Medical Decision Support System for Diagnosis of Neuromuscular Disorders using DWT and Fuzzy Support Vector Machines

Abdulhamit Subasi

International Burch University, Faculty of Engineering and Information Technologies, 71000, Sarajevo, Bosnia and Herzegovina.

E-mail: asubasi@ibu.edu.ba

Abstract

The motor unit action potentials (MUAPs) in an electromyographic (EMG) signal provide a significant source of information for the assessment of neuromuscular disorders. In this work, different types of machine learning methods were used to classify EMG signals and compared in relation to their accuracy in classification of EMG signals. The models automatically classify the EMG signals into normal, neurogenic or myopathic. The best averaged performance over 10 runs of randomized cross-validation is also obtained by different classification models. Some conclusions concerning the impacts of features on the EMG signal classification were obtained through analysis of the classification techniques. The comparative analysis suggests that the fuzzy SVM modelling is superior to the other machine learning methods in at least three points: slightly higher recognition rate; insensitivity to overtraining; and consistent outputs demonstrating higher reliability. The combined model with DWT and Fuzzy-SVM achieves the better performance for internal cross validation (External cross validation) with the area under the ROC curve (AUC) and accuracy equal to 0.996 (0.993) and 97.41% (97%), respectively. These results show that the proposed model have the potential to obtain a reliable classification of EMG signals, and to assist the clinicians for making a correct diagnosis of neuromuscular disorders.

Keywords: Electromyography (EMG); Motor unit action potentials (MUAPs); Discrete Wavelet Transform (DWT); Artificial Neural Network (ANN); Radial Basis Function Networks (RBFN); C4.5 Decision Tree; Linear Discriminant Analysis (LDA); Support Vector Machine (SVM); Fuzzy SVM.

1. Introduction

The Electromyography (EMG) is generated by the electrical activity of the muscle fibers and it can be detected using intramuscular electrodes. Neuromuscular diseases are a group of disorders whose most important characteristic is that they cause muscular weakness and/or muscle tissue wasting. These disorders have an effect on the motor nuclei of the cranial nerves, the anterior horn cells of the spinal cord, the nerve
roots and spinal nerves, the peripheral nerves, the neuromuscular junction, and the muscle itself. In neurogenic disorders, some motor neurons degenerate, the surviving motor neurons grow new axonal sprouts establishing synaptic contacts with the denervated muscle fibers. Needle electromyography is the most suitable way for the detection of changes in motor unit size and its internal structure. Hence needle EMG differentiate between other types of normal and abnormal spontaneous activity. From the point of view of signal analysis, the electrical activity recorded by needle EMG may be pseudorandom and the frequency content of the recorded signal depends on quite a lot of aspects. Since the electrical properties of intramuscular tissues affect volume conduction, the higher frequencies within the signal will be attenuated considerably more than the lower ones. The frequency content of the EMG signal has a physiological significance, and frequency analysis has been used in different studies of neuromuscular disorders. In needle EMG, a shift toward high frequencies is a characteristic feature in myopathies, while a shift toward low frequencies is often well-known in neurogenic conditions [1].

Hence, the resulting information can be used to find out the origin of the diseases, i.e. neurogenic or myopathic [2-4]. It is typical clinical practice to examine the MUAPs from visual inspection and from listening to their audio characteristics. Nevertheless, subjective MUAP assessment, even if satisfactory for the detection of obvious abnormalities, may not be sufficient to describe less apparent deviations or mixed patterns of abnormalities [5]. Consequently, for an effective automated EMG signal classification, a systematic treatment of EMG signals must be carried out. For this reason, a number of computer-based quantitative EMG analysis algorithms have been developed. Pattichis and Pattichis [6] have investigated the usefulness of the wavelet transform (WT), which provides a linear time-scale representation for describing motor unit action potential (MUAP) morphology and three different neural networks, the backpropagation (BP), the radial-basis function network (RBF), and the self-organizing feature map (SOFM). Pattichis and Elia [7] used autoregressive and cepstral analyses combined with time domain analysis in classification of EMG signals. Pattichis et al. [8] used MUAP parameters as input to a sequential parametric pattern recognition classifier. Subasi et al. [9] investigated the practicality of using an autoregressive model and wavelet neural network to extract classifiable features from EMG. Katsis et al. used Support vector machines (SVM) [10], RBFN and DecisionTree (DT) [11] for MUAP classification. Pino et al. [12] used Naive Bayesian (NB), DT, and pattern discovery (PD) classifiers for MUAP classification and characterization. Yan et al. [13] used mutual information-based feature selection in surface EMG and applied fuzzy LS-SVM in motion classification. Kamavuako et al. [14] tried to determine the relationship between grasping force and features of single-channel intramuscular EMG signals. Recently, Dobrowolski et al. [15] used MUAPs decomposition using wavelet and SVM for the classification of neuromuscular disorders.
Nevertheless, in some of the studies done before, the EMG signals were acquired from surface EMG and the classification is realized on the basis of different feature extraction methods. Actually, there exist so many differences between intramuscular EMG (iEMG) signals and surface EMG (sEMG) signals that it cannot be guaranteed that the effective algorithms used in sEMG scenario also work well in iEMG scenario. Besides, the feature extraction methods are also very important in the classification of EMG signals. In this paper, FSVM classifier combined with statistical features extracted from DWT are compared different machine learning methods to classify EMG signals. To contribute to the quantification of the routine needle EMG examination, a methodology has been developed for EMG signal classification which consists of three steps. In the first step, the EMG signals are decomposed into different frequency bands using discrete wavelet transform (DWT). In the second step, statistical features extracted from these subband decomposed EMG signals to get better accuracy for diagnosis of neuromuscular disorder. In the last step, an unknown EMG signal is classified as normal, myopathic or neurogenic using different machine learning methods. Also we compared the performances of different machine learning techniques such as LDA, ANN, RBFN, C4.5 decision tree, SVM and FSVM.

