Konya Mühendislik Bilimleri Dergisi, c. 11, s. 1, 205-219, 2023
Konya Journal of Engineering Sciences, v. 11, n. 1, 205-219, 2023
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1185629

PREDICTION OF CERVICAL DISC HERNIATION DISEASE UTILIZING TRAPEZIUS sEMG SIGNALS WITH MACHINE LEARNING TECHNIQUES BASED ON FREQUENCY DOMAIN FEATURE EXTRACTION

[1]Burak YILMAZ , [2*]Güzin ÖZMEN , [3]Ahmet Hakan EKMEKÇİ

[1] KTO Karatay University, Electrical-Electronics Engineering Department, Konya, TÜRKİYE
[2]Selcuk University, Technology Faculty, Biomedical Engineering Department, Konya, TÜRKİYE
[3] Selcuk University, Medicine Faculty, Neurology Department, Konya, TÜRKİYE
[1]burak.yilmaz@karatay.edu.tr, [2]gozmen@selcuk.edu.tr, [3]hakanekmekci@selcuk.edu.tr

Highlights

- We proposed a method using machine learning to investigate the classification and prediction of CDH disease with resting-state trapezius sEMG.
- We used Savitsky-Golay and Butterworth filters for denoising and then obtained PSD-based features using the Burg method.
- The best classification accuracy of 91.6% was achieved with 10-fold cross-validation using the Tree classifier, moreover, Neural Networks and CN2 rule inducer provided 87.5% classification accuracy for the prediction of CDH disease.

***Corresponding Author:** Güzin ÖZMEN, gozmen@selcuk.edu.tr

Konya Mühendislik Bilimleri Dergisi, c. 11, s. 1, 205-219, 2023
Konya Journal of Engineering Sciences, v. 11, n. 1, 205-219, 2023
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1185629

PREDICTION OF CERVICAL DISC HERNIATION DISEASE UTILIZING TRAPEZIUS sEMG SIGNALS WITH MACHINE LEARNING TECHNIQUES BASED ON FREQUENCY DOMAIN FEATURE EXTRACTION

[1]Burak YILMAZ , [2*]Güzin ÖZMEN , [3]Ahmet Hakan EKMEKÇİ

[1] *KTO Karatay University, Electrical-Electronics Engineering Department, Konya, TÜRKİYE*
[2]*Selcuk University, Technology Faculty, Biomedical Engineering Department, Konya, TÜRKİYE*
[3] *Selcuk University, Medicine Faculty, Neurology Department, Konya, TÜRKİYE*
[1]burak.yilmaz@karatay.edu.tr, [2]gozmen@selcuk.edu.tr, [3]hakanekmekci@selcuk.edu.tr

(Received: 07.10.2022; Accepted in Revised Form: 19.12.2022)

ABSTRACT: Cervical disk herniation (CDH) is a disease that affects the quality of life of many people due to the neck pain it causes. The aim of this study was to develop an automatic prediction system to aid in diagnosis by evaluating the change in the surface electrical activity of the trapezius muscle in SDH disease in order to find an answer to the question: 'Can the surface electromyogram (sEMG) recorded from the trapezius muscle be an effective indicator for the diagnosis of SDH disease?'. To this end, a dataset will be created using preprocessing and feature extraction methods from sEMG signals from CDH patients and healthy individuals. In the first step, the Savitsky-Golay filter is used to denoise the sEMG signals and the dominant frequency signals between 20 and 150 Hz are included in the study using the Butterworth filter design. Twenty PSD-based features in the frequency domain were then obtained from the signals to which we applied the Burg method. Eleven of the most significant features based on the information gain, gain ratio, and Gini values are selected to be submitted to the classifiers. 80% of all new feature areas are used for classification and the rest for prediction. The best classification accuracy of 91.6% was obtained with the Tree classifier using 10-fold cross-validation for classification. In addition, neural networks and CN2 rule inducer provided 87.5% classification accuracy for prediction using 20% of the remaining data that the classifiers had not seen before. The experimental results demonstrate that the trapezius muscle has different surface electrical activity in CDH patients and healthy subjects and that the frequency domain characteristics of this activity are important for disease prediction.

Keywords: *EMG, Machine Learning, Classification, Prediction, Frequency Domain*

1. INTRODUCTION

Cervical disk herniation (CDH) is a condition that causes compression of the spinal cord or nerve roots between the C5-C6 and C6-C7 vertebrae. CDH usually causes pain radiating to the upper extremities and paresthesias felt on the skin [1-3]. CDH is usually characterized by arm and neck pain, and neurologists make the diagnosis through physical examination, imaging, and electrodiagnostic tests such as electromyography (EMG). MRI is the main diagnostic method that guides physicians in diagnosing CDH, but it is a time-consuming and expensive procedure. On the other hand, the needle EMG method is also used, which provides a faster solution in clinical evaluation. However, needle EMG is invasive and may cause various complications. The alternative to needle EMG is surface EMG (sEMG), which allows the measurement of total muscle action potentials on the surface. It is not preferred in clinical assessments because it involves superficial muscles. In this study, we aimed to investigate the usability of sEMG in diagnosing CDH patients for clinical and biomedical applications. Based on extensive anatomical preparations, [4] have described more than twenty muscles involved in head and neck movements. Unfortunately, few of them are superficial enough to be reached with surface electrodes [5]. Following the

***Corresponding Author:** Güzin ÖZMEN, gozmen@selcuk.edu.tr

physical model developed for the neck muscles, Bernhardt et al. [6] identified four muscles in the cervical spine that can be reached by superficial EMG: Semispinalis Capitis, Splenius Capitis, Sternocleidomastoid (SCM), and Trapezius. The trapezius muscle is the most painful muscle due to acute trauma and occupational myalgias. Many studies have emphasized that the trapezius muscle does not contribute to head and neck movements, but when the arms are actively used, this muscle should be examined to monitor the electrical activity of the adjacent muscles [7].

The electromyogram is defined as the graphical representation of the electrical activity that occurs in resting and contracted muscles and provides important information for the diagnosis of abnormalities in both muscles and the motor system [8]. The sEMG is a complex, unstable, and noisy signal. The amplitude of the sEMG signal is arbitrary and can usually be expressed as a Gaussian distribution. The amplitude range of the signal is 0-10 mV (peak-to-peak) or 0-1.5 mV (RMS). The useful energy of the sEMG is in the frequency range of 0-500 Hz, but the dominant energy is in the range of 50-150 Hz [9].

