sciendo

Polish Journal of Medical Physics and Engineering
The Journal of Polish Society of Medical Physics
ISSN 1898-0309, doi: 10.2478/pjmpe-2020-0018

September 2020
Vol 26, Issue 3

Scientific Paper

Diagnosis of amyotrophic lateral sclerosis (ALS) disorders based on electromyogram (EMG) signal analysis and feature selection

Aicha MOKDAD[a], Sidi Mohammed El Amine DEBBAL[a,*], Fadia MEZIANI[a]

[a]*Genie Biomedical Laboratory (GBM), Faculty of Technology, University of Abou Bekr Belkaid Tlemcen, Tlemcen, Algeria*
[*]*E-mail address: adebbal@yahoo.fr*

(received 31 March 2020; revised 13 June 2020; accepted 14 July 2020)

Abstract

Electromyogram signal (EMG) provides an important source of information for the diagnosis of neuromuscular disorders. In this study, we proposed two methods of analysis which concern the bispectrum and continuous wavelet transform (CWT) of the EMG signal then a comparison is made to select which one is the most suitable to identify an abnormality in biceps brachii muscle in the main purpose is to assess the pathological severity in bifrequency and time-frequency analysis applying respectively bispectrum and CWT. Then four time and frequency features are extracted and three popular machine learning algorithms are implemented to differentiate neuropathy and healthy conditions of the selected muscle. The performance of these time and frequency features are compared using support vector machine (SVM), linear discriminate analysis (LDA) and K-Nearest Neighbor (KNN) classifier performance. The results obtained showed that the SVM classifier yielded the best performance with an accuracy of 95.8%, precision of 92.59% and specificity of 92%. followed by respectively KNN and LDA classifier that achieved respectively an accuracy of 92% and 91.5%, precision of 92% and 85.4%, and specificity of 92% and 83%.

Key words: bispectrum; continuous wavelet transforms (CWT); support vector machine (SVM); linear discriminate analysis (LDA); K-Nearest Neighbor (KNN).

Introduction

EMG signal is a small electrical current generated by muscle fibers. For clinical diagnosis, the processing of these signals is the main concern of researchers because various neuromuscular diseases can be determined analyzing the time and frequency domain properties of EMG signals.[1] Amyotrophic Lateral Sclerosis (ALS) must be mentioned when we spoke about neuromuscular disorders because of its rapid progressive by the destruction of nerve cells in the brain and spinal cord.[2] This pathology can be identified by analyzing the change between the healthy EMG signal and compare it with the pathological one which can be better understood by time and frequency analysis of the EMG signals in order to investigate and determine the source of the disorder. These promising results allowed us to introduce the bispectrum and CWT.

Kaplanis et al. have used Higher-order spectra (HOS) to analyze the surface electromyogram (sEMG) signal.[3] They have shown that the level of Gaussianity of sEMG changed as the maximum voluntary contraction (MVC) varied. The signal became less Gaussian at very low and very high MVC, but some were in the middle, sEMG became more Gaussian. The level of non-Gaussianity of EMG signal variation was used to differentiate healthy and pathological EMG signals.[4] Also, F. Meziani et al applied bispectral analysis to follow the pathological severity in two patients and compared their results with a healthy patient, they candidated the bispectrum in discriminating unhealthy and healthy conditions.[5] While other researchers went directly to a time-frequency analysis of this signal by applying CWT as method that quantifies temporal changes of the frequency content of non-stationary signals without losing resolution in time or frequency.[6] CWT show better performing in extracting indices in the time-frequency domain.[7] In reference 8, mother wavelet Daubechies of 4^{th} order with 6^{th} levels of decomposition (db4) has been chosen to implement the wavelet transform, because of its suitability for detecting signal changes and due to its shape, which is similar to the shape of motor unit action potentials.[9-11]

