Available online at www.sciencedirect.com
ScienceDirect
Medical Engineering & Physics 30 (2008) 563–573
Medical Engineering & Physics
www.elsevier.com/locate/medengphy

Motor unit potential characterization using "pattern discovery"

L.J. Pino [a,*], D.W. Stashuk [a], S.G. Boe [b], T.J. Doherty [b]

[a] *Systems Design Engineering, University of Waterloo, Canada*
[b] *Clinical Neurological Sciences, London Health Sciences Centre, Canada*

Received 21 September 2006; received in revised form 24 May 2007; accepted 16 June 2007

Abstract

Typically in clinical practice, electromyographers use qualitative auditory and visual analysis of electromyographic (EMG) signals to help infer if a neuromuscular disorder is present and if it is neuropathic or myopathic. Quantitative EMG methods exist that can more accurately measure feature values but require qualitative interpretation of a large number of statistics. Electrophysiological characterization of a neuromuscular system can be improved through the quantitative interpretation of EMG statistics. The aim of the present study was to compare the accuracy of pattern discovery (PD) characterization of motor unit potentials (MUPs) to other classifiers commonly used in the medical field. In addition, a demonstration of PD's transparency is provided. The transparency of PD characterization is a result of observing statistically significant events known as patterns. Using clinical MUP data from normal subjects and patients with known neuropathic disorders, PD achieved an error rate of 30.3% versus 29.8% for a Naïve Bayes classifier, 30.1% for a Decision Tree and 29% for discriminant analysis. Similar results were found for simulated EMG data. PD characterization succeeded in interpreting the information extracted from MUPs and transforming it into knowledge that is consistent with the literature and that can be valuable for the capture and transparent expression of clinically useful knowledge.

Keywords: Clinical decision support system; Electromyographic signal analysis; Motor unit potential; Neuromuscular disorder; Pattern discovery; Quantitative electromyography; QEMG

1. Introduction

Analysis of motor unit potentials (MUPs) helps clinicians infer if a neuromuscular disorder is present and if it is a result of pathology affecting motor neurons/motor axons (i.e. neuropathic) or related to muscle fiber injury or atrophy (i.e. myopathic). The status quo approach for an electromyographer involves using qualitative auditory and visual analysis of intramuscular needle-detected electromyographic (EMG) signals from voluntarily activated muscles. Qualitative analysis is often prone to error and limited by subjective misinterpretation in part because a large amount of potentially ambiguous information needs to be extracted and analyzed [1,2]. A recent study authored by Kendall et al. showed that faculty and residents (blind to the underlying diagnosis) using video recorded needle based examinations for radiculopathy had an overall 46.9% agreement with the actual diagnosis [3]. In addition, qualitative methods are less able than quantitative methods to provide effective longitudinal comparisons. An electrophysiological characterization that is sensitive to small changes caused by a neuromuscular disease would allow measurement of the degree of the disease involvement and therefore allow evaluation of treatment effectiveness [4].

For an electrophysiological characterization of a neuromuscular system to be sufficiently sensitive, EMG signals must be quantitatively analyzed. One class of quantitative electromyography (QEMG) methods attempts to characterize the individual motor units of a muscle by analyzing isolated MUPs. Usually, an individual motor unit is represented by a MUP template calculated from a train of MUPs generated by the motor unit. The characterization methods discussed in this work can use MUP templates based on any of these QEMG methods (e.g. level or window triggering, or more sophisticated decomposition techniques). Hence-

* Corresponding author. Tel.: +1 519 888 4567x33691; fax: +1 519 746 4791.
E-mail address: ljpino@uwaterloo.ca (L.J. Pino).

1350-4533/$ – see front matter © 2007 IPEM. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.medengphy.2007.06.005

forth, the term MUP will be used to refer to the calculated MUP template unless otherwise mentioned. As such, the focus of this study is on characterization of MUPs as normal, myopathic or neuropathic. Characterization is a clinical term referring to the discernment, description or attribution of distinguishing traits [5]. The degree of involvement of a disease process can be determined if each MUP in a set of MUPs sampled from a muscle is objectively characterized using a numeric value on a continuous scale that can be effectively combined into a neuromuscular characterization. More specifically, characterization of a single MUP using three numeric values reflecting the probability of it being detected from a normal, myopathic or neuropathic muscle is an initial step towards useful, robust neuromuscular characterization. This work describes and evaluates characterization processes that provide quantitative interpretation of information commonly extracted from individual MUPs during a QEMG examination.

The literature describes a number of different processes for characterizing MUPs. Stewart et al. developed and evaluated a system where using a pool of normal subjects the normative ranges were defined as the mean ± 2 standard deviations of the values of several different MUP features [6]. The error rate for characterizing mean MUP feature values as being detected from myopathic muscles was 44% and the neuropathic error rate was 36.7%.

Pattichis et al. [7] applied artificial neural network (ANN) models to the classification of MUPs sampled from normal, neuropathic and myopathic patients and achieved an error rate as low as 10%. The authors found that ANNs easily tended towards over-fitting (i.e., it was difficult to achieve generalization—the ability of the ANN to correctly classify unknown cases based on the training data.) In addition, they are one of the least transparent (basis of their decisions are easily understood) classifiers because of the often-large number of linear transformations applied to the feature values making ANNs essentially black box classifiers.

Pfeiffer & Kunze [8,9] applied linear discriminant analysis (LDA) to classify MUPs by calculating the probability that a particular MUP was detected from a normal, myopathic or neuropathic muscle. The LDA technique works best using continuous feature values and was found to be an accurate method for MUP characterization. However, because of arithmetic transformation of feature values, it is not as easy to understand the basis of decisions made by an LDA classifier as it is with classifiers that simply use logical relationships between features, i.e., for three classes LDA uses two arithmetic expressions to provide a relationship among the features.

Methods to characterize a neuromuscular system and to measure the degree of involvement of a disorder require a MUP characterization process that is accurate, allows the basis of it decisions to be easily understood, produces a numeric value in support of or refutation of a characterization, and is able to achieve generalization. The above techniques do not completely satisfy these important requirements. It is hypothesized, that a pattern discovery (PD) classifier can meet these requirements. Thus, the current study compares the accuracy of a PD classifier when used for the characterization of MUPs (detected in voluntarily active muscles using an intramuscular needle electrode) with other classifiers commonly used for medical decision support. The ability of PD classification to meet the above-mentioned requirements relative to other classifiers is also discussed. The remainder of the paper is structured as follows: a requirements section describing in more detail the criteria needed for effective MUP characterization; a brief description of potential classifiers and how they meet the requirements for MUP characterization; methodology, which formally describes pattern discovery based classification, provides rationale for choosing the other classifiers for comparison and describes how simulated data was generated and clinical data gathered; Section 5 reporting the performance of the different classifiers; and finally Sections 6 (Discussion) and 7 (Conclusions).

