

Effect of multiscale PCA de-noising on EMG signal classification for diagnosis of neuromuscular disorders

Ercan Gokgoz · Abdulhamit Subasi

Received: 21 January 2014 / Accepted: 10 March 2014 / Published online: 3 April 2014
© Springer Science+Business Media New York 2014

Abstract Different approaches have been applied for quantitative analysis of EMG signals. This study introduces the effect of Multiscale Principal Component Analysis (MSPCA) denoising method in ElectroMyoGram (EMG) signal classification. The effect of the MSPCA denoising method discussed on EMG signal classification. In addition, effect of Multiple Single Classification (MUSIC) feature extraction method presented and compared for the classification of EMG signals. The results were accomplished on the basis of EMG signal data to classify into normal, ALS or myopathic. Furthermore, total accuracy of classifiers such as k-Nearest Neighbor (k-NN), Artificial Neural Network (ANN) and Support Vector Machines (SVMs) were discussed. Significant results were found by using MSPCA denoising method. The comparisons between the developed classifiers were based on a number of scalar performances such as sensitivity, specificity, accuracy, F-measure and area under ROC curve (AUC). The results show that MSPCA de-noising has considerably increased the accuracy as compared to EMG data without MSPCA de-noising.

Keywords Electromyography (EMG) · Motor unit action potentials (MUAPs) · Artificial neural network (ANN) · Support vector machine (SVM) · k-Nearest Neighbor (k-NN) · Multiscale Principle Component Analysis (MSPCA) · Multiple Signal Classification (MUSIC)

This article is part of the Topical Collection on *Patient Facing Systems*

E. Gokgoz (✉) · A. Subasi
Faculty of Engineering and Information Technologies, International
Burch University, Francuske Revolucije bb. Ilidza, Sarajevo 71000,
Bosnia and Herzegovina
e-mail: ercangokgoz@gmail.com

A. Subasi
e-mail: asubasi@ibu.edu.ba

Introduction

Electromyography (EMG) signal is a biomedical signal that measures electrical currents generated in muscles during contraction. The nervous system controls the muscle activity both contraction and relaxation. Muscles are being managed by the nervous system and dependent on the anatomical and physiological structure. EMG signals contain noise while traveling on different tissues. Moreover, the EMG signal acquisition collects signals from motor units at a time which may be effected by different signals. That is what makes EMG signal more complex. Motor Unit Action Potentials (MUAPs) in EMG signals provides an important source of information for the diagnosis of neuromuscular disorders [1]. The nature and characteristics of the signal can be understood and hardware integrations can be made for various EMG signal applications [2]. Recent improvements in technologies of signal processing have made it practical and reliable to develop advanced EMG signal analysis [3–5]. Analysis of EMG signals using powerful and advance methodologies is becoming new trend in biomedical signal processing. Because EMG signal analysis is crucial in clinical diagnosis of neuromuscular disorders.

Recently signal processing techniques and machine learning methods have received extensive attention in EMG signal analysis and classification for diagnosis of neuromuscular disorders. Frequently used signal processing techniques are Fourier transform, autoregressive modeling, wavelet transform, time-frequency approaches [6–8]. ANN used to classify motor unit action potentials (MUAP) of muscles [9]. SVM and ANN are utilized together to diagnose neuromuscular disorders [10, 11]. Neuro-Fuzzy systems are also used for diagnosis of neuromuscular diseases [12]. Tuning and upgrading of regularization and the kernel parameters increasing classification accuracy [13, 14]. Machine learning techniques including Artificial Neural Networks (ANN), dynamic

recurrent neural networks (DRNN), support vector machines (SVM) and fuzzy logic systems used for diagnosis of neuromuscular diagnosis. EMG signal decomposition has been done by wavelet spectrum matching and principle component analysis of wavelet coefficients with reasonable accuracy rates [15]. Multilayer perceptron neural networks (MLPNN), dynamic fuzzy neural network (DFNN), adaptive neuro-fuzzy inference system (ANFIS) and combined feature extraction methods (Autoregressive, discrete wavelet transform and wavelet packed energy) presented according to their effect on accuracy in the classification of EMG signals [16]. PSO optimized SVM classification technique improves accuracy of EMG signal classification [17]. Subasi et al. [18] developed feed forward error back propagation artificial neural networks (FEBANN) and wavelet neural networks (WNN) based classifiers and compared according to their accuracy in classification of EMG signals. They used an autoregressive (AR) model feature extraction as an input to classification system. The success rate for the WNN technique was 90.7 % and for the FEBANN technique 88 %.

One of the major problems in the development of an automatic EMG signal classification systems is the noise problem. During the recording process, noise seriously distorts the signal. Noise can be baseline wandering, motion artifact, power-line interference and electrode pop or contact noise. In this research, noise was removed with Multiscale Principal Component Analysis (MSPCA) technique. MSPCA is employed to generate a helpful representation of EMG signals that removes noise from the signal waveform. Besides denoising methods, the feature extraction methods are also very important for higher classification performance. This paper investigates classification performance of different machine learning methods with Multiscale Principal Component Analysis (MSPCA) de-noising technique using intramuscular EMG signals. In this study, EMG data is taken from different subjects and then MUSIC signal processing techniques and machine learning methods applied to classify EMG signal for diagnosis of neuromuscular disorders. EMG signal classification accuracy improved by utilizing MSPCA. The effects of MSPCA de-noising and MUSIC feature extraction methods are compared and discussed using different performance measures.