The remainder of the paper is organized as follows. In the next section, information is given about the subjects and the methods applied in each step of the EMG signal classification process are presented. Section 3 provides a complete experimental study of the different machine learning methods for diagnosis of neuromuscular disorders, in which the impact of feature set and algorithmic issues are compared with respect to classification performance. Section 4 gives some discussion on the results. Finally, the conclusions are summarized in Section 5.

2. Materials and Methods

2.1. Subjects and Data Acquisition

An EMG system (Keypoint; Medtronic Functional Diagnostics, Skovlunde, Denmark) with standard settings was used. The EMG signal was acquired from the biceps brachii muscle using a concentric needle electrode (0.45 mm diameter with a recording surface area 0.07 mm$^2$; impedance at 20 Hz below 200 kΩ). The signal was band-pass filtered at 5 Hz to 10 kHz and sampled at 20 kHz for 5 s with 12-bit resolution. All the measurements from patients and control group were done in Neurology Department of University of Gaziantep. The electrode is usually advanced at least 3-5 mm into the muscle before recording. The electrode is also moved at least 3-5 mm between recordings to make sure that different MUAP’s are recorded. The electrode is advanced until the medial or posterior border of the muscle is reached. The electrode is then pulled out to the fascial and inserted to a new radial direction.
A. Subasi, Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines, Computers in Biology and Medicine 42, 806–815, 2012.

The EMG signals were recorded from the biceps brachii muscle at force levels approximately 30% of maximum voluntary contraction (MVC) under isometric conditions. Patient diagnoses were based on a range of clinical information included a general examination and clinical history of the patient, and EMG and nerve conduction tests. Muscle biopsies were not taken in the majority of cases, on ethical grounds, as they are only considered in EMG clinic in cases where diagnosis is uncertain or for specific clinical reasons. In this study, EMG data collected from 27 subjects have been analysed. Data were recorded from seven healthy subjects (three males, four females) with ages ranging from 10 to 43 years (mean age±standard deviation (S.D.): 30.2±10.8 years), seven myopathic subjects (four males, three females) with ages ranging from 7 to 46 years (mean age±standard deviation (S.D.): 21.5±13.3 years) and thirteen neurogenic subjects (eight males, five females) with ages ranging from 7 to 55 years (mean age±standard deviation (S.D.): 25.1±17.2 years) as in Subasi et al. [9].

2.2. Feature extraction using discrete wavelet transform

The wavelet transform (WT) allows the discrimination of non-stationary signals with different frequency features [16]. The wavelet transform decomposes a signal into a set of basic functions called wavelets. The wavelets are obtained from a single function \( \psi \), the mother wavelet, by dilations and translations as \([17,18]\).

\[
\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi \left( \frac{t-b}{a} \right)
\]

(1)

for positive \( a \). Conventionally, \( a = 1 \) for the mother wavelet and increasing \( a > 1 \) dilates the wavelet, expanding the interval over which it takes non-zero values. This model leads to the definition of the equation for the continuous wavelet transform (CWT):

\[
W(a,b) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{a}} \psi \left( \frac{t-b}{a} \right) dt,
\]

(2)

where \( b \) operate to translate the function across \( x(t) \) just the same as \( t \) does in the equations above, and the variable \( a \) acts to adjust the time scale of the probing function, \( \psi \).

The discrete wavelet transform (DWT) is used to decompose a signal by using filters to extract interesting frequency resolution components within the signal. The DWT possesses compact support in both time and frequency domain [19-22]. The DWT is a signal-processing tool that finds many engineering and scientific applications. DWT analyzes the signal at different frequency bands, with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions called scaling functions and wavelet functions, which are related to low-pass and high-pass filters,
respectively. Each stage of this scheme consists of two digital filters and two down-samplers by 2. The down-sampled outputs of first high-pass and low-pass filters provide the detail, D1 and the approximation, A1, respectively. The first approximation, A1 is further decomposed and this process is continued as in Fig. 1. These approximation and detail records are reconstructed from the Daubechies 4 (DB4) wavelet filter. Details are given in [19-26].

The extracted wavelet coefficients provide a compact representation that shows the energy distribution of the EMG signal in time and frequency. In order to decrease the dimensionality of the extracted feature vectors, statistics over the set of the wavelet coefficients were used [27]. The following statistical features were utilized to represent the time-frequency distribution of the EMG signals:

1. Mean of the absolute values of the coefficients in each sub-band.
2. Average power of the wavelet coefficients in each sub-band.
3. Standard deviation of the coefficients in each sub-band.
4. Ratio of the absolute mean values of adjacent sub-bands.

Features 1 and 2 represent the frequency distribution of the signal and the features 3 and 4 the amount of changes in frequency distribution. Total 15 feature are extracted, which are 4 different values for features (1), (2) and (3); 3 values for feature (4). These feature vectors, calculated for the frequency bands A5 and D3–D5 and were used as an input to classifiers.

2.3. Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA) is a technique that produces discriminant functions that are linear in the input variables. It is also used in a particular sense to refer to the technique in which a transformation is required, which maximises between class separability and minimizes within class variability.

There are different ways of generalising the criterion to the multiclass case. Optimisation of these criteria yields transformations that reduce to Fisher’s linear discriminant in the two-class case and that maximise the between-class scatter and minimise the within-class scatter [28]. Standard LDA classifier of RapidMiner [29] machine learning tool is used in our study.