2. Requirements for MUP characterization

The requirements for MUP characterization are based on ideas developed by Kononenko [10] and Sprogar et al. [11] who describe a set of requirements needed for machine learning systems used in medical decision support.

2.1. Transparent

MUP characterizations need to be presented in a manner that allows a clinician to understand how a characterization was determined. This is especially important when faced with an unexpected characterization that contradicts a physician's initial expectation or intuition.

2.2. Accurate

MUP characterizations need to maximize both specificity and sensitivity beyond what is typically achieved through routine subjective analysis of an EMG signal.

2.3. Report confidence

A confidence measure reports the degree to which a characterization suggested by a classifier is likely to match the 'true' underlying characterization.

2.4. Numeric MUP characterization value

MUP characterizations need to be presented as numeric measures so that individual MUP characterizations can be combined into an overall neuromuscular characterization.

2.5. Mixed mode multi-variate features and interdependency

Discriminatory power can be increased when multiple feature values are considered simultaneously [12]. MUP characterization must be able to handle various mixed mode multi-variate data types that include numeric (continuous and or discrete), Boolean, nominal (categorical labels) and or ordinal (position or rank). The system needs to handle any underlying joint probability distribution of the features and capture their interdependencies. Clinical patterns that offer important clues to the type of a neuromuscular disorder often combine different data types.

2.6. Generalization

A MUP characterization system needs to accurately classify novel patterns that have not appeared in the training data [13].

2.7. Handle missing data

A MUP characterization system needs to be able to handle missing feature values in both the training and test data.

This study investigated the ability of a number of generic classifiers to provide transparency, accuracy, numeric characterization values and handle mixed-mode data types when used to classify electrophysiological data only. The other requirements will be examined in future work. The following section discusses some classifiers capable of handling mixed mode data and their suitability for MUP characterization.

3. Mixed mode data discrete classifiers

Naïve Bayesian (NB), Decision Tree (DT), and pattern discovery (PD) classifiers were designed to handle mixed mode data. They are discrete classifiers since they handle nominal and discrete data types and require continuous features to be quantized (segmenting the range of a feature's values into distinct intervals). NB and DT classifiers are widely used while PD based classifiers have been developed recently.

A NB classifier is built by assuming that all of the features are conditionally independent of each other and produces a joint probability estimate of each class that is the product of the conditional probabilities of each feature given the class. A NB classifier is rather simplified in that it doesn't reveal any interdependencies among the features and does not easily provide a transparent explanation [14].

A DT is a structure of tests done on the features of input data that leads to a classification. Since a single variable at a time is tested this does not readily reveal relationships among the features [15].

3.1. Pattern discovery

Information-theory based statistical inference can be used for the detection of significant patterns. Wong and Wang [16–19] describe the use of information-theory based PD for classification. Their classification algorithm is composed of three parts: discovery of patterns, rule selection, and classification. A significant pattern is said to occur when a set of discrete feature values occur together more often than expected assuming random occurrence. Patterns that include a class label can be used as rules for classification. An example of a rule would be a MUP (labeled myopathic) with low amplitude, low duration and a high number of turns. An information-theoretic measure called weight of evidence (WOE) represents the discriminatory power of a rule. PD classification is transparent because it does not apply transformations to the data and can express the WOE of the support or refutation of each set of rules that contributed to the classification.

4. Methods

This section formally defines pattern discovery based classification, provides a rationale for choosing the other classifiers for comparison, and describes how MUP data was obtained and represented and how classifier performance was measured.

4.1. Pattern discovery based classification

An observation with an unknown label can be classified using PD by discovering significant patterns in a set of training data. Given a training data set of N samples, a sample X is described by M features $X = \{x_1, \ldots, x_M\}$ where each feature x_i, $1 \leq i \leq M$, is a random variable. Each feature can take on a value from its discrete alphabet $\alpha = \{\alpha_i^1, \ldots, \alpha_i^{m_i}\}$ where m_i is the number of characters in the alphabet of the ith feature. A primary event occurs when a single feature x_i takes on a value from α_i. The pth primary event is denoted by $[x_i = \alpha_i^p]$ or simplified to x_{ip}. Let S be a subset of integers $(1, \ldots, R)$ containing $r = |S|$ elements where $1 \leq r \leq R$. X^s is an rth order event that contains r primary events. The number of observations of an rth order event X^s is denoted by o_{X^s}. The number of expected observations of an event is denoted by e_{X^s} and is calculated assuming the features in X^s are uniformly distributed and mutually independent. An event is defined as a pattern when it passes a test of statistical significance using the adjusted residual d_{X^s} [18].

$$d_{X^s} = \frac{(o_{X^s} - e_{X^s})/\sqrt{e_{X^s}}}{\sqrt{v_{X^s}}} \quad (1)$$

where v_{X^s} is the maximum likelihood estimate of the variance of the standardized residual $(o_{X^s} - e_{X^s})/\sqrt{e_{X^s}}$.

At the 95% confidence level, if $d_{X^s} > 1.96$, a positive pattern occurs and if $d_{X^s} < -1.96$, a negative pattern occurs. If

a pattern contains a class label then it is a rule and can be used for classification. A set of class labels $Y = \{y_1, \ldots, y_k, \ldots, y_K\}$ contains K labels and $1 \leq k \leq K$. The discriminatory power of an rth order rule X_l^r can be determined by the WOE, which is the odds of a sample X belonging to class y_k versus not belonging to y_k.

$$\mathrm{WOE} = \log_2 \frac{P(X_l^r | Y = y_k)}{P(X_l^r | Y \neq y_k)} \tag{2}$$

where $Y = y_k$ indicates the sample X is from class y_k, $Y \neq y_k$ indicates the sample X is not from class y_k, X_l^r is an rth order pattern, l is an index indicating the lth, rth order pattern.