The rest of the paper is organized as follows: in the next section, we explained the subjects and present the methods MSPCA, MUSIC, ANN, k-NN, SVM methods. In Section 3 complete experimental results in respect to different classification accuracy measurements such as area under ROC curve, F-measure and classification accuracy is presented. In Section 4, discussion is given on the impact of used denoising and classification methods. Finally, the conclusions are summarized briefly in Section 5.

Materials and methods

EMG data

The EMG signals were recorded and analyzed under usual conditions for MUAP analysis. The recordings were made at low (just above threshold) voluntary and constant level of contraction. Visual and audio feedback was used to monitor the signal quality. A standard concentric needle electrode was used. The EMG signals were recorded from five places in the muscle at three levels of insertion (deep, medium, low). The high and low pass filters of the EMG amplifier were set at 2 Hz and 10 kHz. The material consisted of a normal control group, a group of patients with myopathy and a group of patients with ALS. The control group consisted of 10 normal subjects aged 21–37 years, 4 females and 6 males. 6 out of 10 were in very good physical shape, and the remaining except one was in general good shape. None in the control group had signs or history of neuromuscular disorders. The myopathy group includes 7 patients; 2 females and 5 males aged 19–63 years. All 7 had clinical and electrophysiological signs of myopathy. The ALS group consisted of 8 patients; 4 females and 4 males between 35 and 67 years old. In the meantime, clinical and electrophysiological signs compatible with ALS, 5 of them died in a couple years after beginning of the disorder, supporting the diagnosis of ALS. The brachial biceps and medial vastus muscles were used in this study because they were the most frequently investigated in the two patient groups [19].

MSPCA denoising method

Multiscale Principal Component Analysis (MSPCA) combines the features and capability of Principal Components Analysis (PCA) to decorrelate the variables by obtaining a linear relationship. MSPCA decorrelate autocorrelated measurements by using wavelet analysis to get significant features. MSPCA evaluates the PCA of the wavelet coefficients at each scale and merge the results at related scales. Due to its multiscale manner, MSPCA is convenient for modeling of data for signals whose behavior changes over time and frequency [20].

Signal processing by PCA is extensively used as a classical multiscale signal processing tool [21–23]. PCA has been applied in different fields of science and engineering to better utilize its ability [20, 24, 25]. For biomedical signals like EMG, a robust extension of classical PCA by analyzing shorter signal segments is suggested [26]. It may be used in data reduction, beat detection, classification, signal separation and feature extraction [27]. The desired quality of processed signals is achieved by selecting the principal components (PC) based on energy features in selected wavelet subset matrices. The number of PC selection is based on cumulative

percentage of total variation of variances to control the quality of denoised signals. PCA is performed on these matrices for signal denoising. PCA denoising follows the idea that retaining only the principal components with highest variance to reconstruct the decomposed signal. The choice of multiscale matrices and selection of eigenvalues preserve the desired energy in the processed signals [20, 24, 25]. PCA transforms the data matrix statistically diagonalizing method by the covariance matrix. The process is obtaining the correlation between variables in the data. If the measured and evaluated variables are related, the first few entities present the relationship between the variables. The MSPCA methodology decomposes each variable on a selected type of wavelets. The PCA model is determined one by one for the coefficients at each scale [20].

PCA transforms $n \times p$ data matrix as \mathbf{X} . The variables calculated as a linear weighted sum obtained as,

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T \tag{1}$$

where, \mathbf{P} represents the principal component loadings, \mathbf{T} is the principal component scores, n and p are represents number of dimensions and variables. The principal component loadings represent direction and position of the hyperplane which evaluate the highest probability of residual variance referred measured variables. Multiscale PCA (MSPCA) combines the capability of PCA to extract the crosscorrelation or relationship between the variables with orthonormal wavelets. The measurements for each variable (column) are decomposed to its wavelet coefficients using the same orthonormal wavelet for each variable in order to combine the benefits of PCA and wavelets. This results in transformation of the data matrix, \mathbf{X} into a matrix, \mathbf{WX} , where \mathbf{W} is an $n \times n$ orthonormal matrix representing the orthonormal wavelet transformation operator containing the filter coefficients. The matrix, \mathbf{WX} is as big as the original data matrix size, \mathbf{X} , but related to the wavelet decomposition, the deterministic component for each variable in \mathbf{X} is compressed in a relatively small number of coefficients in \mathbf{WX} , when the stochastic supplement in each variable is approximately decorrelated in \mathbf{WX} , and propagate over all components according to its power spectrum. To make use the multiscale components of the data, PCA of the covariance matrix coefficients at every scale is calculated separately from the other scales. The outcoming scores at every scale are not cross correlated because of PCA, and their autocorrelation is roughly decorrelated because of the wavelet decomposition. Being dependent on the character of the application, a lesser subset of the principal component scores and wavelet coefficients can be chosen at every scale. The amount of principal components, to be absorbed at every scale is not altered by the wavelet decomposition because it does not influence the core relationship connecting the variables at any scale. As a result,

existing techniques like, cross validation, screen test, or parallel analysis may be used for the data matrix in the time domain or to every wavelet coefficient to choose the appropriate amount of components. For choosing the relevant wavelet coefficients, distinct methods may be applied according to the application. Using split thresholds at every scale permits MSPCA to be more responsive to scale-varying signal characteristics like autocorrelated measurements. Limiting of the coefficients at every scale points the area of the time-frequency space. Details are given in [20].