In the literature, some studies classify neuromuscular diseases [10-12], detect muscle activity [13-14], muscle fatigue [15-16], and classification of low back pain [17-18] and neck pain [19-21] using sEMG. To our knowledge, no study classifies CDH patients using surface EMG, except [22]. In addition, some studies draw attention to the upper trapezius muscle during semi-static activities that require repetitive movements of the upper extremity. This is the most common area for muscle pain. This pain sometimes indicates chronic trapezius myalgia or tension neck syndrome [7].

In studies of sEMG or EMG signals, there are several approaches to generating features from the signals. Generally, three approaches are preferred for feature generation, namely time domain, frequency domain, and time-frequency representation [23-24]. Phinyomark et al. evaluated 37 different features based on both time-domain and frequency-domain, and found that the frequency-domain features of sEMG signals are not more redundant than those of the time-domain [23]. However, in the experimental setup of this study, a specific movement was not repeated, but the corresponding muscle performed the task of maintaining a specific weight at a constant height. Therefore, using the frequency domain approach, it was easier to detect the presence of a pain signal associated with the frequency of the signal generated in the muscle in the frequency domain.

This study attempted to develop a predictive model for the detection of CDH disease by analyzing trapezius muscle sEMG data. For this purpose, the study focused on frequency domain features of trapezius muscle signals. The features were tested using classification algorithms. After the designed experimental study, it was found that a clear classification can be made between healthy subjects and subjects suffered from cervical hernia, and it is also possible to develop predictive models for the disease using these models.

There are many current studies in the literature for feature extraction and classification from EMG signals. These methods differ according to experiment design and selected muscle group and, high classification accuracy is obtained. EMG classifications are used in the automated diagnostic system for neuromuscular diseases or used for prosthetic device control. Although needle electrodes are preferred in neuromuscular disorders, there are also few studies using surface EMG. Some of them are presented in Table 1.

Table 1. Similar studies in the literature

		Authors	EMG sample	Feature Extraction	Machine Learning Method	Best Accuracy
Neuromuscular Disorders	Intramuscular EMG (iEMG)	Katsis et al. [25]	Normal, Neuropathic Myopathic	Raw EMG signals	SVM	93% 95% 92%
		Rasheed et al. [26]		Time domain Wavelet domain	Adaptive fuzzy k-NN	93.5% 92.6%
		Gokgoz and Subasi [27]	Normal, ALS, Myopathic	Music	k-NN, SVM, ANN	82.11% 92.55% 90.02%
		Kamali et al. [28]	Normal, Neuropathic Myopathic	Time domain Time-Frequency domain	SVM	97%
		Artameeyanant et al.[29]	Healthy, Myopathic, Amyotrophic	Statistical feature extraction	MLPNN, SVM, k-NN	98.36% 99.17%
		Hazarika et al. [30]	Normal, ALS, Myopathic	Discrete Wavelet Transform-Canonical correlation analysis	k-NN	98.8%
	Surface EMG (sEMG)	Istenič et al. [31]	Muscular Neuronal Disorder	Multiscale entropy	SVM	81.5%
		Barmpakos et al. [32]	Neuropathy Myopathy	Discrete Wavelet Transform, Power spectral density	Random Forest, K-NN	88.8%
Cervical Disc	Surface EMG (sEMG)	Ozmen and Ekmekci [22]	Normal CDH (Fatigue-state)	Short Time Fourier Transform Discrete Wavelet Transform AR model	ANN	99%

2. DATA SET AND DATA PREPARATION

The experimental design used in this study focused on the trapezius muscle and the relationship between the muscle and CDH disease. The trapezius muscle is defined by Moore et al. [33] as a "broad, apartment, superficial muscle extending from the cervical to the thoracic region on the posterior aspect of the neck and trunk" Because the trapezius muscle is superficial and directly connected to the cervical region, this study focused on the basic question: "Can the presence of CDH be detected from the sEMG signals of the upper trapezius muscle?" To answer this question, an experimental study was performed. Detailed information about the experimental study can be found in [22], but the present study only addresses the resting state data recorded from the trapezius muscle. In this prospective study, sEMG data were collected by a physician and a technician in the neurology department of Selcuk College Faculty of Medicine using surface electrodes and a Neoropack Nihon Kohden EMG device in 10 CDH patients (8 males and 2 females, aged 17 to 67 years) and 10 healthy volunteers (4 males and 6 females, aged 19 to 48 years). Participants were selected among right-handed individuals. This study was approved by the local ethics committee (decision no: 2010/33). Each participant completed the informed consent form. For the rest condition, participants were asked to wait 20 seconds without moving the right arm in an upright sitting position. The aim is to investigate the electrical activity of the trapezius muscle while standing in its natural position and to determine the difference in CDH patients.

In this study the raw data were recorded as unsigned 16-bit integers with a sampling frequency of 10 kHz. Thus, each data packet contained 200,000 samples per subject. The data were then segmented into 20 equal parts each. The goal of this division is to take samples of each signal with a duration of 1 second. Figure 1 presents the general flowchart of the entire process. According to the Figure 1 some preprocessing steps are applied to the segmented EMG data. Then the features used for ranking were extracted from the frequency domain features of the pre-processed data. After the ranking test, the most informative features are selected. Finally, the classification algorithms are trained and tested with the cross-validation test and the results are compared. The tested classifiers are Random Forest, Naive Bayes, CN2 rule inducer, Tree, Support Vector Machine (SVM), Neural Networks.

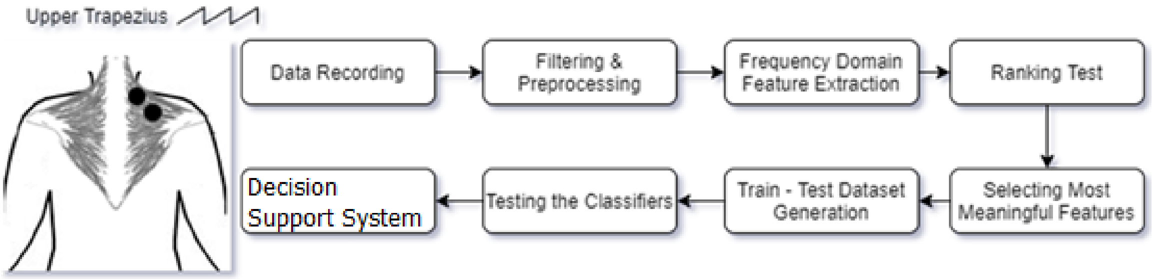

Figure 1. General flowchart of the entire process

Figure 2 presents the pre-processing steps. The first step of preprocessing is to normalize the data in the range of [-1,1]. In general, normalization is not necessary when power spectrum analysis is applied to the sEMG signals [34], but in this study, the experiment consists of no motion, so the signal contains low-frequency components with a high DC basis. The DC base signal was different for each subject. Therefore, to make each signal comparable, it was appropriate to normalize each signal. To separate valuable information from the signal noise, the second step was to apply the Savitsky-Golay (S-G) filter to the signal to smooth the data. Due to the nature of the Savitsky-Golay filter [35], the informative components of the data are preserved while the strong noise is suppressed. In this study, an 8^{th}-order S-G filter with a width of 1023 frames is used.