Note that the range of (2) is $-\infty < \mathrm{WOE} < +\infty$. A rule based on a positive pattern labelled y_k will have a positive WOE and provides support for the classification y_k. A rule based on a negative pattern labelled y_k will have a negative WOE and provides refutation for the classification y_k. The strength of the support (or refutation) is proportional to +WOE (or −WOE). A pattern that is *always* associated with a particular label will produce a WOE of $+\infty$ and a pattern that is *never* associated with a particular label will produce a WOE of $-\infty$. Patterns associated with multiple labels will produce a WOE that is in between these two extremes.

An rth order pattern with a class label where $r < M$ is called a component rule. Examining component rules can enhance transparency by revealing the subsets of features and their strength (or lack thereof) of contribution towards the support or refutation of a classification. The union of disjoint component rules discovered in an observation is called a compound rule and is denoted by X_l^* where

$$X_l^* = \bigcup_{l=1}^{n} X_l^r \qquad \text{and} \qquad X_p^r \cap X_q^r = \emptyset, \quad p \neq q, \quad 1 \leq p, q \leq n \tag{3}$$

where n is the number of component rules in a given compound rule.

The weight of evidence for a sample X is the sum of the WOEs of each component rule that fires in support or refutation of class y_k. The PD classifier calculates K weights of evidence (one for each class) for each input sample. A sample X with an unknown class label is classified as y_k if the compound WOE supporting y_k is greater than the compound WOEs of the other classes.

4.2. Other classifiers considered

The classification performance of three other classifiers were compared with the PD classifier: two discrete classifiers – Weka J48 DT and Weka NB; and one continuous – LDA. Although not as transparent as the other classifiers, LDA was chosen for comparison because it was previously used for MUP classification [8,9]. DT and NB classifiers were chosen because they are commonly used for medical decision support, are regarded as being transparent [10,15], and can handle nominal data. The DT and NB classification error rates were determined using the algorithms of the Weka explorer system [20]. A Matlab function, RAFisher2CDA [21], was modified as per Pfeiffer's method [8] to calculate the Fisher discriminants used for LDA classification.

4.3. MUP data

Both simulated and clinical EMG data were decomposed to calculate MUP templates using decomposition-based QEMG (DQEMG [22]). DQEMG typically finds 51 isolated MUPs produced by a single motor unit, aligns them, and uses a median trimmed average to form the MUP template.

4.3.1. Clinical MUP data

A disposable concentric needle electrode (Model N53153; Teca Corp., Hawthorne, NY) was used to acquire intramuscular signals using DQEMG [22] on a Neuroscan Comperio (Neuroscan Medical Systems, El Paso,Texas) with a bandpass of 10 Hz–10 kHz at a sampling rate of 31.2 kHz as previously described [23–25]. These intramuscular signals were acquired during 30 s voluntary isometric contractions performed at approximately 10% of each individual subjects' maximal voluntary contraction (MVC). Subjects were aided in maintaining a consistent force throughout the 30-s contraction by the presence of a target line on an oscilloscope displaying the force produced by the contraction.

Control and neuropathic MUPs were sampled from the biceps-brachii and first dorsal interosseous muscles. In total, 1649 MUPs were sampled from 16 healthy control subjects (aged 27 ± 4 years) and 427 MUPs were sampled from 14 patients, including 9 patients (aged 52 ± 12 years) with clinically probable or definite amyotrophic lateral sclerosis as defined by the revised El Escorial criteria [26] and 5 patients (aged 37 ± 11 years) with Charcot–Marie–Tooth disease type X confirmed via genetic testing.

4.3.2. Simulated data

EMG signals were simulated using a physiologically based model [27] to help examine the relationship between level of involvement and MUP characterization performance. The simulator was extended to also allow simulation of the affects of neuropathic and myopathic disorders [28]. To simulate a neuropathy, motor units are reorganized progressing from random motor neuron death to random re-innervation of orphaned fibers by nearby surviving motor neurons. To simulate inflammatory myopathy, a small percentage of randomly selected healthy fibers are "infected" and atrophied by a small fraction, and a smaller proportion are hypertrophied by a small fraction. This process is iterated by infecting additional fibers, atrophying and hypertrophying the newly and previously infected fibers until the prescribed level of involvement is reached. A fiber is considered non-functioning (i.e. dead) when its diameter is below a critical threshold.

The simulator models the recruitment of motor units necessary to bring the level of force produced by a diseased muscle up to a prescribed percentage of the MVC force of a healthy muscle. Therefore, a simulated muscle that is normal, or with 25% or 75% motor unit loss at 10% MVC are each modeled as if producing close to the same force. EMG signals detected using a concentric needle at various intramuscular positions of several different muscles during approximately 7–10% MVC were simulated and then decomposed into MUP templates. This method mimics the completion of several EMG studies across different individuals and includes background MUP interference and noise typical in clinical studies. A comparison done by Hamilton-Wright and Stashuk [27] of the quantitative analysis of simulated versus real healthy decomposed MUPs shows good correspondence.

In total, 500 MUPs were extracted from simulated EMG signals of normal muscle, 500 from myopathic muscle and 500 from neuropathic muscle. The myopathic/neuropathic MUPs, in approximately equal proportion, were simulated to come from muscles with 25%, 50% and 75% muscle-fiber/motor-unit loss. Each MUP was labeled as either normal, myopathic or neuropathic allowing comparison with the literature that uses these labels [6–9].

4.4. Feature extraction

Values for the following MUP features, measured from a MUP template, were fed into the four classifier algorithms: amplitude, duration, phases, turns and area-to-amplitude ratio (AAR) otherwise known as thickness. Except for LDA, which requires continuous data, simulated MUP data was quantized into intervals derived using simulated data. Similarly, quantization intervals for clinical data were derived using clinical data. The discrete classifiers (PD, DT and NB) were all presented with the same quantized training and testing data sets. This prevented NB and DT from using their own quantization methods so that the effects of quantization on classification accuracy were not a factor during comparison. Using only three quantization intervals labeled low, medium, and high helps to simplify the visual patterns that explain the results leading to diagrams that are easily recognized and understood by clinicians. The natural logarithm of amplitude was taken before determining the linear discriminant functions to minimize its skewness. Prior to quantization, none of the features were transformed for PD, DT, NB classification.

4.5. Classifier performance

For the clinical MUP data, sensitivity was defined as the total number of MUPs characterized as neuropathic divided by the total number of 'true' neuropathic MUPs. Specificity was defined as total number of MUPs characterized as normal divided by the total number of 'true' normal MUPs. Accuracy was defined as one minus the error rate and was equal to the mean of specificity and sensitivity. The term sensitivity-specificity deviation (SSD) defined as

$$\mathrm{SSD} = \sqrt{\frac{(A - \mathrm{Sens})^2 + (A - \mathrm{Spec})^2}{2}} \tag{4}$$

where A is the accuracy, Sens the sensitivity, and Spec is the specificity, was used to determine how well a characterization method maximized both specificity and sensitivity.