Feature extraction using multiple signal classification (MUSIC) method

EMG signal includes its own signal components. Raw EMG data need to be extracted to define the composed features. Signal should be preprocessed in order to feed classification algorithms to classify efficiently [15]. The frequency content of the signal extracted from raw EMG signal by MUSIC method.

The multiple signal classification is a frequency estimation method as a development of Pisarenko harmonic decomposition which is both first frequency estimation method and based on eigen decomposition [28–30]. The dimensional space is divided into signal and noise components using the eigenvectors of the correlation matrix [31].

$$\mathbf{Q}_s = [\mathbf{q}_1 \mathbf{q}_2 \dots \mathbf{q}_P], \mathbf{Q}_w = [\mathbf{q}_{P+1} \dots \mathbf{q}_M] \tag{2}$$

Length of the time window $M = P + 1$ that means number of complex exponentials are 1 greater than the number of complex exponentials to allow the size of the time window like $M > P + 1$. Hence, dimension of noise subspace is greater than 1. Advantage of larger dimension is procuring an improved and stronger frequency estimation method. Due to the orthogonality between the noise and signal subspaces, all time window frequency vectors of the complex exponentials are orthogonal to the noise subspace from

$$\mathbf{P}_s \mathbf{v}(f_p) = \mathbf{v}(f_p) \quad \mathbf{P}_w \mathbf{v}(f_p) = \mathbf{0}. \tag{3}$$

Thus, for each eigenvector ($P < m \leq M$)

$$\mathbf{v}^H(f_p) \mathbf{q}_m = \sum_{k=1}^M q_m(k) e^{-j2\pi f_p(k-1)} = 0 \tag{4}$$

for all the P frequencies f_p of the complex exponentials. Therefore, if we compute a pseudospectrum for each noise eigenvector as

$$\bar{R}_m(e^{j2\pi f}) = \frac{1}{|\mathbf{v}^H(f) \mathbf{q}_m|^2} = \frac{1}{|\mathbf{Q}_m(e^{j2\pi f})|^2} \tag{5}$$

the polynomial Q_m has $M - 1$ roots. These roots produce P peaks in the pseudospectrum from Eq. (5) [31].

The pseudospectra of $M - P$ noise eigenvectors shares its roots with the signal subspace. The remaining roots of the noise eigenvectors procure at different frequencies. There are no restrictions on the range of roots. If there is some near the group circle, can produce extra peaks in the pseudospectrum.

$$\bar{R}_{music}(e^{j2\pi f}) = \frac{1}{\sum_{m=p+1}^M |v^H(f)q_m|^2} = \frac{1}{\sum_{m=p+1}^M |Q_m(e^{j2\pi f})|^2} \quad (6)$$

The P complex exponentials' frequency estimation taken as the P peaks in pseudospectrum. Since in the Eq. (6) does not include information about the powers of the complex exponentials or the background noise level, pseudospectrum term is used [31].

The presumption in the MUSIC pseudospectrum is the noise eigenvalues equal to $\lambda_m = \sigma^2 w$, which means the noise is white. But in practice the noise eigenvalues will not be equal when prediction method used instead of correlation matrix. The variations become more obvious when the correlation matrix is predicted by small samples of data. Therefore, a little variation on the MUSIC algorithm, known as the *eigenvector (ev) method*, was intended to define potentially different noise eigenvalues [32]. The pseudospectrum is

$$\bar{R}_{ev}(e^{j\omega}) = \frac{1}{\sum_{m=p+1}^M \frac{1}{\lambda_m} |v^H(f)qm|^2} = \frac{1}{\sum_{m=p+1}^M \frac{1}{\lambda_m} |Q_m(e^{j2\pi f})|^2} \quad (7)$$

where λ_m is the eigenvalue corresponding to the eigenvector q_m , for the method. The pseudospectrum is normalized by its corresponding eigenvalue for each eigenvector. The eigenvector and MUSIC methods are identical in case of equal noise eigenvalues ($\lambda_m = \sigma^2 w$) for $P+1 \leq m \leq M$ [31].

The peaks in the MUSIC pseudospectrum corresponds approaches to zero in Eq. (7). As a matter of fact, we may prefer z-transform instead of this denominator which is sum of the z-transforms and formulated as [31].

$$\bar{P}_{music}(z) = \sum_{m=p+1}^M Q_m(z)Q_m^* \left(\frac{1}{z^*} \right) \quad (8)$$

Classification algorithms

k-nearest neighbor (*k*-NN)

K-nearest neighbour (k-NN) is a popular algorithm due to its speed, because there is almost no learning process. Theoretical background of k-NN is pretty simple. There are K training inputs with a specified volume [33].

