table 1: Summary of EMG data for each muscle with sample duration *ns* = 0.5. The total number of seconds of data for each class is provided. Values in parenthesis give the number of unique subjects for each muscle with respect to each class.

analyzing EMG data in the frequency domain. Our methodology is distinct from previous research in the following:

a) We consider EMG data measured at full muscle contraction, which improves the objective evaluation of per-subject and per-muscle characteristics;

b) We classify diagnostic regions of the full EMG signal in the frequency domain, rather than pre-segmented individual MUAP waveforms; and

c) In addition to evaluating classification performance on data across different subjects, we also evaluate the characteristics of different muscles for diagnostic purposes.

In this work, we present a proof-of-concept of a novel methodology for classifying EMG signals of neuromuscular disease in the frequency domain. We are developing this methodology with the goal of classifying EMG signals according to disease severity. However, due to limited data, we demonstrate our methodology on the problem of classifying EMG signals into normal vs. myopathic categories in this work. We provide a rigorous evaluation of generalization capabilities across subjects and muscles via cross-validation, in contrast to a number of existing works [e.g., 8, 9, 5, 6].

2 EMG Data Description

Our EMG data was collected at the EMG Laboratory in the Department of Neurology of the Baylor College of Medicine in Houston, TX, by (or under the direction of) Dr. James Killian, M.D. The data we consider consists of 15 EMG sessions from 8 different subjects measured in one or more different muscles. Three of the subjects are female and the remaining subjects are male. The mean age of the subjects is 56.63 (std. dev=16.4) years. The currently available data are from the biceps brachii, triceps brachii, deltoid and vastus lateralis (VL), selected for their diagnostic utility by the physician. We use the term *trace* to denote a record of a "full" EMG session for a single subject on a single muscle. Each trace is collected using the following methodology: A monopolar needle electrode is inserted into a designated skeletal muscle in a proximal arm or leg. The signal is processed through the differential preamplifier to a Cadwell Sierra EMG machine amplifier (Cadwell Laboratories, WA, USA) which transfers the signal to a computer display and loudspeaker for clinical evaluation. The subject then exerts maximum contraction of the muscle under study as the electrode is moved by several millimeters until an adequate interferential muscle pattern of firing motor units is noted on the screen. A 60 sec sample is then recorded. The process is repeated on 4 to 6 separate muscles and the captured traces from each muscle are stored for subsequent signal analysis.

In a post-labeling session the physician designates each trace as a member of one of the following five classes based upon the observed severity of the pathology in the EMG signal: Healthy/Normal (Nor), Borderline Myopathy (Myo1), Mild Myopathy (Myo2), Moderate Myopathy (Myo3), Severe Myopathy (Myo4). The basis of the clinical diagnostic gradings of abnormal myopathic motor units (individual motor unit with durations of activity under 6ms) is related to the estimated percentage of myopathic units relative to the total number of firing motor units. Borderline: 0-10% abnormal units, mild: 10-25%, moderate: 25-50%, severe: above 50%. This is a subjective grading based on visual and auditory analysis by co-author JK of the different muscle samples. We are developing our methodology with all five classes in mind, and preliminary experiments show reasonable classification results for the five class problem. However, the currently available data is too scarce to represent the nuanced and fuzzy differences of the degrees of myopathy, especially across various muscles. We need significantly more data in order to present a five-class study with high confidence in the results. In this work, we focus on the methodology itself – classification of EMG diagnostic regions of 0.5 - 2.0 second durations (as opposed to individual MUAPs), in the frequency domain – and not the classification method. We demonstrate the effectiveness of our methodology using a baseline classifier (linear SVM) that yields good performance on the two classes (i.e., normal vs. myopathy) for which we have adequate data coverage. Classifier comparisons will be appropriate at a stage of this work where the main obstacle for better results is no longer the scarcity of data.

Portions of the traces are not diagnostic and/or saturated due to insertional activity or instrument tuning effects. To eliminate the non-diagnostic portions of each trace, the physician manually defines the *diagnostic regions* in each trace, which are temporally-contiguous segments of varying length. While automated identification and separation of non-diagnostic regions is important, it is outside the scope of the present work.

3 Methodology

3.1 Data Preprocessing

We split the diagnostic regions of each trace into fixed slices of *ns* seconds in duration. We subsequently refer to each of these slices as a *sample*. Each sample is a *m*-dimensional vector capturing a temporally-contiguous portion of each diagnostic region. We normalize each sample by its L^2 norm which maps the amplitudes of the samples to a common range. This allows us to reconcile, to some degree,

amplitude differences between measurements on different muscles and different subjects at varying contraction levels while retaining other differences of the waveforms. We then map each normalized sample into the frequency domain using the Fast Fourier Transform in MATLAB. We discard the symmetric portion of the frequency-domain samples, resulting in sample vectors of dimensionality $m/2$. Table 1 gives a summary of the samples we consider with sample duration ns = 0.5 sec.