Separating healthy and unhealthy conditions and evaluating the capability of time and frequency domain features is an important issue of many researchers, three popular and widely used classifiers namely Radial basis function network (RBFN), k-nearest neighbor (k-NN) and support vector machine (SVM) were implemented, SVM is declared as a better classifier compared to RBFN and k-NN applying four features: the root mean square, logarithmic root mean square (RMS and logRMS), the centroid of frequency and standard deviation (Std).[12] Also, J. Too et al have applied linear discriminant

© 2020 A. Mokdad, S.M. Debbal, F. Meziani. This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

analysis (LDA) classifier using three selecting features; the mean absolute value, the root mean square (RMS), and wavelength (WL) showed the best classification performance in EMG signal classification application.[13] These results allowed us to implement the SVM, LDA, and KNN in this work.

Our proposed study is to distinguish neuropathy and healthy characteristics applying the bispectrum and CWT of sixteen patients and select the suitable method that indicated the presence of pathology and follow its severity. Then, the study has been reached because of ALS's severity that allowed us to separate healthy and ALS disorder, the performance of time and frequency feature domains are investigated using respectively four features: the entropy, the standard deviation, mean and the median frequency (EP, Std, MNF, MDF). For an optimal evaluation, three popular machine learning algorithms namely SVM, LDA, and KNN are utilized. Finally, the performance of EMG pattern recognition is discussed in order to separate the two cases and select the most appropriate classifier.

Materials and Methods

Database

The material consisted of a normal control group and a group of patients with ALS. The control group consisted of 10 normal subjects aged (2 -37) years, 4 females and 6 males. 6 out of 10 were in very good physical shape, and the remaining except one was in generally good shape. None in the control group had signs or history of neuromuscular disorders. The ALS group consisted of 8 patients; 4 females and 4 males aged 35-67 years. Besides clinical and electrophysiological signs compatible with ALS, 5 of them died within a few years after the onset of the disorder, supporting the diagnosis of ALS. The brachial biceps muscles were used in this study because they were the most frequently investigated in the two patient groups. The EMG signals were recorded from biceps brachii muscle by concentric needle electrode and amplified by an instrumentation amplifier DISA15C01 with a gain of 500. An analog bandpass filter with cut-off frequencies of 2Hz to 10 kHz was used. These analog signals were digitalized by Motorola DSP56ADC16 with 16- bit resolution and sampled at a rate of 23438 Hz.[14]

Analysis of Normal and Pathological EMG Signals

Bispectrum Analysis

In many biomedical applications, bispectrum analysis was performed on the signal of interest. Often the plot of the bispectrum enables us to differentiate different physiological states or pathological conditions.[15]

The bispectrum is the 2D-Fourier transform of the third cumulant function defined as:

$$S_2^x(\omega_1,\omega_2) = \sum_{\tau_1=-\infty}^{\infty} = \sum_{\tau_{2=-\infty}}^{\infty} = \\ = \mathrm{cum}_3^x(\tau_1,\tau_2)\mathrm{Wexp}\,[-j\omega_1\tau_1 + \omega_2\tau_2] \quad \text{Eq. 1}$$

for $|\omega 1| \le \pi$, $|\omega 2| \le \pi$, and $|\omega 1 + \omega 2| \le \pi$

where $W(\tau_1,\tau_2)$ is the 2-dimensional window function that decreases the variance of the bispectrum. In this study, a Hanning window was used, cum_3^x is the 3^{rd} order cumulant.

The 3rd order cumulant of a discrete signal $x(k)$, which is stationary and has a 0 mean, is defined as:

$$\mathrm{cum}_3^x(\tau_1,\tau_2) = \mathrm{cum}\{x(k)x(k+\tau_1)x(k+\tau_2)\} = \\ = \langle x(k)x(k+\tau_1)x(k+\tau_2)\rangle - \langle x(k)\rangle\{\langle x(k)x(k+\tau_1)\rangle + \\ \langle x(k)x(k+\tau_2)\rangle + \langle x(k+\tau_1)x(k+\tau_2)\rangle\} + 2\langle x(k)\rangle^3 \quad \text{Eq. 2}$$

Where $\langle\cdot\rangle$ denotes the expected process, $x(k)$ discrete signal Cumulants can be expressed in terms of moments and vice versa. One can easily calculate cumulants as certain non-linear combinations of moments.