We first point an input by x and the number of inputs in one class as N_k , where $k \in \{1, 2\}$. We use two simple probabilities to quantify [33]. The simple probability equations are described below

$$p(x|k) = \frac{N_k}{N} \quad (9)$$

where $N=N_1+N_2$. The background probability is evaluated by the following equation

$$p(k|x) = \frac{p(x|k)p(k)}{p(x|1)p(1) + p(x|2)p(2)} \quad (10)$$

where $p(k)$ is the a prior probability of the k^{th} class. If $p(1)=p(2)$ we have

$$p(k|x) = \frac{N_k}{N} \quad (11)$$

The query vector is tagged as class 1 If $N_1 > N_2$. If the prior information is updated to $p(1) > p(2)$ then we have

$$p(k|x) = \frac{p(k)N_k}{p(1)N_1 + p(2)N_2} \quad (12)$$

There is also another way to find same result. The probability of the k th class is defined for given a volume V

$$p(x|k) = \frac{N_k}{N} \quad (13)$$

The background probability of the k th class is described as

$$p(k|x) = \frac{\frac{N_k}{V} p(k)}{\frac{N_1}{V} p(1) + \frac{N_2}{V} p(2)} \quad (14)$$

It seems like Bayes rule which is the base for deriving a k-nearest neighbor classification. k-NN has a model selection problem like other machine learning algorithms. k number must be carefully selected based on model specifically. Distance measure formulas should be selected because different formulas may cause different outputs such as Manhattan and Euclidean distance formulas [33].

Artificial neural networks (ANN)

Artificial neural networks (ANNs) are computing systems with many simple, highly interconnected processing elements which are called nodes or artificial neurons to abstractly simulate the organization, structure and operation of the

biological nervous system. A biological neuron collects input signals from near neurons via the dendrites, conjugates the signals, and produces an output response then forward to other neurons via the axon. In addition that, a neuron makes a connection to other neurons by axons with the dendrite tree of a neighboring neuron. The neurons are characterized as mathematical function with numerical inputs and a single output [34]. It is obvious that learning can be modeled mathematically like classification, if the first perceptron has a number of inputs. Learning in ANNs is accomplished through special training algorithms developed based on learning rules presumed to imitate the learning mechanisms of biological systems. There are many architectures and types of neural networks especially for learning functions in signal processing field. The details are well documented in [35, 36].

The perceptron and the biological neuron are comparable in the following ways [6]. The input of the perceptron has different features, thus becomes similar to real dendrite branches of the biological neuron. The synapse connection w_j is modeled as weights which is very important for summation and activation function. The processing of the weighted inputs to calculate the output of the perceptron is done by summation node using following mathematical function which acts like nucleus of the biological neuron:

$$c_i = f(\sum_j w_j x_{ij}) = f(u_i) \tag{15}$$

The activation function f effects neuron's behavior and it is generally has a limited increasing and nonlinear function like sigmoid function:

$$f(u_i) = \frac{1}{e^{-u_i}} \tag{16}$$

The output c_i is result of calculation which is spreading along the axon. In the perceptron, learning quality and rate is equivalent to optimum adjustments of weights to compare the real network output and the expected network output for a particular training input. The perceptron learning algorithm is not able to classify when the data is not separable. But, perceptron could produce much more separating limitations if connected like a network that has links in each other [37]. This neural network design is known as the multilayer perceptron (MLP) which is one of the most common neural network designs. MLP design has a single input layer attached to output layer via a hidden layer. In the network, every node, in each layer, is linked to every node in the next layer. An N-layer MLP network generally has an input layer, N - 1 hidden layers and an output layer. The complexity of the separation boundary depends on the number of hidden layers and nodes per layer. The activation functions create separator boundaries instead of linear function [6].

Support vector machines (SVMs)

Vapnik established Support vector machines (SVMs) in 1992 as a Supervised Learning Method [38]. SVMs are set of supervised machine learning methods which is using in computer science for classification. Although basic SVM is type of tree algorithm, SVMs are capable to classify via different kernel approaches. The main advantage of SVM is having the best generalization capability on statistical learning theory field [39]. Base of SVM is a linear separator by a hyperplane with a given margin sizes. The hyperplane divides into two classes which are linearly dividable groups of points in a multidimensional space. In theory this situation is possible but in daily life both classes may not linearly separable since they mixed up. This case confuses the classifier which is using linear hyperplane. Due to fact that, there is a requirement to apply a nonlinear transformation from input space to feature space [6].

The hyperplane shows evaluated sum of all training samples x_i . Algorithm searching for the next hyperplane to make possible the separation linearly in feature space,

$$f(x) = \sum_{i=1}^n w_i \phi(x_i) + b \geq 1 \quad \text{if } y_i = +1 \tag{17}$$

$$f(x) = \sum_{i=1}^n w_i \phi(x_i) + b \leq -1 \quad \text{if } y_i = -1 \tag{18}$$

Equations (17) and (18) can be used together for a strict linear discriminant function

$$y_i f(x) \geq 1 \tag{19}$$

The closest points to the hyperplane have an important role since they detect the size of class margins and boundaries. If linear separation not applied to both two classes in feature space, strict discriminant function cannot be weakened which is represented in Eq. (19). To make weaker, the soft-margin SVM classifier is defined as [40]:

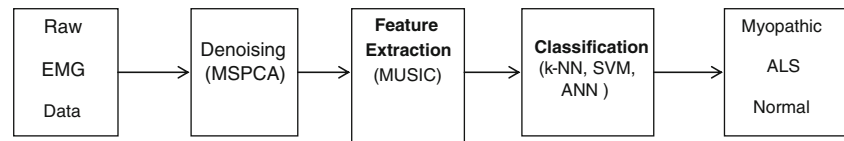
$$\tilde{S}(X, \tilde{S}) = \frac{1}{2} \sum_{i=1}^n w_i^2 + C \sum_{i=1}^n S_i \tag{20}$$

subject to

$$y_i f(x) \geq 1 - S_i \quad S_i > 0 \tag{21}$$

The parameter 'C' is an adjuster for the relationship between the training errors and generalization capability [6].

