



HAL
open science

Automatic classification of intramuscular EMG to recognize pathologies.

Alban Gallard, Konstantin Akhmadeev, Éric Le Carpentier, Raphaël Gross,
Yann Péréon, Yannick Aoustin

► **To cite this version:**

Alban Gallard, Konstantin Akhmadeev, Éric Le Carpentier, Raphaël Gross, Yann Péréon, et al.. Automatic classification of intramuscular EMG to recognize pathologies.. Advanced Structured Materials, Springer Nature, inPress. hal-02362458

HAL Id: hal-02362458

<https://hal.science/hal-02362458v1>

Submitted on 13 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chapter 1

Automatic classification of intramuscular EMG to recognize pathologies

Alban Gallard, Konstantin Akhmadeev, Eric Le Carpentier, Raphaël Gross, Yann Péréon, and Yannick Aoustin

Abstract This paper proposes to assess the relevance of new automated tools for electromyography (EMG) analysis, in order to differentiate neuropathic from myopathic patterns. The challenge is to define the diagnosis with only one iEMG signal per patient. Our proposed method uses the decomposition of the EMG signal to characterize motor unit action potentials (MUAPs). The decomposition of each iEMG signal is carried out with EMGLAB. For each signal, the decomposition provides a code which is used by the automated classification algorithms. We use here the linear Support Vector Machine (SVM) and the Bagging Trees methods. For the learning process we use several EMG signals and in different parts of the muscle. Only one recorded electromyography EMG signal per subject is used for the diagnostic test. We evaluate the k -fold cross-validation and the confusion matrix for both models. The accuracy is 77.3% for the SVM and 68.2% for the Bagging Trees. These are the first developments of this tool to make it useful for clinical practice.

Keywords: Quantitative electromyography·Motor unit action potential· Classification· Support Vector Machine·Bagging Trees.

Alban Gallard, Konstantin Akhmadeev, Eric Le Carpentier, and Yannick Aoustin
Laboratoire des Sciences du Numérique de Nantes, UMR CNRS 6004
1 rue de la Noë, BP 92101 F-44321 Nantes Cedex 3, France
Ecole Centrale de Nantes, Université de Nantes
e-mail: Yannick.Aoustin@univ-nantes.fr

Raphaël Gross
Laboratoire d'Analyse du Mouvement, CHU de Nantes 44000 Nantes
e-mail: Raphael.Gross@chu-nantes.fr

Yann Péréon
Laboratoire d'Explorations Fonctionnelles, Centre de Références Maladies
Neuromusculaires Atlantique-Occitanie-Caraïbes (LEF), Université de Nantes,
44000 CHU de Nantes
e-mail: Yann.Pereon@univ-nantes.fr

1.1 Introduction

Electromyography (EMG) is used in routine in clinical practice. Electrodes are inserted into the muscle (intramuscular EMG, iEMG) or placed on the skin (surface EMG, sEMG) to record the muscular electrical activity. The measured EMG signal is a sum of elementary contributions. Each contribution is a wavelet train produced by a motor unit (MU) in the electrode vicinity. A MU corresponds to a spinal motoneuron (MN) and the muscle fibers it innervates [Heckman and Enoka \(2012\)](#), and the wavelet is called the Motor Unit Action Potential (MUAP). A variation of muscle activation level produces a variation of the number of active MUs in a process called "spatial recruitment", and a variation of the discharge rate (that is the number of MUAPs per time unit) of the active MUs, called "temporal recruitment".

EMG signals study can be used to detect neuromuscular diseases. It provides information in favor of either muscle lesions (so-called myopathic pattern) or nerve lesions (so-called neuropathic pattern). Nerve and muscle lesions affect both individual MUAP characteristics and recruitment pattern during contraction. In routine, neurologists assess both through a visual and auditory analysis, but differentiating normal recordings from myopathic or neuropathic ones can be difficult with possible misdiagnosis [Pereon \(2015\)](#), [Harwood et al. \(2012\)](#).