Table 1
Sensitivity/specificity/accuracy of characterization of clinical data

	Sensitivity (%)	Specificity (%)	Accuracy (%)	SSD (%)
PD	71.7	67.8	69.7	1.9
LDA	65.8	76.2	71.0	5.2
J48	66.3	73.5	69.9	3.6
NB	75.7	64.8	70.2	5.4

The results in the Table are the mean sensitivity, specificity, and accuracy across thirty trials for the clinical MUP data. The SSD is based on the means shown in the Table. The training data for each trial consisted of three hundred randomly chosen samples from each class for a total of 600 samples. After the training data was chosen, the remaining data was used to test the classifiers to establish classification performance. Each trial used the same training and test data across the different classification methods.

5. Results

5.1. Clinical MUP data

PD had an average error rate (and standard deviation) of 30.3% (1.8%), LDA 29.0% (2.1%), J48 DT 30.1% (2.1%) and NB 29.8% (1.8%) across thirty trials. All four characterization methods had similar error rates with no statistically significant differences between each other according to the Tukey post hoc test at a significance level of 0.05. Table 1 shows NB had the highest sensitivity and LDA had the highest specificity and accuracy. However, PD had a significantly lower SSD compared to the other methods. J48 DT had the next best SSD, which was about two times the SSD of the PD method.

5.2. Simulated data

PD had an average error rate (and standard deviation) of 24.3% (1.2%), LDA 23.1% (1.4%), DT 24.6% (1.2%) and NB 27.0% (1.4%) across thirty trials. A Tukey post hoc test showed that LDA had a significant improvement at the 0.05 level compared to the other classifiers. PD and DT did not have statistically significant differences between each other according to the Tukey post hoc test at a significance level of 0.05. The Tukey post hoc test between the other classifiers and NB showed that it had a significant decrease in performance at the 0.05 level.

Tables 2 and 3 show the confusion matrix for PD and LDA simulated-data classifications respectively. The confusion matrices shown were determined as an average of the

Table 2
Pattern discovery confusion matrix (simulated data)

True class	Classified as			per class error rate (%)
	Myopathic	Normal	Neuropathic	
Myopathic	170	20	10	15
Normal	20	144	36	28
Neuropathic	21	39	140	30

The PD confusion matrix shows that myopathic simulated MUPs had the lowest per-class classification error rate of 15%. The neuropathic MUPs had the highest per-class error rate at 30%. This confusion matrix shows the expected per-class distribution of errors. The probability of error that a MUP detected from a myopathic muscle is misclassified as normal is 2 times greater than being misclassified as neuropathic (20/200 versus 10/200). A similar trend appears for neuropathic data where the probability of error of a MUP being misclassified as normal is about 2 times greater than being misclassified as myopathic (39/200 versus 21/200).

Table 3
LDA confusion matrix (simulated data)

True class	Classified as			per class error rate (%)
	Myopathic	Normal	Neuropathic	
Myopathic	176	20	4	12
Normal	17	138	45	31
Neuropathic	15	38	147	27

The LDA confusion matrix shows that there was an increase in the error rate for normal classifications from 28% to 31% and a decrease in error rate from 30% to 27% for neuropathic classifications compared to the PD classifier. The confusion matrix for LDA shows the expected distribution for incorrectly classified MUPs detected from muscles with disorders is 5 times greater for normal (for 'true' myopathic) and about 2.5 times greater for normal (for 'true' neuropathic).

confusion matrices across the thirty trials. The Tables show that the per class error rates for simulated normal and neuropathic are similar with a dramatic improvement for simulated myopathic data.

If only normal and neuropathic simulated-data error rates are compared, Table 2 shows an average error rate across two classes of 29% for the PD classifier and Table 3 shows a two class error rate of 28.8% for the LDA classifier, which are very similar to the error rates of the clinical MUP characterizations of 30.3% and 29.0% respectively. Note that the clinical data did not include any myopathic data.

Fig. 1 shows that as the level of simulated involvement increased, the classification error rate decreased for MUPs detected from simulated myopathic and neuropathic muscles. Neuropathic MUPs had error rates of 48%, 30%, and 21% for levels of involvement of 25%, 50%, and 75% respectively. Myopathic MUPs had error rates of 22%, 13%, and 7% for levels of involvement of 25%, 50% and 75% respectively.

6. Discussion

Table 4 summarizes a comparison of the studied classifiers for four of the identified requirements based on the study results and consideration of the known properties of each

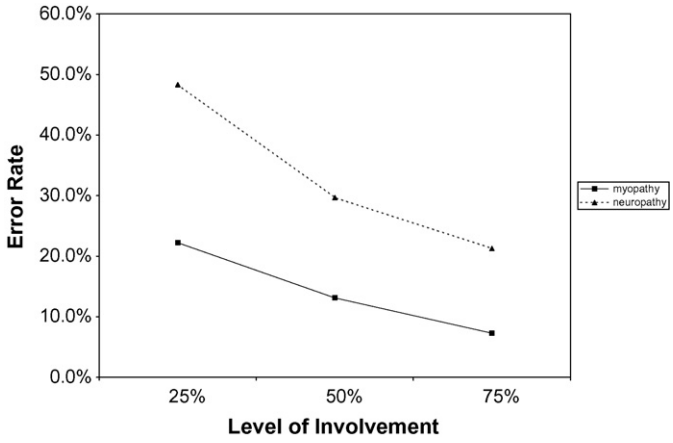

Fig. 1. Error rate vs. level of involvement for simulated MUPs. Myopathic level of involvement refers to percentage of muscles fibers that are no longer functioning. Neuropathic level of involvement refers to percentage of non-functioning motor neurons.

classifier. PD has an advantage with respect to transparency. All four methods have similar accuracy. PD, DT and NB can handle mixed mode data types while LDA cannot because of its inability to handle nominal data. All four classifiers can produce a numeric value for characterization. In addition, while the computational effort required for training for each method can be significant, it is done offline prior to clinical use and only needs to be done once. The computational effort for characterization expended by each method is similar and not consequentially different from current QEMG examinations.