Cumulants can be expressed in terms of moments and vice versa. One can easily calculate cumulants as certain non-linear combinations of moments.

The second and third-order cumulants are:

$$\mathrm{Cum}_1^x = m_1^x \quad \text{Eq. 3}$$

$$\mathrm{Cum}_2^x(\tau) = m_2^x(\tau) - (m_1^x)^2 \quad \text{Eq. 4}$$

$$\mathrm{Cum}_3^x(\tau_1,\tau_2) = m_3^x(\tau_1,\tau_2) - m_1^x[m_2^x(\tau_1) + m_2^x(\tau_2) + \\ m_2^x(\tau_1-\tau_2)] + 2(m_1^x)^3 \quad \text{Eq. 5}$$

The estimate of the bispectrum of a stationary and ergodic random process with the non-parametric approach is given by:

$$B(f_1,f_2) = \langle X(\omega_1)X(\omega_2)X^*(\omega_1+\omega_2)\rangle \quad \text{Eq. 6}$$

where: X(f) is the Fourier transform of a segment (or windowed portion), * denotes a complex conjugate, $\langle\cdot\rangle$ denotes the expected process.

The bispectrum is a function of two frequencies, unlike the power spectrum which is a function of one frequency variable. The frequency f may be normalized by the Nyquist frequency (one half of the sampling frequency) to be between 0 and 1. The bispectrum can be normalized (by power spectra at component frequencies) such that it has a magnitude between 0 to 1and indicates the degree of phase coupling between frequency components.[16,17]

CWT Analysis

In this study, starting from pre-processed EMG signals s(t), we evaluated the Continues Wavelet Transform (CWT) defined as:

$$\mathrm{CWT}\,(a,\tau) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} s(t)\,W^*\left(\frac{t-\tau}{a}\right)dt, a \neq 0 \quad \text{Eq. 7}$$

Where w(t)is a prototype function called mother wavelet, $\boldsymbol{\tau}$ is the translation index, and $\boldsymbol{a}$ is the scale parameter related to the frequency content. Indeed, the CWT may be a useful tool for detecting the presence of muscle activity in EMG signals. In this work, mother wavelet Daubechies of 4^{th} order with 6^{th} levels of decomposition (db4) has been chosen to implement the wavelet transform, CWT decomposes a signal into several

multiresolution components (coefficients) and performs a series of high and low-pass filter operations followed by down-sampling. Thus, the sEMG signal was decomposed into its frequency content form and then was reconstructed.[10]

In principle, this study includes temporal features as the entropy and Std likely to detect any variation of the analyzed EMG signal with respect to the type of pathology following to the extracted parameters and frequency features as the mean and median frequency.

Entropy

The aim set through this parameter is defining the amount of information carried by the signal known as a quantitative measure of disorder, it comprises the calculation of the probability density function. In this work, the entropy is used as feature, it defined by the following formula:

$$\mathrm{Ep}[S] = -\int \mathrm{p}(S)\log\ \mathrm{p}(S)\,\mathrm{d}(S) \qquad \text{Eq. 8}$$

S: continuous aleatory variable: the EMG signal; p (S): Probability density.

Standard Deviation (Std)

The standard deviation is a measure used in digital signal processing to characterize the signal variance. It indicates the dispersion of the values relative to the average, represented by the following formula:

$$\mathrm{Std}(S) = \sqrt{\frac{\sum_{i=1}^{N}(S_i - S_{moy})^2}{N}} \qquad \text{Eq. 9}$$

With S_moy: the average signal; N: number of samples.[12]