EMG can also be used to gain knowledge about the neuromuscular mechanisms involved in the force and movement control. For example, in [Gross et al. \(2016\)](#), EMG data and capture motion data are used to analyze children walking along a curved path. This study investigated the EMG activity in the lower limbs of typically developing children during turning trajectories with increasing curvature resulting from changes of direction towards targets, and evidenced that differences exist in curved compared with straight walking.

Despite the medical doctors' expertise a purely descriptive approach is not sufficient. An automated analysis can lead to an objective diagnosis minimizing observer bias, [Dorfman and McGILL \(1988\)](#). For a few decades now different approaches have been proposed for rendering the EMG examination more quantitative and more automatic to permit precise interpretation of the findings, to minimize observer bias, to facilitate comparative studies over time for different methodologies. An automated analysis relies on a preliminary decomposition of the processed signal, that is the separation of the active MUs contributions together with MUAPs shapes and trains characteristics estimation, to provide EMG features. [Pattichis and al. \(1995\)](#) proposed artificial neural networks for the automatic classification of EMG features, which are recorded from normal individuals and patients suffering with neuromuscular diseases. In the field of machine learning, [Katsis et al.](#) used a Support Vector Machine (SVM) algorithm [Katsis et al. \(2006\)](#), or a radial basis function network and a decision tree [Katsis et al. \(2007\)](#) for automated EMG decomposition and MUAPs classification. Let us remark that several methods exist to carry out this automated EMG decomposition: the Convolution Kernel Compensation (CKC) [Holobar and Zazula \(2007\)](#); [Holobar et al. \(2009\)](#), Monte Carlo Markov Chain based decomposition [Ge et al. \(2011\)](#), Bayesian filtering based on a hidden Markov model of the intramuscular EMG (iEMG) [Monsifrot et al. \(2014\)](#). They allow a full decom-

position of single-channel iEMG signals produced during contractions at moderate force levels but presenting superimposed MUAPs. Farina *et al* Farina *et al.* (2001) propose a tool to evaluate and compare EMG signal decomposition algorithms.

Note that these previous works to help to diagnosis are about the MUAP classification. Another way can be to make the EMG classification. Pattichis and Schizas Pattichis and Schizas (1996) studied a hybrid diagnostic system that combines both neural network and genetics-based machine learning models to diagnose certain neuromuscular diseases from EMG data. 680 MUAP's are collected from twelve healthy subjects, 11 patients with neuropathy and 11 patients with myopathy. The experimental protocol therefore uses 20 MUAPS per patient. Eight people from each group formed the training set, and the other 10 subjects formed the evaluation set. The considered MUAP parameters are duration, spike duration, amplitude, area, spike area, phases, and turns. Each subjects are described mean and the standard deviation of each MUAP parameter in order to characterize the EMG signals and perform the diagnosis. They showed that the diagnostic performance of neural network genetics-based machine learning models is enhanced by the hybrid system. In this paper, a unique EMG signal per subject is used to make a targeted analysis of the MUAP shape, in order to provide a patient diagnosis based on full recorded iEMG signal classification. This paper is also focused on the EMG classification. Sadikoglu *et al* Sadikoglu *et al.* (2017) analyzes the power spectrum of patients' EMG signals to describe parameters such as mean frequency, median frequency, amplitude in terms of root mean square spectral density, minimum and maximum. The classification results distinguish between healthy EMG signals and EMG signals from neuropathic patients. Yousefi and Hamilton-Wright Yousefi and Hamilton-Wright (2014) provide a critical review of several classification methods including Bayesian techniques, neural networks, multilayer perceptrons, fuzzy approaches, SVM, and neuro-fuzzy systems for EMG signal characterization. They conclude that SVM is a remarkably robust classification method in disease diagnosis. Subasi Subasi (2013) proposed to combine the particle swarm optimization (PSO) and SVM to improve the EMG signal classification accuracy. The EMG signals are decomposed into the frequency sub-bands using discrete wavelet transform (DWT). A set of statistical features are extracted from these sub-bands to represent the distribution of wavelet coefficients. The presented results show that the strategy is very efficient. However, the features deduced from the decomposition are not physiologically interpretable.