Figs. 2–4 are example PD characterizations of three different clinical MUPs and Fig. 5 is an example characterization of a simulated MUP. The reporting of the component rules that support each characterization demonstrates the transparency of PD characterization. Furthermore, the rules discovered by PD are consistent with currently used diagnostic criteria [29–32] (see Table 5). These MUP characterizations demonstrate that the PD classifier is capable of capturing knowledge consistent with current practice and expressing it in transparent, understandable terms. The knowledge is captured while simultaneously considering multiple feature values and expressed using common clinical terminology.

Table 4
Summary: classifiers fit to requirements

Requirement	PD	LDA	DT	NB
Transparency	****	**	***	***
Accuracy	***	***	***	***
Mixed-mode data	****	**	****	****
Numeric characterization	****	****	****	****

Number of stars: Four—excellent, three—good, two—fair, one—poor. Accuracy ratings are based on the error rate discussed in Section 5. Transparency is based on the classifier's ability to report strength of support/refutation of a subset of features. Ability to handle mixed mode data and produce a numeric characterization scores are based on examining the methodology of the classifiers.

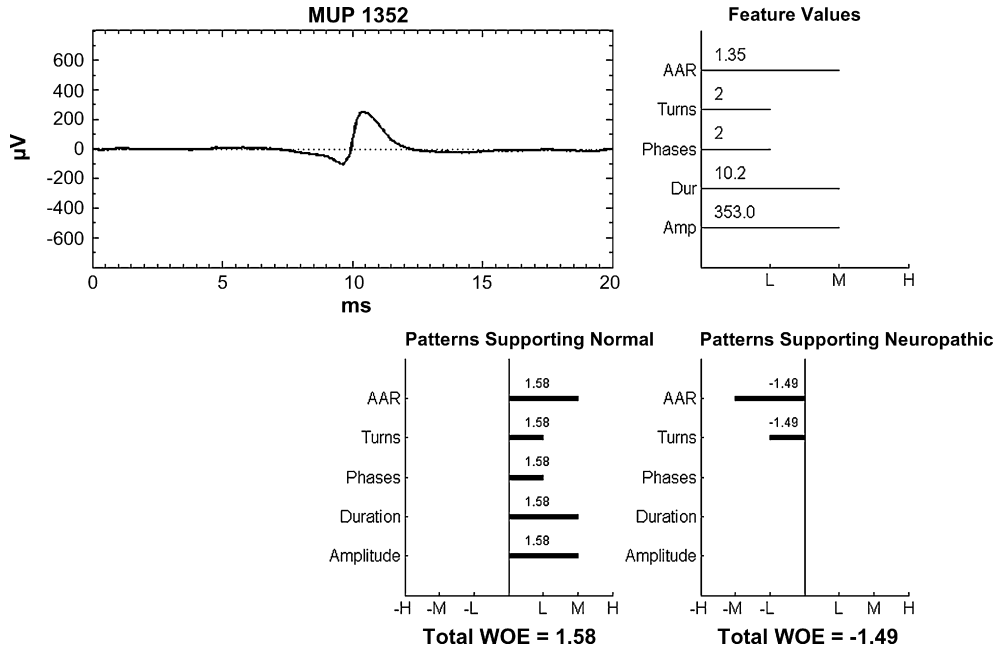

Fig. 2. Evidence supporting characterization of clinical MUP 1352 as normal. The upper lefts of Figs. 2–5 show MUP templates each of which is the median-trimmed average of 51 isolated MUP firings deemed by DQEMG to belong to the same MUP train. Each of the MUP templates is displayed using the same scale. The *y* scale is in microvolts and the *x* scale is milliseconds. The upper right box of each figure shows the result of quantization of the five MUP features into the low, medium or high quantization interval along with their values prior to quantization. For instance in this figure MUP 1352 is quantized as follows: AAR of 1.35 is medium, 2 turns are low, 2 phases are low, duration of 10.2 ms is medium and amplitude of 353 uV is medium. The next two bar plots underneath show the component rules providing evidence for or against the MUP being detected from a normal or neuropathic muscle respectively. Features belonging to the same component rule are indicated by using the same line style (e.g. solid, dashed, dotted, etc.). A component rule is positive if the bars go to the right from the vertical line and a component rule is negative if the bars are left of the vertical. The *x*-axis for a negative rule is also marked with a negative sign for low, medium and high to emphasize that these are rules refuting the characterization.
The bar graph marked as 'Patterns Supporting Normal' shows that all of the features formed a single component rule since they are all solid leading to a WOE of 1.58 that MUP 1352 was detected from a normal muscle. The bar graph marked as 'Patterns Supporting Neuropathic' shows a single component rule for neuropathy. The component rule indicated by the solid line refuting neuropathic is composed of medium AAR, and low turns and has a WOE of −1.49. Since the WOE for normal has the highest value—this MUP was characterized as being detected from a normal muscle.

Transparency is an important part of a clinician's acceptance of a characterization and is critical to their confidence in its veracity [33]. According to Feng [34] transparency to users requires that a characterization system provide logical versus arithmetic expressions of the features. Operators such as "and", "or" and "if-then" are preferred to provide the connections between feature values used to explain characterizations. These expressions are valuable since they provide meaningful explanations and are easier to evaluate. PD characterization provides a transparent explanation of the set of features weighted by their contribution, using both negative and positive rules that refute or support characterization towards the various classes. The system is able to objectively calculate the WOE that a MUP was detected from a normal or neuropathic muscle. This reduces the mental workload of the clinician, which should reduce the number of errors in both individual MUP characterizations as well as overall neuromuscular characterizations. This leads to the expectation that examining clinicians will make fewer errors.

The error rate of the LDA and PD method based on normal and neuropathic clinical data is about 25% lower than the error rate achieved in Pfeiffer's work [8] and about 15% lower than Stewart's work [6]. Pfeiffer's data [8], collected two years apart, was not done consistently. It is not expected that PD will achieve better accuracy than LDA. If PD were tested using the data set in [8] performance similar to LDA would be expected. PD's advantage however is transparency. Stewart's method is used for combining a set of MUPs (representative of different motor units) by obtaining means for each feature [6]. Comparing the mean value of each feature of a set of MUPs to normative ranges does not provide high sensitivity because the significance of 'outlier' MUPs, which may indicate a disorder, is often diminished by the averaging. The poor sensitivity of Stewart's method is supported by Podnar's work, which achieved a sensitivity of 47% [35] (i.e. an error rate of 53%).