Mean and Median Frequency

Two frequency indicators, the mean frequency (MNF) and the median frequency (MDF), are often derived from the Fourier transform. In this work, the MNF and MDF are used as features and expressed as:

$$\mathrm{MNF} = \sum_{j=1}^{L} f_j\, p_j \Big/ \sum_{j=1}^{L} p_j \qquad \text{Eq. 10}$$

$$\sum_{j=1}^{MDF} p_i = \sum_{j=MDF}^{L} p_i = \frac{1}{2}\sum_{j=1}^{L} p_i \qquad \text{Eq. 11}$$

where fj is the frequency value of EMG power spectrum at the frequency bin j, Pj is the EMG power spectrum at the frequency bin j, and L is the length of frequency bin. In the analysis of EMG signal, the MDF and MPF are the two indicators most often used to characterize pathological cases of EMG signal and muscles fatigue.[18,19]

Classification of EMG Signal

To evaluate the capability of time and frequency domain features in differentiating normal and ALS conditions, machine learning algorithms are employed. In this study, three popular and widely used classifiers namely SVM, LDA, and KNN are implemented to EMG signal.

Support vector machine (SVM) has been recognized to be one of the best and efficient machine learning algorithms in EMG pattern recognition.[20] Radial basis function (RBF), linear, polynomial, and Gaussian function are commonly used in SVM. Concerning the classification accuracy, SVM offers a better performance.[12]

Linear discriminate analysis (LDA) is a statistical classification method that covers the boundary points and the different data points lie on the hyperplane. It is a common classification method used in EMG pattern recognition.[21] In the LDA classifier, the number of training samples must much lesser than the number of features and therefore the intraclass scatter matrix would tend to be a singular matrix and the LDA computational will be so demanding.

K-Nearest Neighbor (KNN) is a simple algorithm based on proximity measure. The predicted class of testing data depends on the majority votes of k nearest neighbors measured by Euclidean distance. In this algorithm, the choice of k value is crucial for reducing errors as well as improving the performance; setting k=10.[22]

The proper classifiers are selected, the data is divided into training and testing sets for performance evaluation. In this work, a six-fold cross-validation method is used. Cross-validation is a popular statistical technique that has been widely used in classification and regression.[23] Cross-validation aims to test the whole data set by partitioning the data equally into six parts. Each part is used for testing in succession while the rest is used for a training session.

The classification's performance was evaluated by Accuracy (Acc), Precision (Pr), and Specificity (Sp), defined as:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{Eq. 12}$$

$$Pr = \frac{TP}{TP+FP} \qquad \text{Eq. 13}$$

$$Sp = \frac{TN}{TN+FP} \qquad \text{Eq. 14}$$

Where, TP: true positive, TN: true negative, FP: false positive, FN: false negative.

Results and Discussion

Figure 1 represents the bispectrum of **(a)** normal and **(b)** ALS disorder which showed that the level of gaussianity of EMG changed from normal and neuropathy, in general, the signal became less Gaussian at very low and very high frequencies, but some were in the middle, EMG becoming more Gaussian. The level of the non-Gaussianity of the EMG signal variation allowed us to follow the pathological severity, the nonlinearity and non-Gaussianity concentration of normal (**Figure 1a**), are well centered at lower frequencies than neuropathy activities (**Figure 1b**). This allowed us to conclude that neuropathy is the most severe because of the level of the non-Gaussianity of the EMG signal which tends towards high frequencies. Therefore, to qualify the bispectrum to be a method of analysis the most appropriate that value the degree of pathological severity and agreeable visualization of non-Gaussianity towards high frequencies indicated more accentuated pathological severity.