Fig. 1 Blok diagram of proposed system



Experimental results

The decomposition of EMG signals into integrated MUAPs and their classification into groups of similar characteristics is a common machine learning problem [18]. Complexity of the problem is growing since MUAPs waveforms have unknown and dissimilar shapes. Main goals of analyzing signals are to get better accuracy in shorter time with flexible and useful methods whose reliabilities are accepted. Our EMG data collected from 25 patients. The control group consisted of 10 normal subjects aged 21–37 years, 4 females and 6 males. 6 out of 10 were in very good physical shape, and the remaining, except one, were in general good shape. In control group none of them have any signs or history of neuromuscular disorders. The group with myopathy consisted of 7 patients; 2 females and 5 males aged 19–63 years. All 7 had clinical and electrophysiological signs of myopathy¹⁵. The ALS group consisted of 8 patients; 4 females and 4 males aged 35–67 years [19].

In this study, MSPCA denoising and MUSIC feature extraction method have been applied for EMG signal classification and good accuracy is achieved (Fig. 1). In order to analyze the coherence between method combinations of our approach, the whole EMG data is separated as training and test data since sets must be selected independently from each other. The classification model built via training set and verified by the test set.

Since, k-fold cross-validation method is a well-known and reliable method for performance evaluation, many researchers used the k-Fold cross validation method to decrease bias which is related with random sampling of the produced data sets [18, 16, 41, 42]. K-fold cross validation randomly divide the data into k subsets which are calling as folds and approximately same size. The

cross-validation accuracy (CVA) is the average of the k individual accuracy measures

$$CVA = \frac{1}{k} \sum_{j=1}^k A_i \quad (22)$$

where k (10 in our case) is the number of folds used, and A_i is the accuracy measure of each fold [43]. The training data is used for building of the classification model and the test data is used for validation.

The number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are used to evaluate performance of classifiers. Different definitions are used for different domains. The *sensitivity* and *specificity* are extensively used in diagnostic and detection tests. Sensitivity shows the amount of people with disease, who have a positive test result,

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (23)$$

Specificity shows the amount of people without disease who have a negative test result, which is $1 - FP$ and defined as:

$$Specificity = \frac{TN}{TN + FP} \times 100\% \quad (24)$$

Accuracy is used as a measure, which is:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \times 100\% \quad (25)$$

Table 1 Confusion Matrix for k-NN, ANN, and SVM Classifiers

Actual ^a /Predicted	k-NN			ANN			SVMs		
	ALS	Myopathic	Normal	ALS	Myopathic	Normal	ALS	Myopathic	Normal
ALS	286	8	6	287	6	7	289	0	14
Myopathic	21	255	24	5	246	49	7	273	20
Normal	27	75	198	11	10	279	12	17	271

^a According to the medical expert

Table 2 Comparison of performance results of three different classifiers without MSPCA de-noising

Statistical Parameters	k-NN	ANN	SVMs
Specificity (%)	51,33	61,33	54,66
Sensitivity (Myopathic) (%)	49	34,66	53,66
Sensitivity (ALS) (%)	51,33	56,66	62,66
Total Classification Accuracy (%)	50,55	50,89	57

F-measure is an another measure used to evaluate performance [17]

$$F\text{-measure} = \frac{2 TP}{2 TP + FN + FP} \times 100\% \quad (26)$$

We calculated performances of different machine learning methods. Classification accuracy is the number of correctly classified instances divided by the full number of instances. Results are shown in Table 1.

Results without using MSPCA denoising method are not good enough with the EMG data for any of the classifier. On the other hand, as it can be seen easily by comparing Tables 1, 2 and 3 and Figs. 2 and 3 results with MSPCA denoising are considerably improved. SVMs obtained the best classification accuracy with 92.55 %, ANN is 90.02 % and k-NN is 82.11 % accuracy. This results shows that SVM is better than the others for EMG patterns with MSPCA denoising. As said in literature SVMs have one of the best generalization capabilities. In some cases SVM can classify effectively but in our case which is classification of EMG data, SVMs could not achieve efficient performance. Since, EMG has a complex structure, SVMs and other classifiers need to be feed with more clear data. In spite of the fact that features of EMG data needs to be extracted to obtain better performance from classifiers [39]. Since SVMs have an adjuster parameter for the relationship between the training errors and generalization capability, error rate is not increasing when capability increased [6]. It is obtained that to confirm the optimum number of the parameter within a logical range is higher accuracies obtained by using higher values. The accuracy for SMV with MSPCA denoising

Table 3 Comparison of performance results of three different classifiers with MSPCA de-noising

Statistical Parameters	k-NN	ANN	SVMs
Specificity (%)	66	93	90,3
Sensitivity (Myopathic) (%)	85	82	91
Sensitivity (ALS) (%)	95,3	95,67	96,33
Total Classification Accuracy (%)	82,11	90,02	92,55