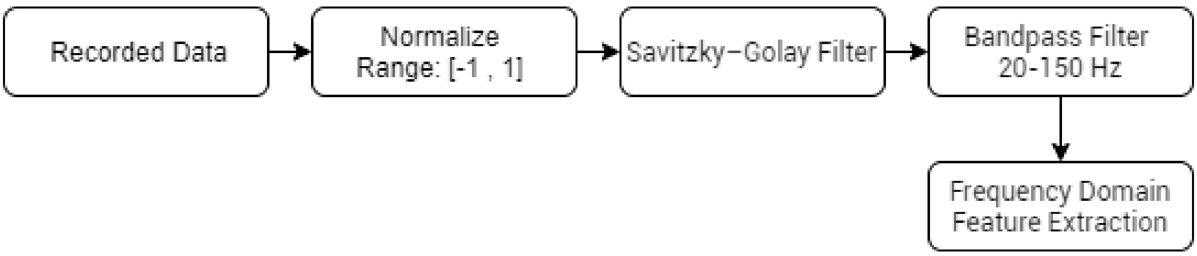

Figure 2. Flowchart of the preprocessing

The next step of preprocessing was bandpass filtering. The structure of the signal was a regular sEMG signal, so the current value of the signal depended on previous values. For this reason, a recursive filtering technique was preferred. In this case, a Butterworth-type filter was the filter of choice (Pauk 2008). Since the data used in this study did not include motion and very low and high frequencies were not meaningful, the passband of the filter was chosen to be in the range of (20 - 150) Hz

The final step of preprocessing is feature extraction from the signal. To extract the features from the signal, Power Spectrum Density (PSD) analysis was applied to the signal using Burg's method [36], so that the features are based on the frequency domain characteristics. In this study, the data length was shortened by dividing the 20-second data into 1-second epochs. Since the FFT is not suitable for determining the PSD in non-stationary signals such as EMG and non-parametric methods such as the periodogram-based Welch method are not preferred because they provide high resolution at long data lengths. Parametric spectral analysis methods such as the AR method are the most effective methods in analyzing PSD for short-time signals, and one of them, the Burg method, gives more stable results than the Yule-Walker and Cov methods [37]. Figure 3 shows the original and preprocessed EMG signal in each control's time and frequency domain.

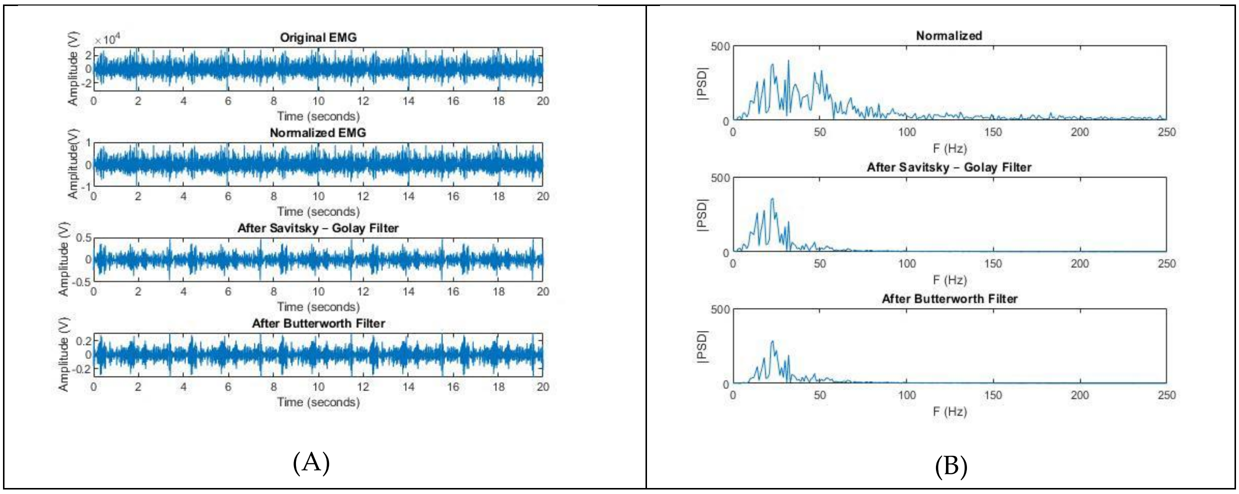

Figure 3. (A) Original and preprocessed EMG in the time domain (B) Original and preprocessed EMG in the frequency domain

3. FEATURE EXTRACTION

The PSD approach provides detailed information about the power distributions of each frequency in the analyzed frequency window. Figure 4 shows the boxplot of the peak frequency values of the sEMG signal for both normal and patient subjects.

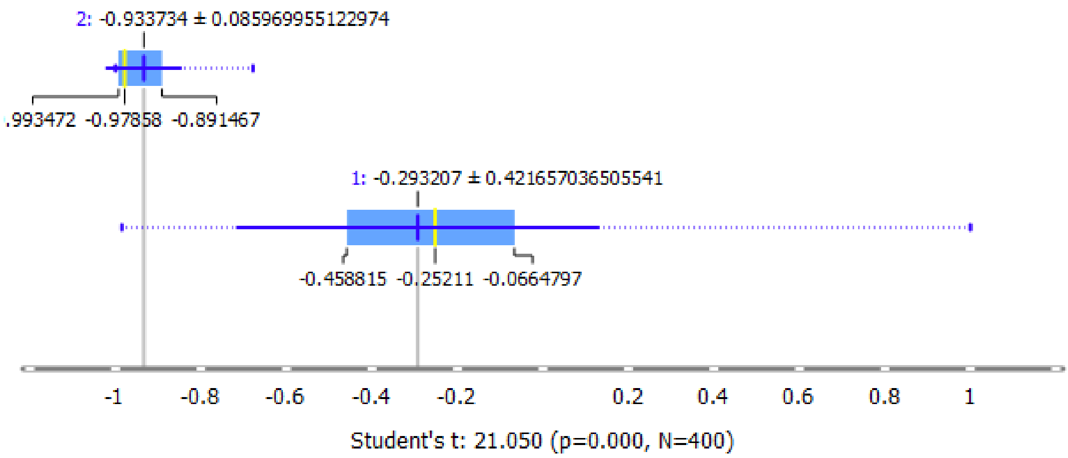

Figure 4. Box plot of the peak frequency distribution of the dataset

In Figure 4, class number 1 represents the subjects with CDH disease and 2 represents the normal subjects.