2.4. Artificial Neural Networks (ANN)

Artificial neural networks (ANNs) are appropriate for solving problems in biomedical engineering and, particularly, in analyzing biomedical signals, because of their variety of applicability and their capability to learn complex and nonlinear relations. ANNs are trained by example instead of rules. When used in diagnosis of neuromuscular disorders, they are not affected by factors such as human fatigue,
emotional states, and habituation. They are capable of rapid identification, analysis of conditions, and diagnosis in real time [30]. The most frequently used training algorithm in classification problems is the backpropagation (BP) algorithm which is used in this work also. There are many different types and architectures of neural networks varying fundamentally in the way they learn; the details of which are well documented in the literature [31, 32]. MLP classifier of RapidMiner [29] machine learning tool was used with a three-layer backpropagation neural network with 9 tanSig neurons in the hidden layer and pureLin neuron in the output layer is used. The network training function is the traingdx. Besides, the learning rate and momentum rate is set to 0.1 and 0.2. The accepted average squared error is 0.005 and the training epochs are 1000. These parameters are obtained by trial and error.

2.5. Radial Basis Function Networks (RBFN)

Since RBFN has a simpler structure and a much faster training process, it is a popular alternative to the MLP. The RBFN has network architecture similar to the classical regularization network [33]. Each node in the hidden layer of RBFN uses an RBF as its nonlinear activation function. The hidden layer carries out a nonlinear transform of the input, and hence the output layer is a linear combiner mapping the nonlinearity into a new space. The biases of the output layer neurons can be modeled by an extra neuron in the hidden layer. The learning of the RBFN necessitates the determination of the RBF centers and the weights. The centers can be placed on a random subset of the training examples, or determined by clustering or by means of a learning procedure. For classification problems, RBF units map input patterns from a non-linearly separable space to a linearly separable space, and the responses of the RBF units form new feature vectors. Each RBF prototype is a cluster serving mainly a certain class. The RBFN is insensitive to the order of the appearance of the adjusted signals, and consequently more appropriate for online or succeeding adaptive adjustment [34]. Standard RBF classifier of RapidMiner [29] machine learning tool was used in the experiments.

2.6. C4.5 Decision Tree (DT)

Any decision tree will increasingly divide the set of training examples into smaller and smaller subsets. If all the samples in each subset had the same category label, we would say that each subset was pure, and could terminate that part of the tree. However, there is a mixture of labels in each subset, and thus for each branch we must make a decision either to stop dividing and accept an imperfect decision, or instead choose a different property and grow the tree further [35].

The C4.5 algorithm has the prerequisite for pruning based on the rules derived from the learned tree. Each leaf node has an associated rule which is the conjunction of the decisions leading from the root node,
through the tree, to that leaf. A technique called C4.5 Rules deletes redundant antecedents in such rules. The information corresponding to nodes near the root can be pruned by C4.5 Rules. This is more general than impurity based pruning methods, which instead merge leaf nodes [35, 36]. Standard J48 classifier of RapidMiner [29] machine learning tool was used in the experiments.

2.7. Support vector machines (SVM)

In this section, some basic work on Support vector machines (SVM) is shortly reviewed, for further details refer to [36-38]. SVM are one of the recently developed classifiers in statistical learning theory [39]. SVM present a solution to the binary classification problem. The algorithm separates the two classes by a hyperplane that maximizes the distance between the hyperplane and the nearest sample of each class. They are focus of interest in biomedical applications, because of their accuracy. Most of the prior classifiers separate classes using hyperplanes. But SVM widen the idea of hyperplane separation to data that cannot be separated linearly, by mapping the predictors onto a new, higher-dimensional space in which they can be separated linearly. It is achievable to separate the classes in any training set while running the risk of finding trivial solutions that over-fit the data. As a result, the SVM algorithm includes methods that intend to avoid this over-fitting. This is accomplished by using an iterative optimisation algorithm. Generally, these misclassifications only arise if the wrong kernel function has been selected or there are cases in different classes. The method’s name derives from the support vectors, which are lists of the predictor values taken from cases that lie close to the decision boundary separating the classes. It is practical to assume that these cases have the greatest impact on the location of the decision boundary. Computationally, finding the best location for the decision plane is an optimisation problem to facilitate a kernel function to create linear boundaries through non-linear transformations [37, 40, 41]. Since RBF kernel function gives better accuracy, it was selected as a kernel function for the SVM classifier. For training the SVM classifier with the RBF kernel function, the kernel parameter $\gamma$ and the trade-off parameter $C$ that could be chosen based on training data. First the original algorithm of SVM was used to find the optimal $\gamma$ value and the trade-off parameter $C$. These parameters were sought in the two-dimension grid by $C = (1, 2, 10, 50, 100, 1000)$ and a set of 100 equidistant values of $\gamma$ between $10^{-2}$ and $10^2$. A low number of support vectors were used as the criterion for choosing $C$ and $\gamma$. For too low values of $C$, a large number of support vectors were obtained. Such a solution was less economical and required more computations for evaluating the decision function. The chosen example values of $C$ between 50 and 500 should provide good and economical solutions. The minimal number of support vectors obtained was 42 corresponding to $C = 300$, $\gamma = 0.2$. Therefore, the RBF kernel parameter $\gamma$ of 0.2 and the trade-off parameter $C$ of 300 were chosen for the RBF kernel function in the SVM classifier.
2.8. Fuzzy Support Vector Machines (FSVM)

SVM is a powerful machine learning method for solving classification problems [37], but there are still some restrictions of this theory. From the training set and formulations, each training sample belongs to either one class or the other. For each class, we can simply verify that all training samples of this class are treated uniformly in the theory of SVM. In many real-world applications, the effects of the training samples are different. Also some training samples are more imperative than others in the classification problem. It would be required that the significant training samples must be classified correctly and would not be cared about some training samples like noise whether they are misclassified or not. This means that, each training sample no more exactly belongs to one of the two classes. It may possibly 90% belong to one class and 10% be meaningless, and it may possibly 20% belong to one class and 80% be meaningless. In other words, there is a fuzzy membership $0 < s_i \leq 1$ associated with each training sample $x_i$. This fuzzy membership can be regarded as the attitude of the related training sample toward one class in the classification problem and the value $(1 - s_i)$ can be regarded as the attitude of meaningless. Lin and Wang [42] extend the concept of SVM with fuzzy membership and make it an FSVM. Even though in the formulation of the problem the fuzzy membership is assumed to be given in advance, it is helpful to have the parameters of membership being automatically setting up in the course of training [42, 43].