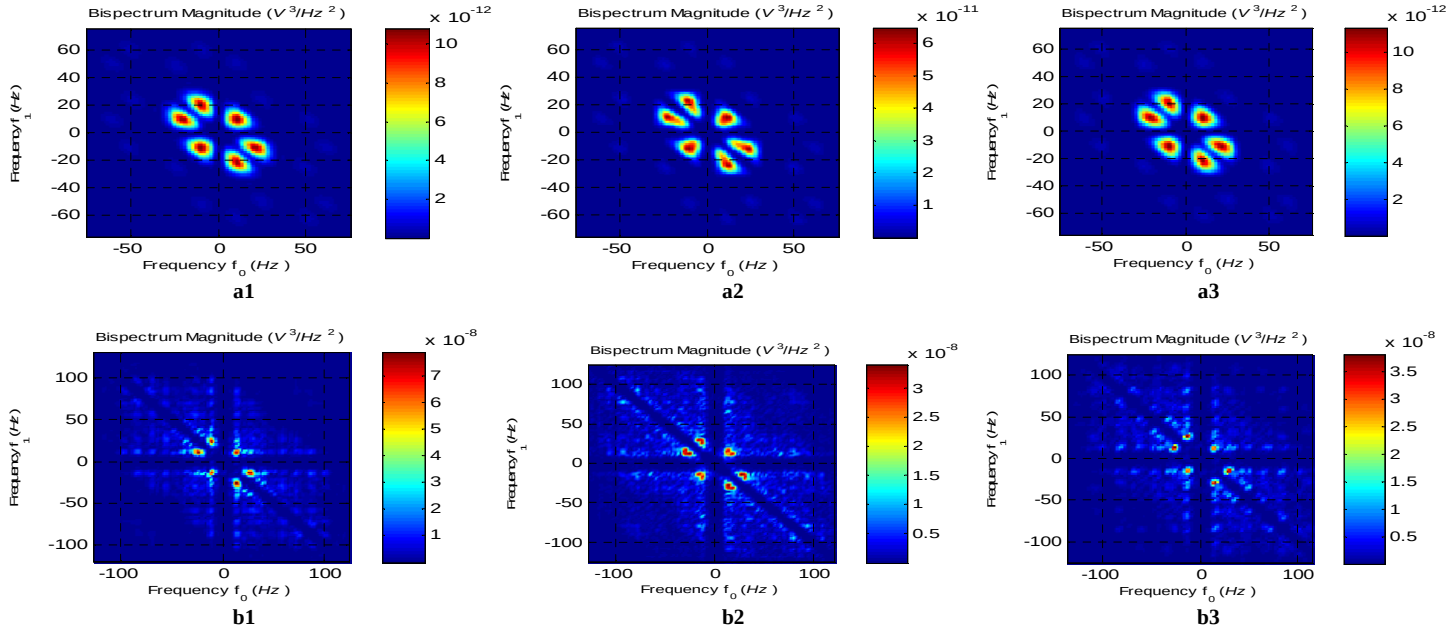


Figure 1. Bispectrum of: (a) healthy and (b) ALS disorders.