In this paper, our goal is to mimic the practitioner behavior, that is a combination of overall aspect of the EMG signal with a more targeted analysis of the MUAP shape, in order to provide a patient diagnosis based on full recorded iEMG signal classification. However our method is based on a unique recorded EMG signal per patient. This method necessarily leads to poorer results than those of Pattichis and Schizas (1996) but it is less tedious at the level of the experimental protocol. Since the purpose of this paper is not to compare decomposition tools, we will use a semi-automatic decomposition by means of the EMGLab software with manual corrections EMG (2007); Florestal *et al.* (2006). Each signal will be encoded by the number of active MU, an average MUAP shape and a firing rate for each active MU. Then, we will focus on the use of machine learning tools for the EMG classification

stage. Two algorithms are considered: the linear SVM [Breiman \(1995\)](#), [Yousefi and Hamilton-Wright \(2014\)](#) and the Bagging Trees [Breiman \(1996\)](#). The bagging method is a suitable mean for increasing efficiency of standard machine learning algorithms, see [Machova et al. \(2006\)](#) and [Gokgoz and Subasi \(2015\)](#).

This paper is outlined as follows: section 1.2 where presents the decomposition of the EMG signal, the Motor Unit, the general features of the myopathic and neuropathic patterns, the features and the classification of the EMG signal; section 1.3 shows the different numerical results; then section 1.4 discusses about these results and finally the section 1.5 offers our conclusions and perspectives.

1.2 Methods

The method is composed of five steps. First, the decomposition of the EMG signals gives the MUAPs. Secondly, we need to understand how the MUs work. We also need to know how the pathologies modify EMG signals and more particularly the MUAPs. With the second and third steps, the best features to classify the signal can be found. Finally, the computation of the machine learning algorithms can be performed with the features created in the previous step.

1.2.1 Data

Miki Nikolic from the Rigshospitalet of Copenhagen presented in his Ph.D thesis [Nikolic \(2001\)](#) an automatic system called EMGPAD to decompose the clinical EMG signal into its constituent motor MUAPs and their corresponding firing patterns (FPs). The signals used for this paper were made available by Miki Nikolic on the website of EMGLab [EMG \(2007\)](#); [Florestal et al. \(2006\)](#). Only the Biceps Bracii muscle has been studied to eliminate variations in EMG in the different muscles.

Five healthy people (age rank 21-37 years) were studied, as well as five patients presenting with myopathic lesions (age rank 19-63 years) and five patients with neuropathic lesions (age rank 35-67 years).

1.2.1.1 Preprocessing

The EMG signal was recorded using intramuscular needle electrodes with a cross-section of 0.07 mm^2 . A surface electrode has been placed on the limb to allow single pole work. The signal was measured for 11.2 s in each case.

The signals were amplified 4000 times and filtered with a bandpass filter between 2 Hz and 10 kHz. Then, they were sampled and digitized. The sampling frequency was 23437 Hz and the digital conversion is encoded with 16 bits processor.

Several measures were taken in different depths and different locations of the Biceps Bracii muscle to explore it globally.

1.2.1.2 Decomposition

The decomposition of the EMG signals was performed semi-automatically using EMGLab [EMG \(2007\)](#).

For simple cases, the automatic decomposition was correct. A manual check was carried out to avoid any errors. For complex cases, the automatic decomposition created too many errors. Automatic decomposition and manual verification take longer than manual decomposition alone.