The major difference among SVM and FSVM is that the cost $C$ of FSVM is multiplied by fuzzy membership $s_i$. In FSVM model, the theory behind is to set a fuzzy membership to every input point and to reformulate SVM in such a way that different input points can make different contributions to the learning of the decision surface. Besides, FSVM is based on the maximization of the margin similar to the classical SVM. On the other hand, it uses fuzzy membership function instead of fixed weights to prevent noisy data points from getting narrower margins. The procedure becomes more complicated for the multi-class classification problems since the outputs could be more than one class and must be divided into $N$ mutually exclusive classes. Actually, there are some ways to solve multi-class classification problems for SVM such as one-against-one (OAO) classifiers and one-against-all (OAA) classifiers [44].

2.8.1. Generating the Fuzzy Memberships

In order to choose the proper fuzzy memberships in a given problem is straightforward. First, the lower bound of fuzzy memberships must be defined, and then, the main property of data set needs to be selected and made connection between this property and fuzzy memberships [42, 43].

Both in binary class and in multi-class classification problem, it is not easy to define the target value of classification and the fuzzy membership function of training data during the implementation of the OAA
classifier in which a key issue is to the success of the classification system. Generally, for a k-class classification problem, the OAA contains k decision functions when the OAA classifier is operated and the jth decision function is trained to separate class j (labeled +1) examples from the rest examples (labeled -1).

To solve the multi-class EMG signal classification problem, OAA-FSVM classifier can be defined as, an EMG signal belongs to class j if the element of column j is 1, while another signal does not belong to class j if the element of column j is -1. We can assign higher weight to each positive example and lower weight to each negative example. Consequently, the membership function \( ss_{ij} \) for each of the n EMG signals in k classes’ problem can be defined as

\[
ss_{ij} = \begin{cases} 
ss_{ij}^+ & \text{if } v_{ij} = 1 \\
ss_{ij}^- & \text{if } v_{ij} = -1
\end{cases}
\]

where

\[0 \leq ss_{ij}^+ \leq 1, \quad 0 \leq ss_{ij}^- \leq 1, \quad i=1,\ldots,n, \quad j=1,\ldots,k\] (4)

For the experimental study, the symbol OAA-FSVM (\( ss_{ij}^+ \), \( ss_{ij}^- \)) can be defined as the fuzzy membership function of OAA-FSVM classifier. The target values of proposed EMG signal classification system will be transformed into multi-dimensional bipolar values. Besides, every training data point has its different weights at each set of classification and the setting of fuzzy membership function is decided by its target value [44].

For training the FSVM classifiers with the RBF kernel functions, the kernel parameter \( \gamma \) and the trade-off parameter \( C \) could be determined based on training data. First the parameters were sought in the two-dimension grid by \( \gamma = (1, 2, 10, 50, 100, 1000) \) and a set of 100 equidistant values of \( \gamma \) between \( 10^{-3} \) and \( 10^3 \) as it is done in SVM. A low number of support vectors were used as the criterion for choosing \( C \) and \( \gamma \). For too low values of \( C \), a large number of support vectors were taken. The selected example values of \( C \) between 50 and 500 give effective solutions. The minimal number of support vectors obtained was 32 corresponding to \( C = 250, \gamma = 0.2 \). Therefore, the RBF kernel parameter \( \gamma \) of 0.2 and the trade-off parameter \( C \) of 250 were chosen for the RBF kernel function in the FSVM classifiers. The value of the tuning membership parameter \( r \) for the FSVM classifier using the RBF kernel function \( (C = 250, \gamma = 0.2) \) was searched in the set of \( r = (1, 1/2, 1/4, 1/8, 1/16) \) based on training data [44]. The value of \( r \) associated with the minimal number of support vectors was selected. The minimal number of support vectors obtained was 32 corresponding to \( r=1 \).

3. Experimental Results
To contribute to the quantification of the routine needle EMG examination, a methodology has been developed for EMG signal classification which consists of three steps. In the first step, the EMG signals are decomposed into different frequency bands using discrete wavelet transform (DWT) as shown in Figure 2. In the second step, statistical features extracted from these subband decomposed EMG signals to get better accuracy for diagnosis of neuromuscular disorder. In the last step, an unknown EMG signal is classified as normal $[0 \ 0 \ 1]$, myopathic $[0 \ 1 \ 0]$ or neurogenic $[1 \ 0 \ 0]$ by different classification methods namely Linear discriminant analysis (LDA), radial basis function networks (RBFN), multilayer perceptron artificial neural networks (MLPANN), C4.5 decision tree (C4.5 DT), Support Vector Machines (SVM) and Fuzzy SVM (Fig.3.). Out of these classifiers, FSVM combined with statistical features extracted from DWT gave the best performance result for EMG signal classification. Besides, this technique tries to extract the most valuable characteristic input features by minimizing redundancy and exclude noise from the EMG signal.

The problem of predicting performance based on limited data is an interesting, and controversial. Different techniques encountered, of which one—cross-validation—is gaining ascendance and is possibly the assessment technique of choice in most practical limited-data situations. Comparing the performance of different machine learning techniques on a particular problem is another matter that is not as easy as it sounds. In most practical classification techniques the cost of a misclassification error depends on the nature of error it is—whether, for instance, a positive example was erroneously classified as negative or vice versa. While doing classification, and evaluating its performance, it is often crucial to take these costs into account. For classification problems, it is natural to measure a classifier’s performance in terms of the error rate [45].