DT, NB and PD based classifiers are suitable for use in MUP characterization; however, PD has an advantage over DT and NB based on its transparency. The low error rate of the LDA method helps to confirm Pfeiffer's conclusion that LDA classification is an accurate method [8]. However, LDA is not very transparent and

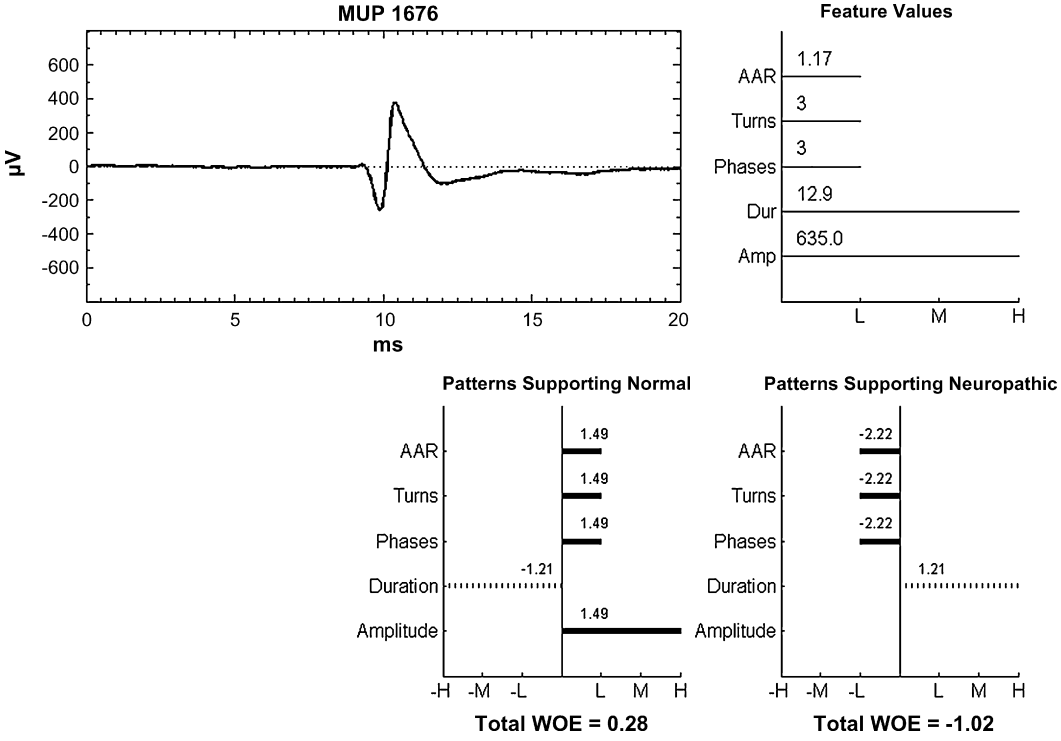

Fig. 3. Evidence supporting characterization of clinical MUP 1676 as Normal. The bar graph in the figure titled as 'Patterns Supporting Normal' shows two component rules making up a compound rule for normal—one rule as a solid line, the other as a dashed line. The component rule indicated by the solid line supporting normal is composed of low AAR, low turns, low phases and high amplitude. The other component rule indicated by a dashed line is composed of high duration and refutes normal with a WOE of −1.21. In the 'Patterns Supporting Neuropathic' bar graph there are two component rules. The component rule indicated by the solid line with low AAR, low turns, low phases has a negative WOE of −2.22. The other component rule indicated by a dashed line is composed of high duration and has a WOE of 1.21. In total, the WOE for being detected from a normal muscle is less than one at 0.28, and from a neuropathic muscle −1.02. The spread between the highest WOE and the second highest WOE for MUP 1676 is 1.3. This indicates a slightly lower level of confidence towards a classification for normal of MUP 1676 than MUP 1352 whose WOE spread was 3.07.

cannot easily handle nominal data types—an important requirement in extending a clinical decision support system to handle other clinical observations. DT classifiers use an error reduction based training algorithm and therefore provide rules that are used to minimize classification error. This can lead to large trees (even with pruning) and over-fitting problems. NB and DT methods, as mentioned previously, cannot find relationships among multiple features. PD rules are determined through observation of statistically significant relationships in the training data without considering classification accuracy and capture the dependencies among features using hypothesis testing. Although accuracy across the four methods did not differ significantly for the clinical data, PD had a significant advantage in its ability to maximize both sensitivity and specificity. A clinic with unknown "costs" for false negative and false positive characterizations would favour a characterization method that maximized both sensitivity and specificity.

The characterization methods had similar error rates for clinical data, and simulated normal and neuropathic MUPs. This suggests that the simulator can be useful for studying the relationships between level of involvement and characterization performance for normal and neuropathic MUPs. However, the lower error rates achieved by the LDA and PD methods for simulated myopathic versus neuropathic data may be an artifact of the simulator. The simulator does not deal with changing muscle density or structure. A separate study using clinical data from both inflammatory and non-inflammatory myopathic disorders is needed.

Although the results shown by Fig. 1 are for simulated data, they are consistent with the expectation that as a disease affects a greater portion of a muscle the probability of detecting MUPs that reflect the effects of the disorder increases thus reducing the number of errors made when categorizing MUPs detected from muscles with higher levels of involvement of a disorder. On the other hand, the reduced probability of detecting a MUP that reflects the affects of a disorder on a muscle during the early stages of involvement suggest that further development of QEMG methods, to combine MUP characterizations and/or use other QEMG based features, is needed to improve both the specificity and sensitivity with which a neuromuscular system can be characterized. In addition, further evaluation with appropriate clinical data is required.