The first step in manually decomposing a signal is to find isolated MUAPs. It is important to check their repetitions to ensure that they are isolated. Afterwards, each time the MUAPs are found in the EMG signal, a mark is placed. This allows you to decompose overlapping MUAPs and find new ones.

The number of signals decomposed for each patient is detailed in the table [1.1](#). There are respectively 50 signals for the groups of healthy subjects and patients with myopathy. There are 45 signals for the neuropathic patient group. Thus, 145 signals were decomposed to feed the machine learning algorithms. Each signal is represented by the number of active MU's and, for each active MU, by the firing rate and the average MUAP.

	Subject numbers	Number of signals	Total
Healthy Subjects	1	17	50
	2	12	
	3	4	
	4	14	
	5	3	
Myopathic Subjects	1	15	50
	2	8	
	3	6	
	4	9	
	5	12	
Neuropathic Subjects	1	12	45
	2	9	
	3	8	
	4	5	
	5	11	

Table 1.1: Number of signals for each healthy subject and patient.

1.2.2 Motor unit (MU)

Studying the creation and propagation of the electrical signal is important to understand the differences between the pathologies. The central nervous system composed of the brain and the spinal cord sends electrical messages to the muscles [Heckman and Enoka \(2012\)](#). These messages are sent by using the motoneurons (MNs) [Silani et al. \(1999\)](#), which are located in the brain stem and in the anterior part of the grey matter of the spinal cord, through the motor axon along the peripheral nerves to the muscles fibers. The muscle fibers innervated by a motoneuron are called muscle unit. The motor unit (MU) defines the association of the motoneuron and the muscles fibers it innervates. The MUs transform the message of the central nervous system to contract a muscle and to extend the opposite one. In that way, a synergistic effect and an antagonistic force are produced [Farina et al. \(2016\)](#).

A MUAP found in the EMG is the sum of the action potentials of all the muscle fibers innervated by a single motoneuron. The electrode location is important because an EMG measures a part of the electrical signal of the muscle fibers of the MUs. Moreover, a fiber closer to the electrodes gives a more important signal.

Waveform characteristics are different for all MUAPs, for example the duration or the amplitude. The number of phases, which is the number of sign changes of the signal and also changes; such as the number of peaks.

The rate at which the MUAPs are discharged is the firing rate. In other words, it is the number of MUAPs per unit of time. It increases with strength.

The number of muscle fibers per MU is different for all muscles. That is why, it is important to always use the same muscle to compare the EMG.

1.2.3 Pathologies

To find the best features for machine learning it is essential to know how the pathologies affect the MUs and the EMG signals. Typically, in case of nerve or root lesions, a lower number of MUs are recruited (poor recruitment because fewer motor neurons are available), and at a higher rate. The chronic denervation-reinnervation process is responsible for the increase in size and duration of MUAPs due to distal reinnervation: this defines the EMG neuropathic pattern. In the case of muscular dystrophy, a lower number of muscle fibres per MU is present, providing MUAPs of low amplitude, short duration, multiple turns, and recruitment is abnormally increased (rich recruitment) in order to provide as much as possible efficient strength: this induces the EMG myopathic pattern. Figure 1.1 shows this phenomenon by simplifying it [Kubis and Lozeron \(2014\)](#).

Table 1.2 summarizes the consequences of the disease on the MUAPs. The healthy case is the standard, which is why there is the sign = at every lines. The + sign means higher and the – sign means lower.