In order to calculate the performance of a classifier on independent dataset called the test set, its error rate on a dataset that played no part in the creation of the classifier need to be assessed. In some circumstances, three different datasets can be mentioned: the training data, the validation data, and the test data. The training data is used to train the classifier. The validation data is used to adjust parameters of the classifier. Then the test data is used to calculate the error rate of the final method. All of the three sets need to be chosen individually: the validation set must be different from the training set to achieve better performance in the selection stage, and the test set must be different from both to get a reliable estimate of the true error rate. If lots of data exists, a large sample can be used for training; then another, independent large sample of different data can be used for testing. As long as both samples are representative, the error rate on the test set will give a true indication of future performance. And the larger the test sample, the more precise the error estimate. The accuracy of the error estimate can be computed statistically [45].

If the amount of data for training and testing is limited, it is common to hold out one-third of the data for testing and use the remaining two-thirds for training. In this case, if the sample used for training (or
testing) is not representative, it should be ensured that the random sampling is done in such a way as to guarantee that each class is appropriately represented in both training and test sets. This technique is called stratification, and it provides only a primitive precaution against uneven representation in training and test sets. A more general approach to mitigate any bias caused by the individual sample chosen for holdout is to repeat the whole process, training and testing, several times with different random samples. In each step, a certain amount of the data is randomly selected for training, possibly with stratification, and the remainder used for testing. The error rates on the different iterations are averaged to yield an overall error rate. This is the repeated holdout method of error rate estimation. In a single holdout process, one might consider swapping the roles of the testing and training data—that is, train the system on the test data and test it on the training data—and average the two results, consequently reducing the effect of uneven representation in training and test sets. On the other hand, a simple alternative forms the basis of an important statistical technique called cross-validation. In cross-validation, the data is split into k approximately equal partitions and each in turn is used for testing and the remainder is used for training. This is called k-fold cross-validation, and if stratification is adopted as well and it is called stratified k-fold cross-validation [45].

Since the usual approach for calculating the error rate of a classification technique given a single, fixed sample of data was to use stratified 10-fold cross-validation, the EMG data has been divided randomly into 10 parts in which the class has been represented in approximately the same amounts as in the full dataset. Tests have also shown that the use of stratification improves results slightly. Therefore the typical assessment technique in situations where only limited data is available is stratified 10-fold cross-validation. Note that neither the stratification nor the division into 10 folds has to be exact, because statistical evaluation is not an exact science. Moreover, stratification reduces the variation, but it certainly does not eliminate it entirely [45].

In order to calculate the performance of the each model, the whole EMG data is divided into training and test sets, and 10-fold cross-validation is used subsequently. The cross-validation accuracy (CVA) is the average of the k individual accuracy measures

\[
CVA = \frac{1}{k} \sum_{j=1}^{k} A_i
\]

(5)

where k (10 in this case) is the number of folds used, and \( A_i \) is the accuracy measure of each fold, \( i = 1, \ldots, k \) [46]. While stratifying the data by three classes, each of the 10 folds contains approximately the same proportions of NOR, MYO and NEU cases as those in the whole data set. K-fold cross validation is used to avoid bias possibly introduced by selection of a specific training and test set. But there are variations of the K-fold cross validation (CV). The most popular approaches include the K-fold CV, in which more than one subject is left out in each loop. This approach provides a slightly more biased error
estimate, but with reduced variability, mainly due to more subjects being left-out [47]. If the results obtained are used only to report prediction error estimates, the CV is called External (ECV), and the second sample is a true test set. If the error obtained from the second sample is used to choose the best of the derived models, then the CV is called Internal (ICV), since it is part of model development, and the second sample is a second training set. Therefore, ICV can lead to erroneous conclusions if one refers to its results as error estimates. On the other hand, ECV gives an unbiased error estimate of a model [47, 48]. The results for ECV and ICV are shown in table 7 and 8.

Receiver operating characteristic (ROC) curve is a graphical technique for evaluating classification performance and represent the performance of a classifier without regard to class distribution or error costs. They plot the number of positives enclosed in the sample on the vertical axis, expressed as a percentage of the total number of positives, versus the number of negatives involved in the sample, defined as a percentage of the total number of negatives, on the horizontal axis. This is one approach of using cross-validation to generate ROC curves. A simpler approach is to accumulate the predicted probabilities for all the numerous test sets, along with the true class labels of the corresponding instances, and make a single ranked list based on this data. This is the easier method to implement, because for each fold of a 10-fold cross-validation, weight the instances for a selection of different cost ratios, train the scheme on each weighted set, count the true positives and false positives in the test set, and plot the resulting point on the ROC axes. The area under the curve (AUC) is used instead because, roughly speaking the larger the area the better the model [45]. In this study, the second method was used in the calculation of AUC.

ROC analysis was used to evaluate the discrimination ability of the classifiers. The classification performance was then measured by the mean area under the ROC curve (AUC). The mean AUC, as an average performance, gives an indication of a typical AUC obtained using the given input data, and indicates how reliably result is estimated [49-54]. The AUC is usually taken as the index of performance because it gives a single measure of overall accuracy that is independent of any particular threshold [41, 51]. In spite of its advantages, the ROC plot does not give a rule for the classification of cases.

The test performances of the models were determined by the computation of the following statistical parameters:

**Specificity:** number of correct classified healthy subjects/number of total healthy subjects.

**Sensitivity (myopathy):** number of correct classified subjects suffering from myopathy/number of total subjects suffering from myopathy.

**Sensitivity (neurogenic):** number of correct classified subjects suffering from neurogenic disorder/number of total subjects suffering from neurogenic disorder.