From the box plot data, it can be seen that there is a shift in peak frequency between healthy and CDH patients, while the PSD data provides more information. In this study, 19 different features were analyzed and selected as features for classification [38-39]. These features are listed in Table 2; the last, 20.

Table 2. Frequency Domain features used in this study.

Feature	Description	Feature	Description
AF1	The frequency at which the PSD of the sEMG signal is maximum	**W1**	Width of the dominant frequency.
AF2	The frequency at which the PSD of the sEMG signal is the second maximum	**W2**	Width of second dominant frequency
P1	The magnitude of Dominant Frequency F1	**iqrF**	The difference between the third quartile and the first quartile range of PSD
P2	The magnitude of the Second Dominant Frequency F2	**fF1**	Number of peaks under dominant frequency
QF1	Quality factor of F1 (QF=F1/wF2)	**fF2**	Number of peaks under the second dominant frequency
QF2	Quality factor of F2 (QF=F2/wF2)	**varF**	The variance of PSD of sEMG signal.
meaF	Mean of all peak values of PSD.	**kurtF**	Kurtosis of PSD
mea1F	Mean of all peak values under dominant frequency.	**skewF**	Skewness of PSD
mea2F	Mean of all peak values under second dominant frequency.	**medF**	Median of all peak values of PSD
Diag	Diagnosis feature (target)		

The ranking values of the PSD-based features of the sEMG signal and the top 10 features are determined based on the information gain, gain ratio, and Gini values calculated previously. Table 3 shows the main features selected for classification. The most important features are highlighted in the table with a darker background. Information gain was chosen as the main ranking method in selecting the characteristics, and 0.25 was used as the threshold. The features that are below the threshold of 0.25 are rejected.

Table 3. Selected features

	#	***Info. gain***	***Gain ratio***	***Gini***
medF	**1**	0.5210	0.2605	0.2644
iqrF	**2**	0.4521	0.2261	0.2425
Mean	**3**	0.4505	0.2253	0.2459
mea1F	**4**	0.4370	0.2185	0.2369
meaF	**5**	0.4370	0.2185	0.2369
varyF	**6**	0.4249	0.2124	0.2321
P1	**7**	0.4197	0.2098	0.2300
AF1	**8**	0.4197	0.2098	0.2300
mea2F	**9**	0.4051	0.2025	0.2238
fF1	**10**	0.2853	0.1561	0.1752
QF1	**11**	0.1604	0.0802	0.1046
P2	**12**	0.1266	0.0633	0.0826
skewF	**13**	0.1220	0.0610	0.0815
AF2	**14**	0.1097	0.0548	0.0729
kurtF	**15**	0.1066	0.0533	0.0717
W2	**16**	0.1058	0.0529	0.0701
QF2	**17**	0.0863	0.0432	0.0572
W1	**18**	0.0832	0.0416	0.0556
fF2	**19**	0.0499	0.0250	0.0337

Statistical information of the most significant features is also represented in Table 4. The entire dataset has no missing values.

Table 4. Statistical information of the features

Name	*Center*	*Dispersion*	*Min*	*Max*
fF1	-0.535	-0.816883	-1	1
mea2F	-1.17e-16	∞	-2.806e-16	2.806e-16
AF1	-0.61347	-0.720118	-1	1
P1	-0.671574	-0.592841	-1	1
varyF	-0.738759	-0.529157	-1	1
meaF	-0.475416	-1.17414	-1	1
mea1F	-0.475416	-1.17414	-1	1
Mean	-0.721069	-0.517747	-1	1
iqrF	-0.177202	-3.29478	-1	1
medF	-0.450080	-1.292014	-1	1

4. CLASSIFICATION AND PREDICTION

After turning the raw signals into a dataset, the dataset is separated into two parts. The first part consists of 80% of the entire data and is used to train the classifier algorithms. The second part which consists of 20% of the entire data is used for testing the classifiers after training. The classifiers which the dataset represented to are listed as Random Forest, Naive Bayes, CN2 rule inducer, Decision Tree, Support Vector Machine (SVM), and neural network algorithms [40-41]. During training, 10-fold cross-validation was applied to the classifiers. Classification algorithms require some predetermined parameters for the training process. Therefore, each classifier is developed with a specific architecture with its characteristics. Decision trees also called C4.5, use a divide-and-conquer algorithm to generate an initial tree. It requires a certain set S of cases. The dataset can be discrete or continuous [40]. Naïve Bayes is a supervised learning method that is very easy to construct. The method does not require complex parameters or schemes. It can be easily used with huge datasets [40]. Due to the nature of the algorithm, the Naive Bayes algorithm did not require extra parameters in this study. CN2 algorithm based on ID3 algorithm like the Decision Trees. According to the definition algorithm, "CN2 produces an ordered list of if-then rules" [41]. Random forest is a tree-based algorithm. The algorithm uses several tree predictors. Each tree is represented with a random vector which is independently taken from the same distribution in the forest. The algorithm has a generalization error limit. Equal tree affects the generalization error. Hence when the trees in the forest reach a result, they are voted for the most possible class [42]. SVM works on hyperplanes which are high- or infinite-dimensional space [43]. It requires fewer samples to train, is sensitive to dimensionality, and is considered a very robust algorithm for classification tasks [40]. Neural networks or Artificial Neural Networks (ANN) algorithms are based on interconnected nodes of computation units that mimic real-life neurons. In the architectural design, layers are created by nodes and each node is connected to the previous and next layers with weight coefficients. Training the ANN means finding the proper weight value for each node that performs the classification task. The design parameters of each algorithm are explained in Table 5.