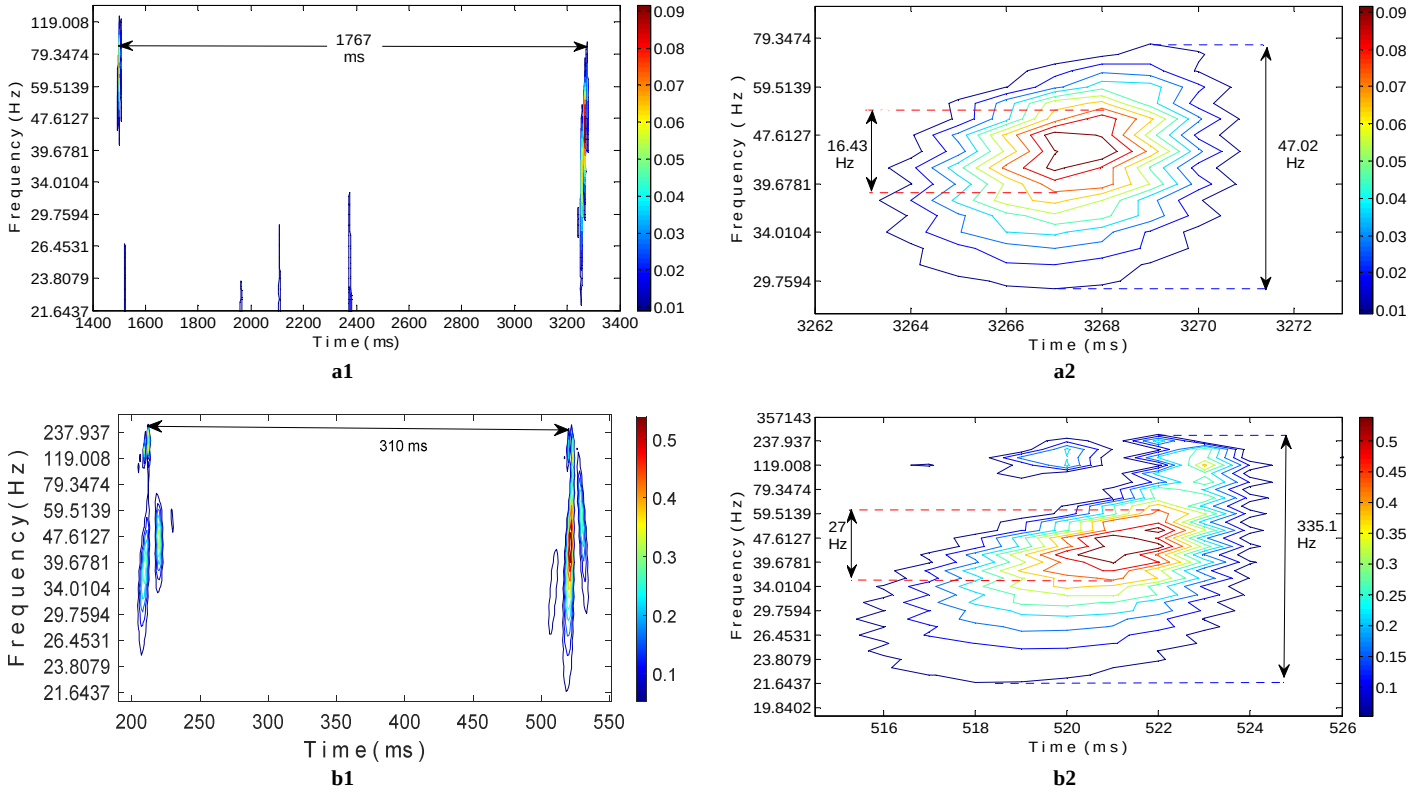


Figure 2. Scalogram percentage of energy for each wavelet of :(a) healthy, (b) ALS disorder.

A suitable comparison of the neuropathy compared to healthy cases by a bispectral analysis indicated a slight increase of the bispectral magnitude in the neuropathic cases which reached the 8×10^{-8}. In healthy conditions, the magnitude value ranges from 6.5×10^{-8} to 11×10^{-12}, which allowed us to track the evolution of pathology by extraction of the bispectral magnitude and to candidate the bispectrum analysis to be reliable for determining the severity of pathology while keeping the interval of the beginning and the end of its bispectral magnitude of these two pathologies.

Figure 2 shows the time-frequency representation of EMG signals obtained from the time-frequency methods, namely, CWT using to address the goal of determining variations in biceps brachii muscle of patients which are dependent on healthy and unhealthy conditions. In this study, the analysis is made by calculating the time interval between the two peaks in the 2 cases, it is represented by **Figures 2a(1)**, **2b(1)**. The

extraction of the most energetic contour from these figures allowed us to extract a corresponding parameter. **Figures 2a(2)**, **2b(2)** represent the calculated parameter that serves to differentiate these two cases, it is the mean frequency of the muscular activity that evolved progressively from the healthy patient towards the other which has the neuropathy, the same parameter was used for the study of the most and least energetic contours percentage and to confirm again the interest of this parameter in the normal case analysis and even pathological cases because of the existence of a correlation between this already calculated parameter and the severity of the pathology. Thus, to consider them as a basic indicator for assessing the degree of pathological severity, the study of the maximum energy percentage of these contours confirmed that the neuropathy with the highest percentage of energy evolved from 10% to 55%, and to consider them as the most severe with a mean percentage of 32.5%, the healthy contours percentage reached till 9% starting with 1% with a mean percentage of 5%.