**Total classification accuracy:** number of correct classified subjects/number of total subjects.
In general, all techniques accomplished a good performance with mean AUCs up to 0.996. A slightly lower performance was observed when applying LDA as compared to other machine learning methods. Tables 1-6 show confusion matrices for each classifier and Tables 7-10 shows the ICV and ECV performances of these classifiers using the statistical features of DWT subbands as input. The ICV and ECV results do not show much more difference (1-2 %) in the FSVM, SVM and ANN. For ICV (ECV) test the correct classification rate of 91.66% (89.6%) for LDA, 94.08% (93 %) for RBFN, 94.25% (93.66) for MLPANN, 94.83% (93.3%) for C4.5 DT, 96.75% (96%) for SVM and 97.41% (97%) for FSVM have been reached after the first step for all classification techniques. As it can be seen from tables, the specificity, sensitivity values gave similar results. The EMG signal classification rate of RBFN was superior to the classification rate of LDA in a normal group; but classification rate of LDA was superior to the classification rate of RBFN for pathological cases. If EMG signal classification accuracy obtained using RBFN, MLPNN and C4.5 DT are compared; they are almost equal classification accuracies. Although classification accuracy of LDA is the worst, it has superior sensitivity (for both myopathic and neurogenic pathological cases) compared to the other classification techniques. FSVM has the best classification performance if compared to the other classification techniques.

Accurate identification of EMG signal is important for both diagnosis and treatment evaluation. The FSVM classifier identified the groups with an overall accuracy of 95.4% (ROC area=0.965). Classification accuracy improved significantly when statistical features of DWT subbands were used, providing 97.41% accuracy. This effect also resulted in an improvement of ROC area (AUC=0.996) and the accuracy of FSVM was higher than with an LDA, RBFN, C4.5 DT, MLPNN and SVM based classifier. The enhanced classification accuracy of the FSVM using statistical features of DWT subbands as basic EMG signal parameters makes it an attractive alternative for diagnosing of neuromuscular disorders and evaluating the effectiveness of treatment.

4. Discussion

Similar studies of LDA, C4.5. DT, MLPNN and SVM as an EMG signal classifier can be found in different studies [10-15]. If we compare the results taken from this study with the previous results [6-15], the results taken from this study is 3% better than the previous results. Because statistical features extracted from DWT subbands of EMG signals eliminates the noise in the data, extracts the best features for EMG data and provides unbiased estimator. Hence the procedure represented in this study contributes better results than the previous studies. When differentiating EMG signals, performance can be expressed using cross-validation accuracy rates and ROC area. Classification performance has been found to be significantly enhanced when statistical features of DWT subbands and FSVM method are used to train and
test the classifiers. Feature selection and classification method appears, therefore, to be important when dealing with EMG signal classification; this can also reduce the dimensionality of input data with less computational complexity.

The main advantage of the SVM approach is that this flexibility is largely controlled by the algorithm itself on the training data. For the RBF kernel, two parameters have to be selected beforehand: the trade-off parameter $C$ and the kernel parameter $\gamma$. They can be optimised for an optimal generalisation performance in the traditional way by using an independent test set or k-fold cross-validation [54]. On the other hand, the generalisation behaviour can be estimated directly and thus the parameters can be chosen solely based on the training data. In order to test these last two criteria to parameter selection, they were applied on a single realisation of a small training set and subsequently evaluated on a large validation set of samples. It can be seen that the parameter $C$ produces the maximal amplitude of any Gaussian basis function in the decision function. For a chosen width parameter $\gamma$, $C$ therefore controls the maximal distance at which a support vector can contribute significantly to the decision function. Hence, for too low values of $C$, a large number of support vectors were obtained. Such a solution is less economical and needs more computations for evaluating the decision function, as it corresponds to an RBF-network with a larger number of nodes. Therefore, a low number of support vectors can be used as criterion for choosing $C$. Examining the number of support vectors as function of $C$ and $\gamma$ within a reasonable range shows that starting from a sufficiently high value of $C$ the number of support vectors does not decrease anymore. Empirically, it is found that even for lower values of $C$ a good generalisation ability is achieved, provided that the width $\gamma$ is readjusted appropriately. For different settings of $\gamma$, different final values for the margin on the generalisation error are obtained. This can be used for choosing $\gamma$ for a given value of $C$ [55-57].

Concerning the estimation of the prediction error by 10-fold cross-validation, it can be seen that 10-fold CV error has little bias for EMG data set. It would be of significance to consider this for the current problem, where there is also feature-selection bias to be corrected for. As to be expected for training samples of twice or near twice the size, the (estimated) bias was smaller: between 1 and 3% for the EMG data set. The estimated prediction rate according to 10-fold CV error remains essentially constant as statistical features are deleted in the classifier. Hence feature selection provides essentially little improvement in the performance of the classifier for the EMG data set; also it confirms that the number of features does not affect the prediction error which is almost zero. The 10-fold CV error, which has the selection bias removed, is approximately equal to 1% for 15 features. Although the feature vectors have been generated independently of the class labels, a classifier can be formed that has not only an average zero error but also an average CV error close to zero. Also it is important to note that if a test set is used to estimate the prediction error, then there will be a selection bias if this test set was used also in the feature
selection process. Thus the test set must play no role in the feature selection process for an unbiased estimate to be obtained. Concerning the former approach, an internal cross-validation (ICV) is not sufficient. Hence, an external cross-validation (ECV) must be done whereby at each stage of the validation process with the deletion of a subset of the observations for testing, the classifier must be trained on the retained subset of observations by performing the same feature selection procedure used to train the rule in the first instance on the full training set [47]. Hence besides ICV, ECV is used to get sufficient results (see tables 8-10).