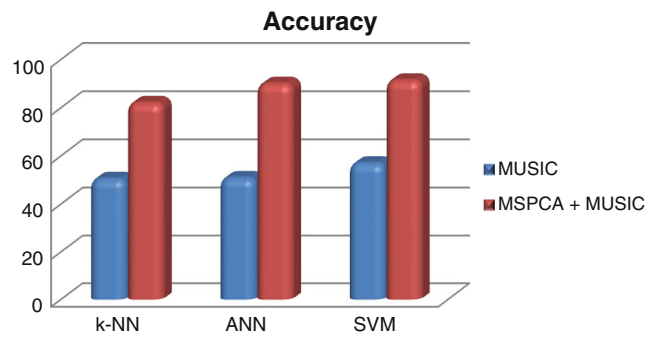


Fig. 2 Graphical representation of results of both approaches

is 92.55 %, conversely 57 % without MSPCA. This proves the great advantage of using MSPCA denoising.

For Myopathy, highest sensitivity obtained with SVMs in both with and without MSPCA denoising. But for ALS, highest sensitivity obtained by ANN which shows that for ALS, ANN can perform better than others. Highest specificity obtained by ANN in both cases which means ANN is compatible to classify normal EMG data. MSPCA perform best on ANN with almost 40 % increment which was following by SVMs with 35 % increment. According to fact that better total accuracy can be expected from SVMs with MSPCA denoising and ANN has better compatibility with SVM.

The classification performance also measured via receiver operating characteristic (ROC) curve which is graphical technique. It is produced by plotting true positives as percentage of all positives in the sample versus negative ones using cross-validation. For each fold of cross-validation it counts the true and false positives in the test set and plot the results to ROC axes [44]. Then, the classification performance can be measured by the mean area, which is the area under the ROC curve (AUC). The mean of area which is under curve as an average gives result using drawn and calculated axis points to evaluate how reliable the estimated results. Correspondingly, the quality of the estimation of a curve depends on the number tested thresholds. The AUC is generally taken as the index of performance because it generates one accuracy output which not

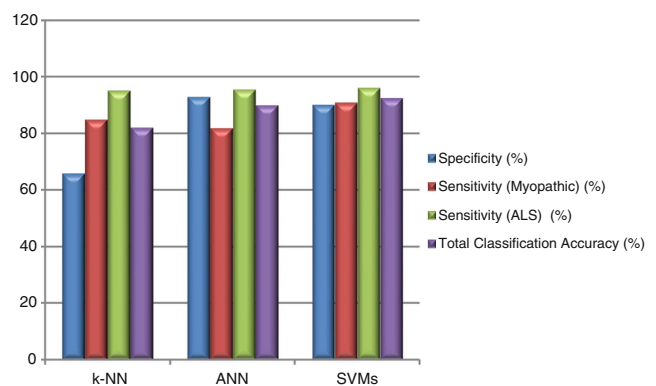


Fig. 3 Graphical representation of evaluation performance of classifiers

Table 4 AUC results for different classifiers

	k-NN	ANN	SVM
Normal	0.88	0.952	0.923
Myopathic	0.89	0.958	0.962
ALS	0.956	0.979	0.984
Total	0.909	0.963	0.956

depends on any defined threshold [45–47]. It can be seen from Table 4 that AUC results are close to previous results but not in the same order. Total AUC of ANN is slightly better than other classifiers with 0.963. Since there is coherence between normal data and ANN, 0.952 which is best for normal data obtained. SVM is following ANN with 0.956. Even if results little higher, best classification rate also obtained by ANN for Normal data, comparing to SVM. Since results are very close to each other, it is Acceptable.

F-measure is another performance measurement which supports our results. As shown from Table 5 best total accuracy obtained by using SVM which is following by ANN. Although total accuracy of SVM is better than ANN, ANN performed surprisingly efficient for ALS data with 0.952 which is little better than SVM.

Discussion

Other significant EMG classification studies can be find in literature such as [3, 48–50, 39, 16]. In some studies EMG signal data obtained from surface and classified using some other techniques while some of them using simulated EMG datasets. We used intramuscular EMG signal data which gathered using needle from muscle. Since collection method is different, it cannot be said both of the data is EMG data and by using same methods, same accuracy rate can be achieved.

SVM approach can be applied to classifying EMG signals due to their wide generalization capability, feasibility and flexibility compared to traditional classification techniques. Advantage of the capability provides higher classification accuracies and a lower sensitivity. Effects of the cost parameter C and kernel parameter γ values for EMG signal dataset are discovered. C organizes the range of support vector which can provide to the decision function. As a result, by using low

Table 5 F-measure of different classifiers

	k-NN	ANN	SVM
Normal	0.75	0.879	0.9
Myopathic	0.799	0.875	0.925
ALS	0.902	0.952	0.951
Total	0.817	0.902	0.925

values of C , big numbers of support vectors were obtained and more computations required determining the decision function. Hence, low number of support vectors can be used as parameter while choosing C . It is obtained that to confirm the optimum number of support vectors which are as function of C and γ within a logical range, while applying sufficiently high values of C , number of support vectors does not decrease anymore. Higher accuracies obtained by using higher C values. As an experiment, it is found that for lower values of C good generalization ability is achieved. For different γ values, different margin sizes are obtained on the generalization [40, 51, 52]. Different kernels, C and gamma values were tried to find the best accuracy results, the best accuracy obtained by using poly kernel with degree two, C value 650 and gamma value 0.001.