Table 5. Classifiers' Design parameters

Decision Tree: **Tree size:** 7 nodes, 4 leaves **Edge widths:** Relative to parent **Target class:** None **Pruning:** at least two instances in leaves, at least five instances in internal nodes, maximum depth 100 **Splitting:** Stop splitting when the majority reaches 100% (classification only)	*CN2 Rule Inducer:* **Rule ordering:** unordered **Covering algorithm:** exclusive **Gamma:** 0.7 **Evaluation measure:** Laplace **Beam width:** 6 **Minimum rule coverage:** 2 **Maximum rule length:** 5 **Default alpha:** 1.0 **Parent alpha:** 1.0	*Random Forest:* **Number of trees:** 14 **Maximal number of considered features:** unlimited **Replicable training:** No **Maximal tree depth:** unlimited **Stop splitting nodes with maximum instances:** 5
Support Vector Machine **SVM type:** SVM, C=0.8, ε=0.1 **Kernel:** Polynomial, $(\text{auto}\ x \cdot y + 0.0)^{3.5}$ **Numerical tolerance:** 0.001 **Iteration limit:** 100	*Neural Network* **Hidden layers:** 20, 20 (2 hidden layers with 20 nodes each) **Activation:** tanh **Solver:** Adam **Alpha:** 0.0001 **Max iterations:** 200 **Replicable training:** True	*Naïve Bayes:* The naive Bayes algorithm did not require extra parameters in this study

5. RESULTS and DISCUSSION

The classification algorithms are trained and tested with the Stratified 10-fold Cross validation test, the average over classes results is shown in Table 6. The metrics used to evaluate the models are listed as Area Under Curve (AUC), Classification Accuracy (CA), F1, Precision, and Recall [44].

Table 6. Testing results of algorithms via 10-Fold cross-validation.

Model	*AUC*	*CA*	*F1*	*Precision*	*Recall*
Random Forest	**0,9643**	**0,9333**	**0,9334**	**0,9344**	**0,9333**
Tree	**0,9266**	**0,9333**	**0,9332**	**0,9340**	**0,9333**
Neural Network	**0,9568**	**0,9222**	**0,9223**	**0,9225**	**0,9222**
AdaBoost	0,9107	0,9111	0,9111	0,9111	0,9111
SVM	0,9464	0,8889	0,8890	0,8926	0,8889
CN2 rule inducer	0,9501	0,8667	0,8655	0,8722	0,8667
Naive Bayes	0,9375	0,8667	0,8667	0,8667	0,8667

During the cross-validation, only 80% of the entire data set is used for both training and testing. The rest of the data was never represented to the classifiers until the training of the models finished. After the training, the rest of the data is presented to the trained classifiers, and predicted results are compared with the actual results. The prediction results of the classifiers are shown in Table 7.

Table 7. Prediction results of the classifiers

Model	*AUC*	*CA*	*F1*	*Precision*	*Recall*
Neural Network	**0.972**	**0.875**	**0.878**	**0.891**	**0.875**
CN2 rule inducer	**0.875**	**0.875**	**0.876**	**0.878**	**0.875**
Naive Bayes	**0.903**	**0.825**	**0.830**	**0.861**	**0.825**
Tree	0.850	0.800	0.805	0.827	0.800
Random Forest	0.893	0.775	0.782	0.836	0.775
SVM	0.937	0.650	0.655	0.786	0.650

Given Tables 6 and 7, the following can be said: the determined frequency characteristics are suitable for an unambiguous classification. Considerable success has been achieved. Considering that the signals are obtained while the studied muscle maintains a certain position and does not move, although the signal is stationary, it can be said that it contains classifiable frequency components. The most obvious reason for this result is that the peak frequency of the signals obtained from CDH patients is slightly shifted towards high frequencies.

When examining Table 6, the top 3 algorithms are Random Forest, Tree (decision tree), and Neural Network algorithms by classification accuracy. However, when examining Table 7, the top 3 algorithms are Neural Network, CN2 rule inducer, and Naive Bayes algorithms. This situation can be interpreted as follows: Random Forest and Tree algorithms are very similar; they are based on the same methods that include pruning steps. The neural network algorithm, on the other hand, is suitable for continuous data and no pruning is applied. Since the prediction data set has not yet participated in the training, tree-based algorithms can prune the branches that will process this data during modeling. In contrast, since artificial neural networks are more immune to this situation, they were more successful in the prediction experiment. However, the CN2 algorithm, whose basic logic is similar to tree-based algorithms, was also successful because it focuses on creating rules rather than branches. To understand the effects of feature selection, an additional experiment was also applied to the entire dataset without feature ranking. In this final experiment, all the features in Table 3 were used in the training and testing process. Table 8 shows the test results of the classifiers with all features.

Table 8. Classification results with no feature selection

Model	*AUC*	*CA*	*F1*	*Precision*	*Recall*
Tree	**0,9289**	**0,9167**	**0,9142**	**0,9266**	**0,9167**
Neural Network	**0,9424**	**0,8833**	**0,8828**	**0,8828**	**0,8833**
AdaBoost	**0,8643**	**0,8833**	**0,8815**	**0,8846**	**0,8833**
Random Forest	0,9360	0,8500	0,8456	0,8546	0,8500
Naive Bayes	0,9248	0,8500	0,8519	0,8656	0,8500
CN2 rule inducer	0,8314	0,8333	0,8239	0,8511	0,8333
SVM	0,6769	0,8167	0,8043	0,8386	0,8167

Classification using all features has shown a lower success rate than classification using selected features Although it is thought that a more detailed classification will be made when more features are used in classification, it seems that considering the features with low information gain negatively affects the success of the classification. When the overall results are evaluated, the most successful algorithms are tree-based algorithms. Two classification results can be given as examples. Figure 5 represents the decision tree result for the experiments and