Based on the results of the present work and familiarity in EMG signal classification problems, the followings can be emphasized:

i. The high classification accuracy of the FSVM classifier gives insights into the features used for defining the EMG signals. The conclusion drawn in the applications demonstrated that the DWT coefficients are the features, which well represent the EMG signals, and by the usage of these features a good distinction between classes can be obtained.

ii. LDA, RBFN, MLPANN, C4.5 DT, SVM based classifiers are appropriate for use in diagnosis of neuromuscular disorders; but, FSVM has an advantage over other classification methods based on its higher classification accuracy.

iii. C4.5 DT classifier uses an error reduction based training algorithm and as a result provide rules that are used to reduce classification error. This can results to large trees and over-fitting problems. C4.5 DT method cannot find relationships among multiple features.

iv. SVM is based on preprocessing the data to represent patterns in a high dimension — typically much higher than the original feature space. With an appropriate nonlinear mapping to a sufficiently high dimension, data from different categories can always be separated by a hyperplane. As a result, while the original features bring sufficient information for good classification, mapping to a higher dimensional feature space make available better discriminatory evidence that are absent in the original feature space. The problem of training an SVM is to select the nonlinear functions that map the input to a higher dimensional space. Often this choice will be informed by the designer’s knowledge of the problem domain. Polynomials, Gaussians or other basis functions might be used in the absence of such information. The dimensionality of the mapped space can be arbitrarily high. For training the SVM, appropriate kernel parameters C, and γ were selected by using trail and error method. The optimal C, and γ values can only be ascertained after trying out different values. In addition, the choice of γ parameter in
the SVM is crucial in order to have a suitably trained SVM. The SVM has to be trained for different kernel parameters until to get the best result.

v. Although accuracy across the five methods did not differ significantly for the clinical data, FSVM had a significant advantage in its ability to maximize the classification accuracy. The classification results and the values of statistical parameters indicated that the FSVM had considerable success in the EMG signals classification by comparing with other classification methods.

vi. The FSVM classifier is robust to dynamic noise by setting lower fuzzy membership for the data points with the noises or outliers, whereas the SVM classifier is sensitive to the noise. Some data points with outliers and noises could be support vectors in the SVM classifier.

vii. The proposed FSVM classifier has lower computation cost than the SVM classifier due to the fewer support vectors, which is critical for online classification of EMG signals. The present study has shown that the proposed FSVM classifier is very effective for diagnosis of neuromuscular disorders.

5. Conclusion

In this paper, we developed an efficient combination of classifier and features, which proved by the different experiments is applicable for the classification of the EMG signals. This was accomplished using LDA, RBFN, MLPNN, C4.5 DT, SVM and FSVM techniques with statistical features extracted from DWT subbands. Because the experiments proved, the combination represented as the Fuzzy SVM and the statistical features extracted from DWT subbands can achieve a better performance than others classifier over the three EMG signal patterns: normal, myopathic and neurogenic. The experiment shows that the feature set extracted from DWT has superior subject-independent, and intrinsic excellent separability in contrast to those conventional known time and frequency domain features. Also as a classifier, the Fuzzy SVM classifier demonstrated better generalization ability and more rapid execution speed than the other classifier. Besides the proposed Fuzzy SVM classifier with statistical features extracted from DWT subbands meets the requirements for normal, myopathic and neurogenic MUAP characterization and is capable of classifying the EMG signals with a high degree of accuracy and repeatability, invariant of the level of motor impairment. Furthermore the proposed Fuzzy SVM classifier shows promise as a clinically useful method of providing numerical inputs to the next step of the interpretation phase of an EMG examination. This demonstrates that the Fuzzy SVM classifier can be valuable for the capture and expression of knowledge useful to a clinician. These results provide encouragement to develop and evaluate a Fuzzy SVM method for quantifying the level of contribution of a neuromuscular disorder.

Acknowledgements
A. Subasi, Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines, Computers in Biology and Medicine 42, 806–815, 2012.

The author thanks to Dr. Mustafa Yilmaz at University of Gaziantep, Neurology Department for providing the EMG data utilized in this research. This research has been supported by International Burch University (IBU Project no: IBU2010–PRD001).

References


51. J. A. Swets, ROC analysis applied to the evaluation of medical imaging techniques, Invest Radiol 14(2) (1979) 109–121.


Table 1. Confusion Matrix for LDA classifier

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Myopathic</th>
<th>Neurogenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>310</td>
<td>79</td>
<td>11</td>
</tr>
<tr>
<td>Myopathic</td>
<td>6</td>
<td>394</td>
<td>0</td>
</tr>
<tr>
<td>Neurogenic</td>
<td>4</td>
<td>0</td>
<td>396</td>
</tr>
</tbody>
</table>

*aAccording to the medical expert.

Table 2. Confusion Matrix for RBFN classifier

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Myopathic</th>
<th>Neurogenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>349</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>Myopathic</td>
<td>6</td>
<td>393</td>
<td>1</td>
</tr>
<tr>
<td>Neurogenic</td>
<td>8</td>
<td>5</td>
<td>387</td>
</tr>
</tbody>
</table>

*aAccording to the medical expert.

Table 3. Confusion Matrix for MLP classifier

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Myopathic</th>
<th>Neurogenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>369</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>Myopathic</td>
<td>16</td>
<td>384</td>
<td>0</td>
</tr>
<tr>
<td>Neurogenic</td>
<td>19</td>
<td>3</td>
<td>378</td>
</tr>
</tbody>
</table>

*aAccording to the medical expert.
Table 4. Confusion Matrix for C4.5 DT classifier

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Myopathic</th>
<th>Neurogenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal a</td>
<td>368</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>Myopathic a</td>
<td>13</td>
<td>387</td>
<td>0</td>
</tr>
<tr>
<td>Neurogenic a</td>
<td>15</td>
<td>2</td>
<td>383</td>
</tr>
</tbody>
</table>

aAccording to the medical expert.