3.2 Classification

In this study, we consider the problem of classifying the frequency-domain samples as Normal or Myopathic. To achieve this, we group all of the samples labeled Myo1-Myo4 into a single superclass Myo*. As table 5 shows, several muscles do not have any representative samples from some classes, and the remaining muscles are poorly represented in terms of the number of samples – particularly the Normal and the borderline myopathy (Myo1) classes, which represent only 14.94% and 9.2% of the total samples, respectively. To help mitigate this issue, we first balance the sampling distributions of the five (Normal, Myo1-Myo4) classes by augmenting the training data with $Nresamp_j = Nmax - N_j$ samples, sampled with replacement, from the training samples of each class j, where $Nmax$ is the number of samples of the class with the maximum number of samples, and N_j is the number of samples in class j. This balancing step ensures that samples of varying severity are equally represented, but leads to a sampling bias between the Normal vs. Myo* superclass. Consequently, we perform an additional balancing step by adding $Nnormal = Nall - NMyo*$ samples from the normal class to the training set, as before, sampling with replacement, where $Nall$ is the total number of samples, and $NMyo*$ is the number of samples in the Myo* class. After balancing, we have a total of 524 samples for the Normal and Myo* classes, with the Myo* class consisting of 131 samples of each of the Myo1-Myo4 classes, respectively.

EMG signals may vary between different subjects or on different muscles. Consequently, it is crucial to evaluate classification accuracy when data from different subjects and/or muscles is used as training and test data. To achieve this, after balancing the samples as described above, we perform ten cross-validation splits, where in each split we use data from half of the subjects for test data, and divide the remaining samples into training ($3/8^{th}$ of the total samples) and validation ($1/8^{th}$ of the total samples) sets. We ensure by random stratified sampling that the training, test and validation sets each contain instances from each of the Normal and Myo* classes and from each muscle group. The classifier we use is a linear Support Vector Machine (SVM). We select the SVM regularization parameter C from the set {0.01, 0.1, 1, 10, 100, 1000} that yields the highest accuracy on the validation set. We report the mean and standard deviation of classification accuracies produced on the test data in each split.

4 Classification Results and Evaluation

4.1 Classification Accuracy vs. Sample Duration

We first evaluate the classification accuracy with respect to the sample duration ns. We consider ns values in the set {0.05, 0.1, 0.2, 0.5, 1, 2}. Table 2 gives the number of balanced samples and the dimensionality m of each sample for each value of ns, and the corresponding mean and standard deviation of classification accuracies across the ten cross-validation splits. We observe that classification accuracy increases with increasing sample duration. The standard deviation also typically decreases, with the exception of ns=2, where the high dimensionality and small quantity of samples produce slightly less stable results. However, this generally suggests that longer sample durations are desirable, despite the high dimensionality of the resulting feature space. Additionally, our results indicate that it is possible to predict the presence or absence of myopathies from relatively short portions of a full EMG trace.

4.2 Per-class, Per-muscle and Per-subject Evaluation

We now evaluate the performance of our methodology on the individual classes, muscles and subjects we consider in this work. For this evaluation we fix the sample duration ns to 0.5, as this duration consists of a reasonable number of samples (1024) to evaluate, at fairly high dimensionality (16000 dimensions/sample) and yields very good classification accuracies (90.4% average).

With respect to the Normal vs. Myo* classes, we observe considerably higher classification accuracy on the Myo* class (mean=0.959, stddev=0.023) than on the normal class (mean=0.822, stddev=0.070). This is due to the fact that our data includes significantly fewer subjects with normal conditions. When we consider individual muscles (Table 3), we observe that the samples from the biceps and deltoid muscles tend to be misclassified more often than the triceps and VL muscles. A possible reason for this is that the biceps and deltoid muscles appear similar to one another in terms of EMG signals, but appear different from the triceps and VL muscles. This is also suggested by the results in Bischoff et al. [1], but further investigation on additional data is necessary to confirm this hypothesis in our case.

ns	#	$m/2$	Accuracy
0.0	10528	1600	0.760 (0.058)
0.1	5256	3200	0.815 (0.059)
0.2	2616	8000	0.878 (0.042)
0.5	1048	16000	0.904 (0.033)
1	512	32000	0.966 (0.028)
2	256	64000	0.971 (0.041)

Table 2: Number of balanced samples and sample dimensionality ($N/2$) with respect to sample duration ns, and corresponding mean and standard deviation of classification accuracies

Table 4 gives the classification accuracies for the individual subjects and their respective traces. Most notable are the results for subject S10, whose biceps and deltoid traces are classified with 28.5 and 16.1% less than their respective mean muscle accuracies (as shown in Table 3). Subject S10 represents a case where some muscles exhibit no observable pathology, while other muscles show signs of myopathy. While it is difficult to state conclusively without data from additional patients with similarly mixed pathologies, accord-

Bicep	Deltoid	Tricep	VL
0.907 (0.087)	0.852 (0.072)	1.000 (0.000)	1.000 (0.000)

Table 3: Per-muscle accuracies from all subjects for *ns*=0.5

ing to the physician, this case may be a result of a borderline myopathy, and the training labels may need revision once sufficient evidence is available.