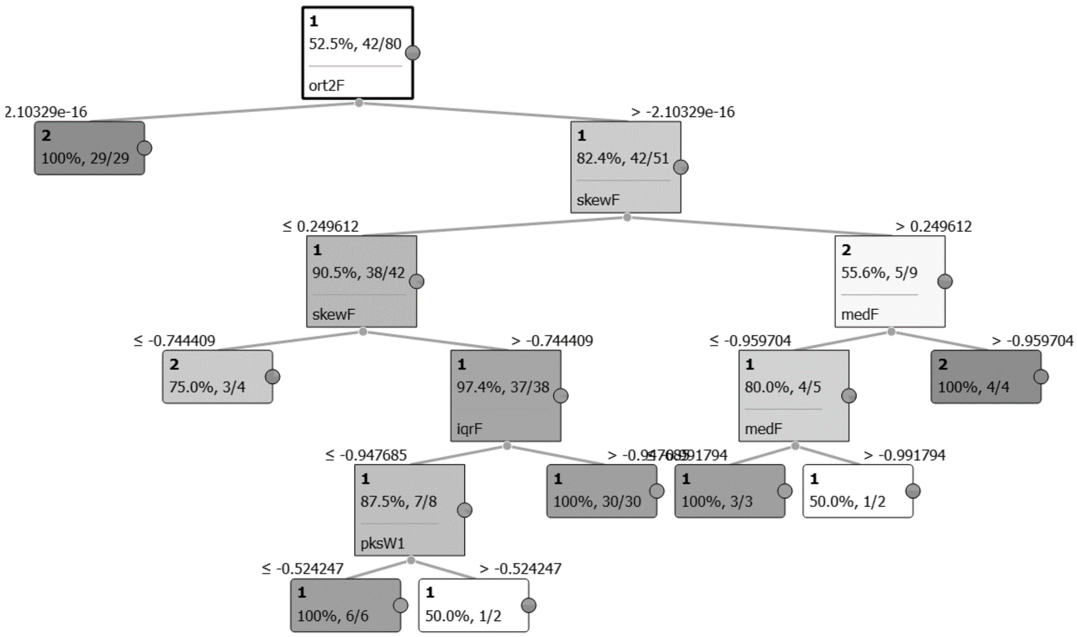

Figure 5. Decision Tree result with the selected features

Table 9 represents the rules generated by the CN2 Rule Inducer algorithm.

Table 9. Induced rules for the selected features.

	IF conditions	THEN class	Distribution	Probabilities [%]	Quality	Length
1	QF2≥-0.8524 AND ort2F≥-2.1015e-16	1	[13, 0]	93:7	0.933	2
2	iqrF≥-0.7337 AND skewF≥-0.7148 AND ort2F≥-1.7791-16	1	[22, 0]	96:4	0.958	3
3	ort2F≥-2.1031e-16 AND iqrF≤-0.9476 AND QF2≥-0.9892 AND medF≥-0.9936	1	[5, 0]	86:14	0.857	4
4	medF≥-0.9937 AND medF≤-0.9925 AND iqrF≥-0.9638	1	[2, 0]	75:25	0.750	3
5	ort2F≤-2.1031e-16 AND pksW2≥-0.9865	2	[0, 25]	4: 96	0.963	2
6	skewF≥0.4860 AND pksW1≥-0.97145	2	[0, 4]	17:83	0.833	2
7	pksW2≤-0.9732 AND medF≥-0.9363	2	[0, 4]	17:83	0.833	2
8	skewF≤-0.7148 AND pksW2≥-0.6413	2	[0, 2]	25:75	0.750	2
9	QF2≥-0.9894 AND QF2≤-0.9653 AND pksW1≥-0.5121	2	[0, 2]	25:75	0.750	3
	TRUE	1	[42, 38]	52:48	0.524	

6. CONCLUSION

This study focused on two questions, first question is “Could sEMG signal recorded from the trapezius muscle be an effective method in the diagnosis of CDH disease” and the second question is “what kind of classifiers are suitable for this task?”. For this purpose, PSD-based features were preferred, because, in the experiment, muscles remained at a certain position, so a time series approach was not suitable for this task. Feature extraction was not enough precursor for the classification task, so feature selection was also very important. During the analysis of the features, it has been obvious that the peak frequency value plays an important role in the features. So, when the feature selection metrics were applied to the features, the top 10 of them were based on peak frequency values.

The final step was evaluating the performance of different classifier algorithms. There are many methods used for the classification of sEMG data [45-49]. In this study, tree-based algorithms and hyperplane-based algorithms were used. In comparison, an interesting situation has occurred. After the training and cross-validation process, the most successful algorithms were tree-based algorithms, depending on the results it can be said that the frequency features obtained from sEMG signals are more suitable to be classified with the tree-based algorithms like Decision Tree or Random Forest. Hence the success rates of all algorithms were good enough to classify the subjects accurately.

There are many current studies in the literature for feature extraction and classification from EMG signals. These methods differ according to experiment design and selected muscle group and generally, high classification accuracy is obtained. EMG classifications are generally used in the automated diagnostic system for neuromuscular diseases or used for prosthetic device control.

In conclusion, according to the results, this study claims that Cervical Disc Herniation affects the electrical activity of the trapezius muscle utilizing resting-state sEMG signals. Our findings are compatible with the literature and provide higher classification accuracy than studies using surface EMG in muscle disorders [31-32].

The main contribution of this study is high classification and prediction values that will be beneficial with non-invasive sEMG in the clinical assessment before MRI in the diagnosis of Cervical Hernia patients. It is considered that this study is an important step in determining the source of muscle pain occurring in the neck region.

The main limitation of this study is the small number of subjects. When the number of subjects is increased, the results will become more reliable, and an automated diagnosis system could be designed.

Declaration of Ethical Standards

The authors of this manuscript declare that they comply with all ethical standards.

Credit Authorship Contribution Statement

Burak Yılmaz: Software, Visualization, Investigation. Writing- Reviewing and Editing, **Güzin Özmen:** Conceptualization, Methodology, Data curation, Writing- Original draft preparation. **Hakan Ekmekçi:** Supervision, Data Curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding / Acknowledgements

This work was not supported by any funding sources.

Data Availability

The data set in this study has not shared in any kind of media yet.

REFERENCES

[1] Mochida, Kiyoshi, Hiromichi Komori, Atsushi Okawa, Takeshi Muneta, Hirotaka Haro, and Kenichi Shinomiya. "Regression of Cervical Disc Herniation Observed on Magnetic Resonance Images." *Spine* 23 (9): 990–95. https://doi.org/10.1097/00007632-199805010-00005, 1998.

[2] Yeung, Jacky T., John I. Johnson, and Aftab S. Karim. "Cervical Disc Herniation Presenting with Neck Pain and Contralateral Symptoms: A Case Report." *Journal of Medical Case Reports* 6 (1): 166. https://doi.org/10.1186/1752-1947-6-166, 2012.

[3] Sharrak, Samir, and Yasir Al Khalili. "Cervical Disc Herniation." *StatPearls [Internet]*, 2020.

[4] Kamibayashi, Lynne K, and Frances J.R. Richmond. "Morphometry of Human Neck Muscles." *Spine* 23 (12): 1314–23. https://doi.org/10.1097/00007632-199806150-00005, 1998.

[5] Takebe, Kyoichi, Mathias Vitti, and John V Basmajian. "The Functions of Semispinalis Capitis and Splenius Capitis Muscles: An Electromyographic Study." *The Anatomical Record* 179 (4): 477–80. https://doi.org/10.1002/ar.1091790407, 1974.