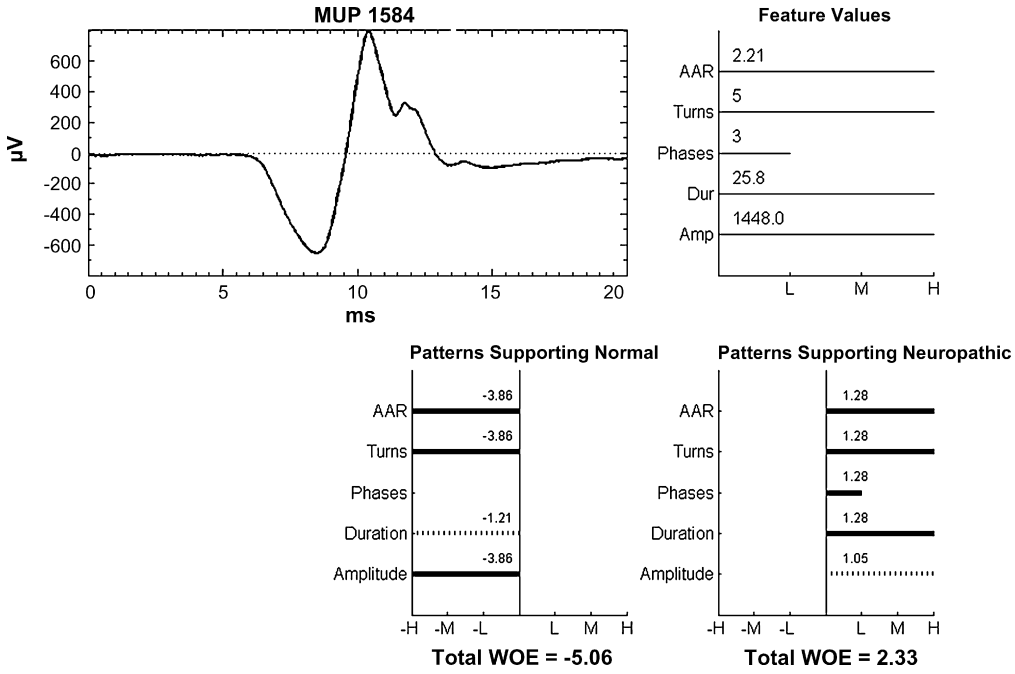

Fig. 4. Evidence supporting neuropathic characterization of clinical MUP 1584. The bar graph in the figure titled as 'Patterns Supporting Neuropathic' shows a compound rule comprised of two component rules that provide evidence that MUP 1584 was detected from a neuropathic muscle. A component rule (dashed line) of high amplitude lends a WOE of 1.05 and a component rule (solid line) of high AAR, high turns, low phases, and high duration provides a WOE of 1.28. In total, the two component rules provide WOE of 2.33 that this MUP was detected from a neuropathic muscle.

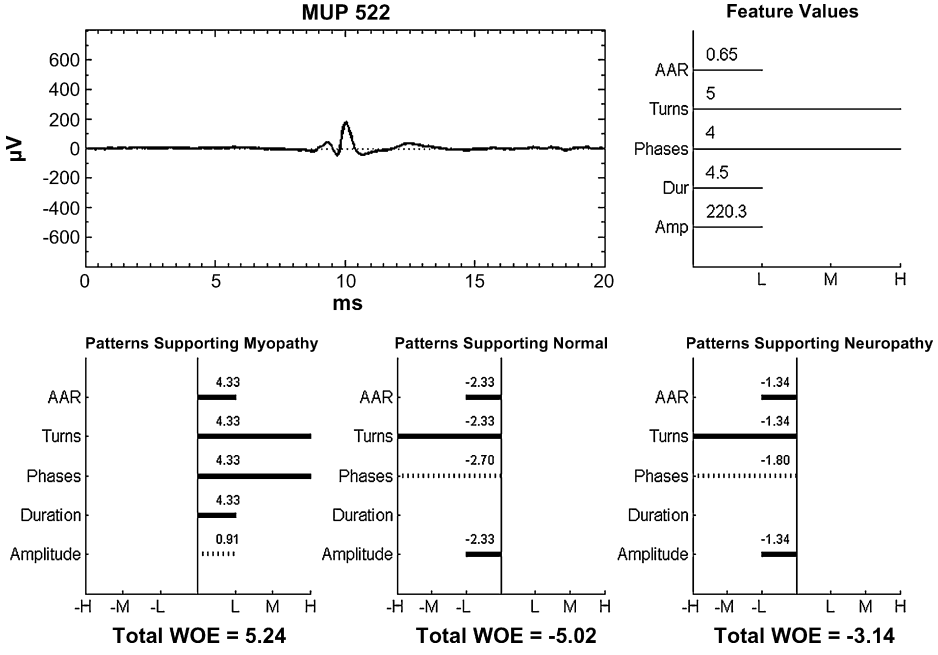

Fig. 5. Evidence supporting myopathic characterization of simulated MUP 522. The bar graph in the figure titled as 'Patterns Supporting Myopathic' shows that the WOE of the first component rule indicated by a solid line of low AAR, high turns, high phases and low duration plus the WOE from the second component rule (dashed line) of low amplitude provides a total WOE of 5.24 that MUP 522 was detected from a myopathic muscle. The compound rules supporting normal and neuropathic both have a large spread (10.26 and 8.16 respectively) below the WOE value for myopathic; an indication of high confidence that this MUP was detected from a myopathic muscle.

Table 5
Current clinical criteria

No.	Rule description	Indication of	Source
1	Decreased duration, decreased amplitude, polyphasic	Myopathy	[32]
2	Decreased AAR	Myopathy	[31]
3	Increased duration, increased amplitude, polyphasic	Neuropathy	[32]
4	Increased amplitude (and/or area) with normal or increased AAR	Neuropathy	[31]
5	Increased amplitude, high number of turns, may not be polyphasic	Neuropathy	[29]

Clinical MUP 1352 (Fig. 2) with medium duration and amplitude, low number of turns, low phases and medium AAR was characterized by the PD classifier as being normal. Clinical MUP 1584 (Fig. 4) with high duration and amplitude, low phases and high turns and high AAR was characterized by the PD classifier as being neuropathic agreeing with rules 4 and 5. Simulated MUP 522 (Fig. 5) with low duration and amplitude, high numbers of turns and phases and low AAR was characterized by PD as being myopathic agreeing with rules 1 and 2.

7. Conclusions

Unlike the other classifiers investigated, the PD classifier is able to explain itself by reporting the sets of feature values, along with a strength-of-evidence measure, supporting or refuting its characterizations. This work has demonstrated that the PD classifier meets the requirements for normal and neuropathic MUP characterization through its abilities to report its characterizations in a transparent manner, handle mixed mode data, discover dependencies among features, provide numerical characterization values, and achieve a similar level of characterization accuracy as state of the art classification methods. The PD classifier shows promise as a clinically useful method of providing numerical inputs to the next step of the interpretation phase of a QEMG examination (neuromuscular characterization).

The PD classifier succeeds in interpreting the information extracted from quantitative MUP analysis and transforming it into knowledge that is consistent with current clinical practice. This demonstrates that the transparency provided by the PD classifier can be valuable for the capture and expression of knowledge useful to a clinician.

Characterization of clinical and simulated, normal and neuropathic, MUP data had very similar error rates. These results provide encouragement to develop and evaluate a PD method for quantifying the level of involvement of a neuromuscular disorder—ultimately fulfilling one of the roles for future QEMG examinations envisioned by Swash [4].