Subject	Accuracy	Trace	Class	Trace Accuracy
S02	0.936	Biceps	Myo*	0.936 (0.050)
S03	0.958	Deltoid	Myo*	0.937 (0.055)
		Triceps	Myo*	1.000 (0.000)
S04	1.000	VL	Myo*	1.000 (0.000)
S07	0.986	Biceps	Myo*	0.972 (0.043)
		Deltoid	Myo*	0.984 (0.025)
		VL	Myo*	1.000 (0.000)
S08	0.888	Deltoid	Nor	0.888 (0.007)
S09	0.975	Biceps	Myo*	0.951 (0.068)
		Deltoid	Myo*	1.000 (0.000)
		Triceps	Myo*	1.000 (0.000)
S10	0.789	Biceps	Myo*	0.622 (0.171)
		Deltoid	Nor	0.691 (0.056)
		VL	Myo*	1.000 (0.000)
S15	0.852	Deltoid	Nor	0.852 (0.028)

Table 4: Per-subject/trace classification accuracies for *ns*=0.5.

5 Discussion and Future Work

In this work, we evaluated a novel methodology for classifying contiguous, fixed-duration samples of EMG signals in the frequency domain. By considering, as training samples, Fourier transforms of normalized, fixed-length segments of diagnostic regions of the full signals (as opposed to extracted MUAPs) measured at full contraction, we demonstrated high average generalization performance by a linear SVM classifier across individual subjects and different muscles. The average classification accuracy on test data increases from 80% to 97% with the duration of the samples (0.1 to 2 sec, respectively) while the reliability, determined from ten cross-validation folds, simultaneously increases (standard deviation decreases). Our analysis also suggests that detecting the presence of myopathy can be accomplished with very short duration samples of a full EMG trace.

The long-term, primary goal of our work is to develop a system that captures the physician's capability to diagnose a variety of neuromuscular disorders from EMG data, as well as to distinguish among the severity degrees of diseases such as the classes of myopathies listed in Section 2. While our classification accuracies are fairly high, this is of course a two-class case. Classifying the samples according to their severities is a more challenging task, and will require more elaborate and sophisticated experiments.

We also aim to classify EMG signals of patients with neurogenic disorders using our methodology. Because our methodology yields comparable results to previous analyses considering EMG data from myopathic and neurogenic diseases (e.g., 6), and based upon our preliminary experiments with 5 classes, we anticipate our method will generalize well to such scenarios.

While the results presented here are encouraging, much additional analysis and development is needed in order to achieve the above goals and to make our system useful for clinicians. This includes systematically designed experiments with increasing amounts and complexity of data (increased variety of subjects, muscles, diseases), testing increasingly sophisticated classification techniques to better align with real-life circumstances such as highly imbalanced sample sets, and intelligent identification of feature subsets necessary for producing high-quality (high-accuracy and high-fidelity) classifications. For fully automated processing, developing techniques to segment an EMG signal into diagnostic and non-diagnostic regions, or to incorporate learning constraints to identify various non-disease-related conditions are also necessary.

Acknowledgements The authors thank Penny Gregg at the EMG Laboratory of the Department of Neurology, Baylor College of Medicine, for her assistance with data collection, and Rice University graduate students Kai Du and Du Nguyen for data preprocessing and software modification efforts in the early stages of this work.

References

[1] C. Bischoff, E. Stälberg, B. Falck, and K.E. Eeg-Olofsson. Reference values of motor unit action potentials obtained with multi-MUAP analysis. Muscle & Nerve, vol. 17, no. 8, pp. 842–851, Aug. 1994.

[2] A. S. Blum and S. B. Rutkove. The clinical neurophysiology primer, vol. 388. Humana Press, 2007.

[3] F. Buchthal, An introduction to electromyography. Copenhagen: Gyldendal, 1957.

[4] C. I. Christodoulou and C. S. Pattichis. Unsupervised pattern recognition for the classification of EMG signals. IEEE Transactions on Biomedical Engineering, vol. 46, no. 2, pp. 169–178, Feb. 1999.

[5] N. F. Güler and S. Koçer, Classification of EMG Signals Using PCA and FFT, J Med Syst, vol. 29, no. 3, pp. 241–250, Jun. 2005.

[6] N. F. Güler and S. Koçer, Use of Support Vector Machines and Neural Network in Diagnosis of Neuromuscular Disorders, J. Med Syst, vol. 29, no. 3, pp. 271–284, Jun. 2005.

[7] R. Merletti and D. Farina. Analysis of intramuscular electromyogram signals. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 367 no. 1887, pp. 357–368, Jan. 2009.

[8] C. S. Pattichis, C. N. Schizas, and L. T. Middleton, Neural network models in EMG diagnosis. Biomedical Engineering, IEEE Transactions on, vol. 42, no. 5. May 1995.

[9] C. S. Pattichis and C. N. Schizas, "Genetics-based machine learning for the assessment of certain neuromuscular disorders.," Neural Networks, IEEE Transactions on, vol. 7, no. 2, pp. 427–439, Jan. 1996.

[10] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin. Techniques of EMG signal analysis: detection, processing, classification and applications. Biological Procedures Online, vol. 8, no. 1, pp. 11–35, Dec. 2006