[6] Bernhardt, P, H-J. Wilke, K.H. Wenger, B Jungkunz, A Böhm, and L.E. Claes. "Multiple Muscle Force Simulation in Axial Rotation of the Cervical Spine." *Clinical Biomechanics* 14 (1): 32–40. https://doi.org/10.1016/S0268-0033(98)00031-X, 1999.

[7] Sommerich, Carolyn M., Sharon M.B. Joines, Veerle Hermans, and Samuel D. Moon. "Use of Surface Electromyography to Estimate Neck Muscle Activity." *Journal of Electromyography and Kinesiology* 10 (6): 377–98. https://doi.org/10.1016/S1050-6411(00)00033-X, 2000.

[8] Bronzino, Joseph D. *Biomedical Engineering Handbook 2*. Vol. 2. Springer Science & Business Media, 2000.

[9] Luca, Carlo J De. "The Use of Surface Electromyography in Biomechanics." *Journal of Applied Biomechanics* 13 (2): 135–63. https://doi.org/10.1123/jab.13.2.135, 1997.

[10] Subasi, Abdulhamit. "Classification of EMG Signals Using PSO Optimized SVM for Diagnosis of Neuromuscular Disorders." Computers in Biology and Medicine 43 (5): 576–86. https://doi.org/10.1016/j.compbiomed.2013.01.020, 2013.

[11] Subasi, Abdulhamit, Emine Yaman, Yara Somaily, Halah A. Alynabawi, Fatemah Alobaidi, and Sumaiah Altheibani. "Automated EMG Signal Classification for Diagnosis of Neuromuscular Disorders Using DWT and Bagging." *Procedia Computer Science* 140: 230–37. https://doi.org/10.1016/j.procs.2018.10.333, 2018.

[12] Jose, Shobha, S Thomas George, M S P Subathra, Vikram Shenoy Handiru, Poornaselvan Kittu Jeevanandam, Umberto Amato, and Easter Selvan Suviseshamuthu. "Robust Classification of Intramuscular EMG Signals to Aid the Diagnosis of Neuromuscular Disorders." *IEEE Open Journal of Engineering in Medicine and Biology* 1: 235–42. https://doi.org/10.1109/OJEMB.2020.3017130, 2020.

[13] Akef Khowailed, Iman, and Ahmed Abotabl. "Neural Muscle Activation Detection: A Deep Learning Approach Using Surface Electromyography." *Journal of Biomechanics* 95 (October): 109322. https://doi.org/10.1016/j.jbiomech.2019.109322, 2019.

[14] Azzerboni, Bruno, Giovanni Finocchio, Maurizio Ipsale, Fabio La Foresta, and Francesco Carlo Morabito. "A New Approach to Detection of Muscle Activation by Independent Component Analysis and Wavelet Transform." In *Italian Workshop on Neural Nets*, 109–16. Springer. https://doi.org/10.1007/3-540-45808-5_11, 2002.

[15] Karthick, P.A., Diptasree Maitra Ghosh, and S Ramakrishnan. "Surface Electromyography Based Muscle Fatigue Detection Using High-Resolution Time-Frequency Methods and Machine Learning Algorithms." *Computer Methods and Programs in Biomedicine* 154 (February): 45–56. https://doi.org/10.1016/j.cmpb.2017.10.024, 2018.

[16] Subasi, Abdulhamit, and M Kemal Kiymik. "Muscle Fatigue Detection in EMG Using Time–Frequency Methods, ICA and Neural Networks." *Journal of Medical Systems* 34 (4): 777–85. https://doi.org/10.1007/s10916-009-9292-7, 2010.

[17] Silva, Luís, João Rocha Vaz, Maria António Castro, Pedro Serranho, Jan Cabri, and Pedro Pezarat-Correia. "Recurrence Quantification Analysis and Support Vector Machines for Golf Handicap and Low Back Pain EMG Classification." *Journal of Electromyography and Kinesiology* 25 (4): 637–47. https://doi.org/10.1016/j.jelekin.2015.04.008, 2015.

[18] Ostojić, Saša, Stanislav Peharec, Vedran Srhoj-Egekher, and Mario Cifrek. "Differentiating Patients with Radiculopathy from Chronic Low Back Pain Patients by Single Surface EMG Parameter." *Automatika* 59 (3–4): 400–407. https://doi.org/10.1080/00051144.2018.1553669, 2018.

[19] Jiménez-Grande, David, S. Farokh Atashzar, Eduardo Martinez-Valdes, Alessandro Marco De Nunzio, and Deborah Falla. "Kinematic Biomarkers of Chronic Neck Pain Measured during Gait: A Data-Driven Classification Approach." *Journal of Biomechanics* 118 (March): 110190. https://doi.org/10.1016/j.jbiomech.2020.110190, 2021.

[20] Walsh, Kevin A, Sean P Sanford, Brian D Collins, Noam Y Harel, and Raviraj Nataraj. "Performance Potential of Classical Machine Learning and Deep Learning Classifiers for Isometric Upper-Body Myoelectric Control of Direction in Virtual Reality with Reduced Muscle Inputs." *Biomedical Signal Processing and Control* 66 (April): 102487. https://doi.org/10.1016/j.bspc.2021.102487, 2021.

[21] Kumar, Shrawan, and Narasimha Prasad. "Cervical EMG Profile Differences between Patients of Neck Pain and Control." *Disability and Rehabilitation* 32 (25): 2078–87. https://doi.org/10.3109/09638288.2010.481029, 2010.

[22] Ozmen, Guzin, and Ahmet Hakan Ekmekci. "Classification of Cervical Disc Herniation Disease Using Muscle Fatigue Based Surface EMG Signals by Artificial Neural Networks." *International Journal of Intelligent Systems and Applications in Engineering* 4 (5): 256–62. https://doi.org/10.18201/ijisae.2017533901, 2017.

[23] Phinyomark, Angkoon, Pornchai Phukpattaranont, and Chusak Limsakul. "Feature Reduction and Selection for EMG Signal Classification." *Expert Systems with Applications* 39 (8): 7420–31. https://doi.org/10.1016/j.eswa.2012.01.102, 2012.

[24] Qin, Pengjie, and Xin Shi. "Evaluation of Feature Extraction and Classification for Lower Limb Motion Based on SEMG Signal." *Entropy* 22 (8). https://doi.org/10.3390/E22080852, 2020.

[25] Katsis, C.D., Y. Goletsis, A. Likas, D.I. Fotiadis, and I. Sarmas. "A Novel Method for Automated EMG Decomposition and MUAP Classification." *Artificial Intelligence in Medicine* 37 (1): 55–64. https://doi.org/10.1016/j.artmed.2005.09.002, 2006.