Future work will focus on extending PD by examining methods of combining sets of MUPs from the same muscle to determine a muscle characterization. Also, future work will examine confidence measures of PD for accuracy and correlation to level of involvement. The goal is to develop and evaluate a system that meets the majority of identified requirements based on testing of clinical data from patients with confirmed diagnoses. If this proves to be satisfactory then a trial on how the system affects physician performance compared to the status quo would be recommended.

Conflict of interest

None.

References

[1] Tversky A, Kahnemann D. Judgement under uncertainty. Science 1974;185:1124–31.
[2] Wickens CD, Hollands JG. Engineering psychology and human performance. 3rd ed. NJ 07458: Prentice-Hall Inc.; 2000.
[3] Kendall R, Werner RA. Interrater reliability of the needle examination in lumbosacral radiculopathy. Muscle Nerve 2006;34(2):238–41.
[4] Swash M. What does the neurologist expect from clinical neurophysiology? Muscle Nerve 2002;999(S11):S134.
[5] Stedman TL. Stedman's medical dictionary. 27th ed. Philadelphia: Lippincott Williams & Wilkins; 2000.
[6] Stewart CR, Nandedkar SD, Massey JM, Gilchrist JM, Barkhaus PE, Sanders DB. Evaluation of an automatic method of measuring features of motor unit action potentials. Muscle Nerve 1989;12(2):141–8.
[7] Pattichis CS, Schizas CN, Middleton LT. Neural network models in EMG diagnosis. IEEE Trans Biomed Eng 1995;42(5):486–96.
[8] Pfeiffer G. The diagnostic power of motor unit potential analysis: an objective Bayesian approach. Muscle Nerve 1999;22:584–91.
[9] Pfeiffer G, Kunze K. Discriminant classification of motor unit potentials (MUPs) successfully separates neurogenic and myogenic conditions, a comparison of multi- and univariate diagnostic algorithms for MUP analysis. Electroencephalogr Clin Neurophysiol 1995;97:191–207.
[10] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med 2001;23(1):89–109.
[11] Sprogar M, Lenic M, Alayon S. Evolution in medical decision making. J Med Syst 2002;26(5):479–89.
[12] Chan KM. Needle EMG abnormalities in neurogenic and muscle diseases. In: Brown WF, Bolton CF, Aminoff MJ, editors. Neuromuscular function and disease. Philadelphia, PA: Elsevier Science; 2002. p. 359–68.
[13] Duda R, Hart PE. Pattern classification. 2nd ed. John Wiley and Sons, Inc.; 2001.
[14] Szafron D, Greiner R, Lu P. TCXplain: transparent explanation of naïve bayes classifications. Report TR03-09. University of Alberta, 2003.
[15] Bull C, Chiogna M, Franklin R, Spiegelhalter D. Expert derived and automatically generated classification trees: an example from pediatric cardiology. In: Proceedings of the 1993 conference on computers in cardiology, 5–8 September. 1993. p. 217–20.
[16] Wong AKC, Wang Y. Pattern discovery: a data driven approach to decision support. IEEE Trans Syst Man Cybern Part C: Appl Rev 2003;33(1):114–24.
[17] Wong AKC, Wang Y. High-order pattern discovery from discrete-valued data. IEEE Trans Knowled Data Eng 1997;9(6):877–93.
[18] Wong AKC, Wang Y. Discovery of high order patterns. In: Proceedings of the 1995 IEEE international conference on systems, man and cybernetics, 2. 1995. p. 1142–7.
[19] Wang Y. High-order pattern discovery and analysis of discrete-valued data sets. Ph.D. Thesis, University of Waterloo, 1997.
[20] Witten IH, Eibe F. Data mining: practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufman; 2005.
[21] Trujillo-Ortiz A, Hernandez-Walls R, Perez-Osuna S. RAFisher2cda: canonical discriminant analysis. A MATLAB file 2004 [WWW document].
[22] Stashuk DW. Decomposition and quantitative analysis of clinical electromyographic signals. Med Eng Phys 1999;21(6–7):389–404.

[23] Boe SG, Stashuk DW, Brown WF, Doherty TJ. Decomposition-based quantitative electromyography: effect of force on motor unit potentials and motor unit number estimates. Muscle Nerve 2005;31(3):365–73.
[24] Boe SG, Stashuk DW, Doherty TJ. Motor unit number estimation by decomposition-enhanced spike-triggered averaging: control data, test–retest reliability, and contractile level effects. Muscle Nerve 2004;29(5):693–9.
[25] Boe SG, Stashuk DW, Doherty TJ. Within-subject reliability of motor unit number estimates and quantitative motor unit analysis in a distal and proximal upper limb muscle. Clin Neurophysiol 2006;117(3):596–603.
[26] Brooks BR, Miller RG, Swash M, Munsat TL. El escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord 2000;1(5):293–9.
[27] Hamilton-Wright A, Stashuk DW. Physiologically based simulation of clinical EMG signals. IEEE Trans Biomed Eng 2005;52(2):171–83.
[28] Hamilton-Wright A, Stashuk DW, Doherty TJ. Simulation of disease effects on muscle structure, activation and acquired electromyography. Muscle Nerve 2003;28:S126.
[29] Stalberg E, Nandedkar SD, Sanders DB, Falck B. Quantitative motor unit potential analysis. J Clin Neurophysiol 1996;13(5):401–22.
[30] Stalberg E, Erdem H. Quantitative motor unit potential analysis in routine. Electromyogr Clin Neurophysiol 2002;42(7):433–42.
[31] Nandedkar SD, Barkhaus PE, Sanders DB, Stålberg EV. Analysis of amplitude and area of concentric needle EMG motor unit action potentials. Electroencephalogr Clin Neurophysiol 1988;69(6):561–7.
[32] Buchthal F. An introduction to electromyography. Copenhagen: Scandinavian University Books; 1957.
[33] Strat TM, Lowrance JD. Explaining evidential analyses. Int J Approx Reason 1989;3(4):299–353.
[34] Feng C, Michie D. Machine learning of rules and trees. In: Michie D, Spiegelhalter DJ, Taylor CC, editors. Machine learning, neural and statistical classification. Herfordshire, UK: Ellis Horwood Limited; 1994.
[35] Podnar S. Criteria for neuropathic abnormality in quantitative anal sphincter electromyography. Muscle Nerve 2004;30(5):596.