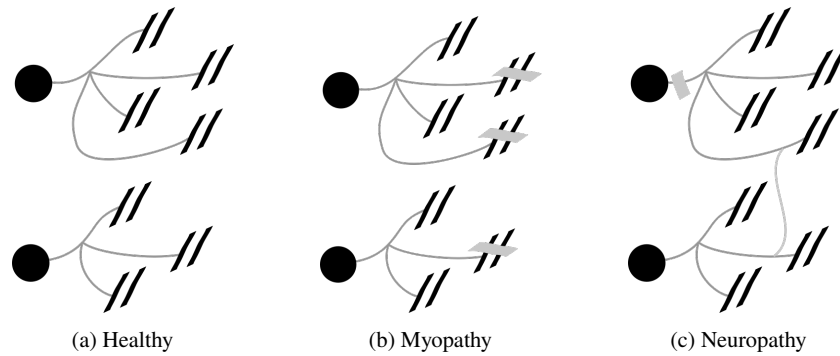


Fig. 1.1: Diagram of the MUs. The black circles represent the motoneurons and the pairs of black lines are the muscle fibers. The grey lines represent the connection between a motoneuron and the muscle fibers.

	Healthy	Myopathy	Neuropathy
# active MUs	=	+	-
MU firing rate	=	+	-
MUAP duration	=	-	+
MUAP amplitude	=	-	+
MUAP # phases	=	+	+
MUAP # turns	=	+	-

Table 1.2: Consequences of the pathologies (# means number).

1.2.4 Features

The features of the Table 1.2 can be used to differentiate the pathologies. The number of active MUs is the first feature of the signal, but the others are defined for each MU. To obtain a unique length of the feature vector, the minimum, the mean and the maximum values are used.

The vector of features for a signal, considering the set of MUAPs, has 16 elements: the number of MUAPs and the minimum, mean and maximum of the duration, amplitude, firing (discharge) rates, number of phases and number of turns.

In this vector, some of the features can be useless to differentiate the pathologies. A T-test is used to find them. A T-Test rejects or not the null hypothesis: there is no significant statistical difference between the two distributions. In that case, the two distributions are one feature for a pathology and the same feature with another pathology [Rice \(2006\)](#).

The T-test is composed of three variables. The first one is the significant value α that determines the percentage of rejection of the hypothesis. The two others are the T-value TV and the degree of freedom of the T-test df . Let us consider two distributions *Healthy Subjects–Myopathic Subjects*, *Healthy Subjects–Neuropathic Subjects*, or *Myopathic Subjects–Neuropathic–Subjects*. The values of T-test and

df are calculated with the mean of the distribution \bar{x} and \bar{y} , their standard deviation σ_x and σ_y , their number of data n and m :

$$TV = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \quad (1.1)$$

$$df = n + m - 2 \quad (1.2)$$

TV is after compared with the result found with df and the significant value α in the table of T-Test. The value of α is chosen equal to 0.05. If it is higher, the null hypothesis will be rejected and the features must be kept. Otherwise, it will not be rejected and the features can be deleted.

The T-test shows that the number of active MUs is not significant.

The correlation analysis indicates the strength of the relationship between two features. The correlation coefficient $\rho_{A,B}$ determines the correlation between the two distributions A and B. In our case, A and B are two different features with all pathologies. The following equation gives $\rho_{A,B}$:

$$\rho_{A,B} = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{A_i - \bar{a}}{\sigma_A} \right) \left(\frac{B_i - \bar{b}}{\sigma_b} \right) \quad (1.3)$$

Where \bar{a} and \bar{b} are the means of the distributions, σ_A and σ_B their standard deviations and N their number of data.

This analysis shows the mean of the duration, the mean of the number of phases and the mean of the number of turns are highly correlated to the other features.

The 12 features remaining are:

- Minimum and maximum of the duration.
- Minimum, mean and maximum of the amplitude.
- Minimum and maximum of the number of phases.
- Minimum and maximum of the number of turns.
- Minimum, mean and maximum of the firing rate.

1.2.5 Classification

With the features found in the previous step, the machine learning models can be studied. The two models used are the Support Vector Machine (SVM) and the Bagging Trees. The first one has been chosen for its simplicity and the second one for its speed and its good performance.