Regarding with presented results in Table 1 and Table 2 it can be said that denoising EMG data with MSPCA is efficiently improved classification accuracy. After denoising with MSPCA accuracy uprised from 57 to 92.55 % for SVM, from 50.89 to 90.02 % for ANN, from 50.55 to 82.11 % for k-NN. Consequently, we can say that EMG data is more coherent to classify with SVM classifier either not denoised or denoised. In total classification accuracy SVMs achieved best performance although it wasn't best on specificity. Best specificity achieved by ANN both with and without MSPCA denoised EMG data. While specificity of SVM is 90.3 % ANN achieved the best with 93 %. According to fact that it can be said normal data should be classify with ANN to get highest accuracy.

Conclusion

In this paper, we find out considerable accompanies between classifiers, denoising and decomposition methods, which proved by the results. Novelty of this study is achieving significantly better performance by denoising with MSPCA over the three EMG signal patterns: normal, myopathic and ALS. This was concluded using classifiers such as ANN, k-NN, SVM and MUSIC decomposition method with MSPCA denoising. During investigation of accompany between used methods, we comprehend how to get higher accuracy by rearranging kernel parameters of SVM. Generally it is better not to use default parameter values of SVM kernels. The classification accuracy with MSPCA improved to 92.55 % which was 57 % without MSPCA using same MUSIC and SVM combination. Classification accuracy of other classifiers also increased such as ANN and k-NN. Improvement of classification accuracy for SVM is 35.55 %, for ANN is 39.87 %, for k-NN is 31.66 %. MSPCA denoising can become helpful for diagnosing of neuromuscular disorders regarding these results and methods.