The analysis of the EMG signal's contours using the Daubechies continuous wavelet of 4^{th} order has been used extensively to evaluate the degree of pathological severity, and even to extract the rate of its evolution by a study of energy percentage of the averaged frequencies.

From **Figure 2**, we noticed evidently that the average frequency of the whole contour (the least energetic) is 47 Hz with the energy percentage of 1%, and 16 Hz for a contour of 8%, the energy percentage in healthy patients can reach up to 9%.

The analysis of the neuropathy contours is used to evaluate the pathological degree's evolution, that means that a percentage of 10% of the complete contour having an average frequency of 335 Hz and 27 Hz for a percentage of 55% of the most energetic contour and finally to infer that neuropathy spreads with an energy percentage of 55% and it's already an advanced stage through the studied cases.

And here is the interest of the analysis by the continuous wavelet is to evaluate exactly the degree of the pathological severity by the calculation of energy percentage in healthy and ALS patients, these result qualified this analysis to be the most effective not only for the analysis of normal cases and even pathological cases because it served to follow the pathology evolution's level which was inaccessible by applying the bispectrum where it just used to locate the level of gaussianity of the EMG signal by calculation the bispectral magnitude.

Table 1. Performance comparison of time and frequency feature's extraction domain.

Full Classification of features	Accuracy	Precision	Specificity
SVM	95.8%	92.59%	92%
LDA	91.5%	85.4%	83%
KNN	92%	92%	92%

The following section consist to separate the data into healthy and neuropathy conditions, the results of the performance features are discussed. In the experiment, four features are extracted (EP, Std, MNF, MDF) from time and frequency domain to classify the EMG signal into normal and ALS classes. The classification performance was compared by using the SVM, LDA, KNN classifier.

As shown in **Table 1**, the proposed method with SVM classifier yielded a higher accuracy (Acc=95.8%) than respectively KNN and LDA classifier that achieved an accuracy of (Acc =92%, precision of 92%, specificity of 92% and Acc=91.5%, precision of 85.4%, specificity of 83%). Also, this proposed extracted feature achieved a higher performance with SVM, KNN, and LDA that qualify the selected features as the most powerful in separating healthy and unhealthy conditions (see **Table 1**).

Conclusion

The use of bifrequency and time-frequency analysis on EMG signals using respectively the bispectrum and CWT has effectively discriminated neuropathy from healthy patients. Based on these results, it can be mentioned that the CWT had good resolution and high performance for visualization of the neuropathy activity because it's not just given differentiation between healthy and pathological cases but also its evolution's degree which was inaccessible by the bispectrum analysis and at the end to qualify the CWT to be the successful analysis methods. Results using time and frequency feature domains were able to discriminate healthy and unhealthy conditions. Also, it shows that a patient with neuropathy can be successfully separated from healthy; this is an important fact applying three popular classifiers (SVM, LDA, and KNN), ours results qualify SVM classifier because it yielded a higher accuracy of 95.8% and has increased the total capability for obtaining high accuracy in diagnosis.

References

1. Rissanen SM, Kankaanpaa M, Tarvainen MP, et al. Extraction of Typical Features from Surface EMG Signals in Paarkinson's Disease. 11th International Congress of Parkinsons's Disease and Movement Disorders, Istanbul, Turkey, 2007.
2. Rowland LP. Amyotrophic Lateral Sclerosis: Theories and Therapies. Ann. Neurol. 1994;35(2):129-130. doi: 1O.1002/ana.410350202.