Table 5. Confusion Matrix for SVM classifier

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Myopathic</th>
<th>Neurogenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal a</td>
<td>374</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>Myopathic a</td>
<td>9</td>
<td>391</td>
<td>0</td>
</tr>
<tr>
<td>Neurogenic a</td>
<td>4</td>
<td>0</td>
<td>396</td>
</tr>
</tbody>
</table>

aAccording to the medical expert.

Table 6. Confusion Matrix for FSVM classifier

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Myopathic</th>
<th>Neurogenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal a</td>
<td>382</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Myopathic a</td>
<td>8</td>
<td>392</td>
<td>0</td>
</tr>
<tr>
<td>Neurogenic a</td>
<td>5</td>
<td>0</td>
<td>395</td>
</tr>
</tbody>
</table>

aAccording to the medical expert.
A. Subasi, Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines, Computers in Biology and Medicine 42, 806–815, 2012.

Table 7. EMG signal classification results for different classifiers using Internal Cross Validation.

<table>
<thead>
<tr>
<th>Statistical Parameters</th>
<th>LDA</th>
<th>RBFN</th>
<th>MLP</th>
<th>C4.5 DT</th>
<th>SVM</th>
<th>FSVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity (%)</td>
<td>77.5</td>
<td>87.3</td>
<td>92.3</td>
<td>92</td>
<td>93.5</td>
<td>95.5</td>
</tr>
<tr>
<td>Sensitivity (Myopathic) (%)</td>
<td>98.5</td>
<td>98.3</td>
<td>96</td>
<td>96.8</td>
<td>97.8</td>
<td>98</td>
</tr>
<tr>
<td>Sensitivity (Neurogenic) (%)</td>
<td>99</td>
<td>96.8</td>
<td>94.5</td>
<td>95.8</td>
<td>99</td>
<td>98.8</td>
</tr>
<tr>
<td>Total Classification Accuracy (%)±SD</td>
<td>91.66±2.82</td>
<td>94.08±2.84</td>
<td>94.25±1.2</td>
<td>94.83±2.89</td>
<td>96.75±2.58</td>
<td>97.41±2.2</td>
</tr>
</tbody>
</table>

Table 8. EMG signal classification results for different classifiers using External Cross Validation.

<table>
<thead>
<tr>
<th>Statistical Parameters</th>
<th>LDA</th>
<th>RBFN</th>
<th>MLP</th>
<th>C4.5 DT</th>
<th>SVM</th>
<th>FSVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity (%)</td>
<td>78</td>
<td>80</td>
<td>91</td>
<td>88</td>
<td>94</td>
<td>95</td>
</tr>
<tr>
<td>Sensitivity (Myopathic) (%)</td>
<td>95</td>
<td>97</td>
<td>94</td>
<td>96</td>
<td>96</td>
<td>98</td>
</tr>
<tr>
<td>Sensitivity (Neurogenic) (%)</td>
<td>96</td>
<td>98</td>
<td>96</td>
<td>96</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>Total Classification Accuracy (%)±SD</td>
<td>80.6±2.8</td>
<td>82 ±2.9</td>
<td>89.66±2.2</td>
<td>85.3±2.91</td>
<td>92±2.45</td>
<td>94±2.4</td>
</tr>
</tbody>
</table>
Table 9. AUC results for different classifiers using Internal Cross Validation.

<table>
<thead>
<tr>
<th></th>
<th>LDA</th>
<th>RBFN</th>
<th>MLP</th>
<th>C4.5 DT</th>
<th>SVM</th>
<th>FSVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0.935</td>
<td>0.963</td>
<td>0.951</td>
<td>0.948</td>
<td>0.959</td>
<td>0.993</td>
</tr>
<tr>
<td>Myopathic</td>
<td>0.99</td>
<td>0.985</td>
<td>0.988</td>
<td>0.983</td>
<td>0.985</td>
<td>0.997</td>
</tr>
<tr>
<td>Neurogenic</td>
<td>0.985</td>
<td>0.977</td>
<td>0.964</td>
<td>0.965</td>
<td>0.995</td>
<td>0.998</td>
</tr>
<tr>
<td>Total</td>
<td>0.943</td>
<td>0.975</td>
<td>0.968</td>
<td>0.965</td>
<td>0.979</td>
<td>0.996</td>
</tr>
</tbody>
</table>

Table 10. AUC results for different classifiers using External Cross Validation.

<table>
<thead>
<tr>
<th></th>
<th>LDA</th>
<th>RBFN</th>
<th>MLP</th>
<th>C4.5 DT</th>
<th>SVM</th>
<th>FSVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0.952</td>
<td>0.965</td>
<td>0.962</td>
<td>0.885</td>
<td>0.946</td>
<td>0.988</td>
</tr>
<tr>
<td>Myopathic</td>
<td>0.975</td>
<td>0.985</td>
<td>0.979</td>
<td>0.977</td>
<td>0.988</td>
<td>0.995</td>
</tr>
<tr>
<td>Neurogenic</td>
<td>0.986</td>
<td>0.994</td>
<td>0.981</td>
<td>0.937</td>
<td>0.990</td>
<td>0.997</td>
</tr>
<tr>
<td>Total</td>
<td>0.971</td>
<td>0.981</td>
<td>0.974</td>
<td>0.933</td>
<td>0.974</td>
<td>0.993</td>
</tr>
</tbody>
</table>
A. Subasi, Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines, Computers in Biology and Medicine 42, 806–815, 2012.
Figure 1. Sub-band decomposition of DWT implementation; $h[n]$ is the high pass filter, $g[n]$ the low pass filter.
A. Subasi, Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines, Computers in Biology and Medicine 42, 806–815, 2012.
Figure 2. EMG signal and its discrete wavelet decomposition into approximate (A1-A5) and detailed (D1-D5) coefficients.
A. Subasi, Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines, Computers in Biology and Medicine 42, 806–815, 2012.

Figure 3. Blok diagram of proposed system.