References

1. McGill, K. C., Optimal resolution of superimposed action potentials. *IEEE Trans. Biomed. Eng.* 49(7):640–650, 2002.
2. Kimura, J., *Electrodiagnosis in diseases of nerve and muscle: principle and practice*, 3rd edition. Oxford University Press, Newyork, 2001.
3. Katsis, C. D., Goletis, Y., Likas, A., Fodiatis, D. I., and Sarmas, I., A novel method for automated EMG decomposition and MUAP classification. *Artif. Intell. Med.* 37(1):55–64, 2006.
4. Loudon, G. H., Jones, N. B., and Sehmi, A. S., New signal processing techniques for the decomposition of EMG signals. *Med. Biol. Eng. Comput.* 30(6):591–599, 1992.
5. Pattichis, C. S., and Pattichis, M. S., Time-scale analysis of motor unit action potentials. *IEEE Trans. Biomed. Eng.* 46(11):1320–1329, 1999.
6. Begg, R., Lai, D. T. H., and Palaniswami, M., *Computational intelligence in biomedical engineering*. CRC Press, Boca Raton, 2008.
7. Subasi, A., Application of classical and model-based spectral methods to describe the state of alertness in EEG. *J. Med. Syst.* 29(5):473–486, 2005.
8. Subasi, A., and Kiyimik, M. K., Muscle Fatigue detection in EMG using time–frequency methods, ICA and neural networks. *J. Med. Syst.* 34(4):777–785, 2010.
9. Schizas, C. N., Pattichis, C. S., and Middleton, L., Artificial neural nets in computer-aided macro motor unit potential classification. *IEEE Eng. Med. Biol. Mag.* 9(3):31–38, 1990.
10. Güler, N. F., and Koçer, S., Use of support vector machines and neural network in diagnosis of neuromuscular disorders. *J. Med. Syst.* 29(3):271–284, 2005.
11. Akben, S. B., Subasi, A., and Tuncel, D., Analysis of repetitive flash stimulation frequencies and record periods to detect migraine using artificial neural network. *J. Med. Syst.* 36(2):925–931, 2012.
12. Koçer, S., Classification of Emg signals using neuro-fuzzy system and diagnosis of neuromuscular diseases. *J. Med. Syst.* 34(3):321–329, 2010.
13. Huang, C. L., and Wang, C. J., A GA-based attribute selection and parameter optimization for support vector machine. *Expert Syst. Appl.* 31(2):231–240, 2006.
14. Ho, S. Y., Shu, L. S., and Chen, J. H., Intelligent evolutionary algorithms for large parameter optimization problems. *IEEE Trans. Evolut.* 8(6):522–541, 2004.
15. Reaz, M. B. I., Hussain, M. S., and Mohd-Yasin, F., Techniques of EMG signal analysis: detection, processing, classification and applications. *Biol. Proced. Online* 8:11–35, 2006.
16. Subasi, A., Classification of EMG signals using combined features and soft computing. *Appl. Soft Comput.* 12(8):2188–2198, 2012.
17. Subasi, A., Classification of EMG signals using PSO optimized SVM for diagnosis. *Comput. Biol. Med.* 43(5):576–586, 2013.
18. Subasi, A., Yilmaz, M., and Ozcalik, H. R., Classification of EMG signals using wavelet neural network. *J. Neurosci. Methods* 156(1–2):360–367, 2006.
19. Nikolic, M., *Findings and firing pattern analysis in controls and patients with myopathy and amyotrophic lateral sclerosis*. Faculty of Health Science, University of Copenhagen, Copenhagen, 2001.
20. Bakshi, B. R., Multiscale PCA with application to MSPC monitoring. *AIChE J* 44(77):1596–1610, 1998.
21. Chawla, M. P. S., A comparative analysis of principal component and independent component techniques for electrocardiograms. *Neural Comput Appl Springer-Verlag London Limited* 18(6):539–556, 2009.
22. Hotelling, H., Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24(6):417–441, 1933.
23. Jolliffe, I. T., *Principal component analysis*, 2nd edition. Springer, New York, 2002.
24. Bakshi, B. R., Multiscale analysis and modeling using wavelets. *J. Chemom.* 13(3–4):415–434, 1999.
25. Misra, M., Yue, H. H., Qin, S. J., and Ling, C., Multivariate process monitoring and fault diagnosis by multi-scale PCA. *Comput. Chem. Eng.* 26(9):1281–1293, 2002.
26. Kotas, M., Application of projection pursuit based robust principal component analysis to ECG enhancement. *Biomed Signal Process Control* 1(4):289–298, 2006.
27. Castells, F., Laguna, P., Sormmo, L., Bollmann, A., and Roig, J., Principal component analysis in ECG signal processing. *EURASIP J Adv Signal Process* 2007:98–119, 2007.
28. Pisarenko, V. F., The retrieval of harmonics from a covariance function geophysics. *J. R. Astron. Soc.* 33(3):347–366, 1973.
29. Schmidt, R. O., Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag* 34(3):276–280, 1986.
30. Bienvenu, G., and Kopp, L., Optimality of high resolution array processing using the eigensystem approach. *IEEE Trans Acoust Speech Signal Process* 31(5):1235–1248, 1983.
31. Manolakis, D. G., Ingle, V. K., and Kogon, S. M., *Statistical and adaptive signal processing*. Artech House, London, 2005.
32. Johnson, D. H., and DeGraaf, S., Improving the resolution of bearing in passive sonar arrays by eigenvalue analysis. *IEEE Trans Acoust Speech Signal Process* 30(4):638–647, 1982.
33. Yang, Z. R., *Machine learning approaches to bioinformatics*. World Scientific Publishing Co. Pte. Ltd., Singapore, 2010.
34. Rosenblatt, F., The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* 65:386–408, 1958.
35. Subasi, A., and Erçelebi, E., Classification of EEG signals using neural network and logistic regression. *Comput. Methods Prog. Biomed.* 78(2):87–99, 2005.
36. Haykin, S., *Neural networks: a comprehensive foundation*. Prentice-Hall, Saddle River, 1999.
37. Minsky, M., and Papert, S., *Perceptrons*. MIT Press, Cambridge, 1988.
38. Vapnik, V., *Estimation of dependencies based on empirical data*, springer, 2006.
39. Dobrowolski, A. P., Wierzbowski, M., and Tomczykiewicz, K., Multiresolution MUAPs decomposition and SVM-based analysis in the classification of neuromuscular disorders. *Comput. Methods Prog. Biomed.* 107(3):393–403, 2011.
40. Cristianini, N., and Shawe-Taylor, J., *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, Surrey, 2000.
41. Sandvig, J. J., Mobasher, B., and Burke, R., A survey of collaborative recommendation and the robustness of model-based algorithms. *IEEE Computer Soc. Technical Committee Data Eng.* 31(2):3–13, 2008.
42. Zweig, M. H., and Campbell, G., Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.* 39(8):561–577, 1993.
43. Zhang, R., McAllister, G., Scotney, B., McClean, S., and Houston, G., Combining wavelet analysis and Bayesian networks for the classification of auditory brainstem response. *IEEE Trans. Info. Tech. Biomed* 10(3):458–467, 2006.
44. Witten, I., and Frank, E., *Data mining, practical machine learning tools*. Morgan Kaufmann Publishers (Elsevier), San Francisco, 2005.
45. Hanley, J. A., and McNeil, B. J., The meaning and use of the area under a receiver operating characteristic(ROC). *Radiology* 143:29–36, 1982.
46. Obuchowski, N. A., Receiver operating characteristic curves and their use in radiology. *Radiology* 229:3–8, 2003.
47. Swets, J. A., ROC analysis applied to the evaluation of medical imaging techniques. *Investig. Radiol.* 14:109–121, 1979.
48. Pino, L. J., Stashuk, D. W., Boe, S. G., and Doherty, T. J., Motor unit potential characterization using pattern discovery. *Med. Eng. Phys.* 30(5):563–573, 2008.

49. Yan, Z., Wang, Z., and Xie, H., The application of mutual information-based feature selection and fuzzy LS-SVM-based classifier in motion classification. *Comput. Methods Prog. Biomed.* 90(3):275–284, 2008.
50. Kamayuko, E. N., Farina, D., Yoshida, K., and Jensen, W., Relationship between grasping force and features of single channel intramuscular EMG signals. *J. Neurosci. Methods* 185(1):143–150, 2009.
51. Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. J., and Smeyers-Verbeke, J., *Handbook of chemometrics and qualimetrics*. Elsevier, Amsterdam, p. 1997, 1997.
52. Belousov, A. I., Verzakov, S. A., and Frese, J., A flexible classification approach with optimal generalisation performance: support vector machines. *Chemom. Intell. Lab. Syst.* 64(1): 15–25, 2002.