[26] Rasheed, Sarbast, Daniel Stashuk, and Mohamed Kamel. "A Software Package for Interactive Motor Unit Potential Classification Using Fuzzy K-NN Classifier." *Computer Methods and Programs in Biomedicine* 89 (1): 56–71. https://doi.org/10.1016/j.cmpb.2007.10.006, 2008.

[27] Gokgoz, E., & Subasi. "Comparison of Decision Tree Algorithms for EMG Signal Classification Using DWT." *Biomedical Signal Processing and Control* 18: 138–44. https://doi.org/10.1016/j.bspc.2014.12.005, 2015.

[28] Kamali, Tahereh, Reza Boostani, and Hossein Parsaei. "A Multi-Classifier Approach to MUAP Classification for Diagnosis of Neuromuscular Disorders." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22 (1): 191–200. https://doi.org/10.1109/TNSRE.2013.2291322, 2014.

[29] Artameeyanant, Patcharin, Sivarit Sultornsanee, and Kosin Chamnongthai. "An EMG-Based Feature Extraction Method Using a Normalized Weight Vertical Visibility Algorithm for Myopathy and Neuropathy Detection." *SpringerPlus* 5 (1): 2101. https://doi.org/10.1186/s40064-016-3772-2., 2016

[30] Hazarika, Anil, Lachit Dutta, Meenakshi Boro, Mausumi Barthakur, and Manabendra Bhuyan. "An Automatic Feature Extraction and Fusion Model: Application to Electromyogram (EMG) Signal Classification." *International Journal of Multimedia Information Retrieval* 7 (3): 173–86. https://doi.org/10.1007/s13735-018-0149-z, 2018.

[31] Istenič, Rok, Prodromos A. Kaplanis, Constantinos S. Pattichis, and Damjan Zazula. "Multiscale Entropy-Based Approach to Automated Surface EMG Classification of Neuromuscular Disorders." *Medical & Biological Engineering & Computing* 48 (8): 773–81. https://doi.org/10.1007/s11517-010-0629-7, 2010.

[32] Barmpakos, Dimitrios, Prodormos Kaplanis, Stavros A. Karkanis, and Constantinos Pattichis. "Classification of Neuromuscular Disorders Using Features Extracted in the Wavelet Domain of SEMG Signals: A Case Study." *Health and Technology* 7 (1): 33–39. https://doi.org/10.1007/s12553-016-0153-3, 2017.

[33] Moore, Keith L, A M R Agur, and Arthur F Dalley. *Essential Clinical Anatomy*. Baltimore, MD: Lippincott Williams & Wilkins, 2011.

[34] Halaki, Mark. "Normalization of EMG Signals: To Normalize or Not to Normalize and What to Normalize To?" In , edited by Karen Ginn ED1 - Ganesh R Naik, Ch. 7. Rijeka: IntechOpen. https://doi.org/10.5772/49957, 2012.

[35] Awal, Abdul, Sheikh Shanawaz Mostafa, and Mohiuddin Ahmad. "Performance Analysis of Savitzky-Golay Smoothing Filter Using ECG Signal." *International Journal of Computer and Information Technology* 01 (02): 24–29, 2011.

[36] Bucy, R S. "Burg Technique." In *Lectures on Discrete Time Filtering*, 47–54. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4613-8392-5_5, 1994.

[37] Subasi, A. Practical guide for biomedical signals analysis using machine learning techniques: A MATLAB based approach. Academic Press, 2019

[38] Altunkaya, Sabri, Sadık Kara, Niyazi Görmüş, and Saadetdin Herdem. "Comparison of First and Second Heart Sounds after Mechanical Heart Valve Replacement." *Computer Methods in Biomechanics and Biomedical Engineering* 16 (4): 368–80. https://doi.org/10.1080/10255842.2011.623672, 2013.

[39] Altunkaya, Sabri. "Yaşlılarda Düşme Riskini Değerlendirmek Için İvmelenme Sinyalinin Frekans Domeni Özellikleri." *European Journal of Science and Technology*, no. August: 150–55. https://doi.org/10.31590/ejosat.779590, 2020.

[40] Wu, Xindong, Vipin Kumar, Quinlan J. Ross, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, et al. *Top 10 Algorithms in Data Mining*. *Knowledge and Information Systems*. Vol. 14. https://doi.org/10.1007/s10115-007-0114-2, 2008.

[41] Clark, Peter, and Tim Niblett. "The CN2 Induction Algorithm." *Machine Learning* 3 (4): 261–83. https://doi.org/10.1023/A:1022641700528, 1989.

[42] Breiman, Leo. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324, 2001.

[43] Balamareeswaran, M., and D. Ebenezer. "Denoising of EEG Signals Using Discrete Wavelet Transform Based Scalar Quantization." *Biomedical and Pharmacology Journal* 8 (1): 399–406. https://doi.org/10.13005/bpj/627, 2015.

[44] Powers, David M. W. "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation," no. January 2011. https://doi.org/10.9735/2229-3981, 2020.

[45] Chowdhury, Suman Kanti. "Discrete Wavelet Transform Analysis of Surface Electromyography for the Objective Assessment of Neck and Shoulder Muscle Fatigue Discrete Wavelet Transform Analysis of Surface Electromyography for the Objective Assessment of Neck and Shoulder Muscle Fatigue", Statler College of Engineering and Mineral Resources, Master Thesis, https://doi.org/10.33915/etd.4841, 2012.

[46] Gokgoz, Ercan, and Abdulhamit Subasi. "Effect of Multiscale PCA De-Noising on EMG Signal Classification for Diagnosis of Neuromuscular Disorders." *Journal of Medical Systems* 38 (4): 31. https://doi.org/10.1007/s10916-014-0031-3., 2014.

[47] Pauk, J. "Different Techniques for EMG Signal Processing." *Journal of Vibroengineering* 10 (4): 571–76, 2008.

[48] Akhundov, Riad, David J. Saxby, Suzi Edwards, Suzanne Snodgrass, Phil Clausen, and Laura E. Diamond. "Development of a Deep Neural Network for Automated Electromyographic Pattern Classification." *Journal of Experimental Biology* 222 (5): 1–5. https://doi.org/10.1242/jeb.198101, 2019.

[49] Siqueira, Ailton Luiz Dias, and Alcimar Barbosa Soares. "A Novel Method for EMG Decomposition Based on Matched Filters." *Revista Brasileira de Engenharia Biomedica* 31 (1): 44–55. https://doi.org/10.1590/2446-4740.0643, 2015.