3. Kaplanis PA, Pattichis CS, Hadjileontiadis LJ, Panas SM. Bispectral analysis ofsurface EMG. In: Proceedings of the 10th Mediterranean Electrotechnical Conference, vol. II. p. 770-773, 2000.
4. Nazarpour K, Sharafat AR, Firoozabadi SMP. Application of higher order statisticsto surface electromyogram signal classification. IEEE Trans Biomed Eng. 2007;54(10):1762-1769.
5. Meziani F, Rerbal S, Debbal SM. Spectro-temporal analysis of electromyogram signals (EMGs). Int J Med Eng Inf. 2019;22(2).
6. Raez MB, Hussain MS, Mohd-Yasin F. Techniques of EMG signal analysis: detection, processing, classification and applications. Biol Proced Online. 2006;8:11-35.
7. Phinyomark A, Phukpattaranont P, Limsakul C. Feature Reduction and Selection for EMG Signal Classification. Expert Systems with Applications. 2012;39(8):7420-7431.
8. Strazza A, Verdini F, Burattini L, et al. Time-frequency Analysis of Surface EMG signals for Maximum Energy Localization during Walking. In: Eskola H., Väisänen O., Viik J., Hyttinen J. (eds) EMBEC & NBC 2017. IFMBE Proceedings. 2018;65.
9. Farge M. Wavelet Transforms and their Applications to Turbulence. Ann Rev Fluid Mech. 1992;24:395-457.
10. Strazza A, Verdini F, Burattini L, et al. A Time-Frequency Approach for the Assessment of Dynamic Muscle Co-contractions. IFMBE Proceedings. 2019;68/2:223-226
11. Ismail AR, Asfour SS. Continuous Wavelet Transform Application to EMG Signals During Human Gait. Conference Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers (Cat. No.98CH36284), Pacific Grove, CA, 1998, pp. 325-329.
12. Mishra B, Wadhwani AK, Singh S. EMG Signal Classification for Neuromuscular Disorder using Soft-Computing Techniques. IJIRMPS. 2019;7(1):24-27.
13. J Too, Abdullah AR, Tengku Zawawi TNS, et al. Classification of EMG Signal Based on Time Domain and Frequency Domain Features International Journal of Human and Technology Interaction. 2017;1(1):25-29.
14. Nikolic M. Detailed analysis of clinical electromyography signals: Emg decomposition, findings and firing pattern analysis in controls and patients with myopathy and amytrophic lateral sclerosis. Ph.D. dissertation, 2001. Online. Available: http://www.emglab.net
15. Chua KC, Chandran V, Acharyaa UR, Lima CM. Application of higher order statistics/spectra in biomedical signals. Med Eng Phys. 2010;32(7):679-689.
16. Nikias CL, Raghuveer MR. Bispectrum estimation a digital signal processing framework. Proceedings of the IEEE. 1987;75(7):869-891.
17. Nikias CL, Petropulu AP. Higher-order spectra analysis: a nonlinear Signal Processing Framework. Prentice-Hall, Englewood Cliffs, NJ; 1993.
18. Mishra A, Mishra V, Yadav VK. Comparison of normal and pathological gait using EMG signal. Int J Adv Res Sci Eng. 2017;6(2):442-451.
19. Cabrera C. Analyse du signal myoélectrique pour l'évaluation de la fatigue périphérique chez des nageurs de haut niveau en demi-fond. 2014.
20. Karthick PA, Ghosh DM, Ramakrishnan S. Surface electromyography-based muscle fatigue detection using high-resolution time-frequency methods and machine learning algorithms. Comput Methods Programs Biomed. 2018;154:45-56.
21. Khan S, Hussain M, Aboalsamh H, Bebis G. A comparison of different Gabor feature extraction approaches for mass classification in mammography. Multimedia Tools and Applications. 2017;76(1):33-57.
22. Benazzouz A, Guilal R, Amirouche F, et al. EMG Feature Selection for Diagnosis of Neuromuscular Disorders. 2019 International Conference on Networking and Advanced Systems (ICNAS). 2019.
23. Xing K., Yang P, Huang J, et al. A real-time EMG pattern recognition method for virtual myoelectric hand control. Neurocomputing. 2014;136:345-355.