1.2.5.1 Linear SVM

The SVM separates the data in several parts. For the linear SVM, the separation is an hyperplane. Its equation is:

$$h(\mathbf{x}) = w_0 + \mathbf{x}^\top \mathbf{w} \quad (1.4)$$

The SVM uses margins that have to be the largest to decrease the error. Usually, the data can not be separated. To have less errors, the SVM attributes a slack variable ξ_i to each point. Its value depends on the classification. It will be null, if it is correct; less than 1, if it is in the margin; and more than 1, if it is false. The problem to find the parameters \mathbf{w} , w_0 and ξ_i is solved thanks the following minimization:

$$\min_{w_0, \mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (1.5)$$

With the conditions:

$$\begin{aligned} \xi_i &\geq 0 \quad \forall i \\ y_i(w_0 + \mathbf{x}_i^\top \mathbf{w}) &\geq 1 - \xi_i \quad \forall i \end{aligned}$$

The value of the parameter C is chosen depending on the importance of the slack variables that we want. This problem can be solved with the Lagrangian function where the Lagrange multipliers are maximized.

1.2.5.2 Bagging Trees

The Bagging Trees is a model that uses several Trees. A Tree is a succession of choices. A tree takes one of the features and a value for it. Then a tree compares the data with the value and separates it in two parts. For each part, the Tree does the same thing until a shutoff parameter like a number of points is reached. The Bagging Trees does not use all data for each Tree, but only a part that is different for all Trees.

When the Bagging Trees tests a new data, it finds the number $P_{d,t}$ of all pathologies d for each Tree t . Let N_t be the number of trees. After, the Bagging Trees assembles the N_t trees:

- Major voting.

$$\operatorname{argmax}_d \sum_t (d = \operatorname{argmax}_d P_{d,t}) \quad (1.6)$$

- Sum rule.

$$\operatorname{argmax}_d \sum_t P_{d,t} \quad (1.7)$$

- Mean rule.

$$\operatorname{argmax}_d \frac{\sum_t P_{d,t}}{N_t} \quad (1.8)$$

- Product rule.

$$\operatorname{argmax}_d \prod_t P_{d,t} \quad (1.9)$$

We will use the major voting is usually used for the classification case [Hakim et al. \(2017\)](#) with 30 trees and 100 possible choices.

1.3 Results

The quality of the decomposition can be defined with the signal to interference ratio (SIR) and the standard deviation (SD). These criteria are described for the healthy subjects, the neuropathic patients, and the myopathic patients in table 1.3.

	Healthy	Myopathy	Neuropathy
SIR	8.57 dB	7.53 dB	9.33 dB
SD	3.37	4.04	3.82

Table 1.3: Decomposition: Quality indices for the healthy subjects, the neuropathic patients, and the myopathic patients.

We can remark that the SIR is lower for myopathy because the magnitude signals are lower. Conversely, SIR is higher for neuropathy because the magnitude of the signals are higher.

The data are split up in two sets. The first, called a training set contains 85% of the data and it is used to train the model. The other one contains 15% of the data and is called the test set. After the training, the model uses the data in the test to verify the number of misclassified signals. The elements of both parts are randomly selected.

k -fold cross-validation method is a reliable method for predicting the error rate of a classification technique [Sandvig et al. \(2008\)](#), [Gokgoz and Subasi \(2015\)](#). k -fold cross-validation arbitrarily divides the data in a given number of subsets, the "folds". For the number of employed folds $k = 10$ we have respectively for each classifier:

- for SVM classifier, Cross-validation accuracy: 68.92%, with STD=13.5.
- for Bagging Trees with 30 trees and and 100 split nodes, Cross-validation accuracy: 84.55%, with STD=8.3.

The Confusion Matrix gives the errors between each pathology for the prediction and the actual pathologies according to expert. The Confusion Matrix for the tests of the SVM and Bagging Trees is represented in tables 1.4 and 1.5. To understand the gathered data let us consider of the healthy subjects in table 1.4. The numbers in blue bold illustrate the agreement between the prediction and reality. There are 8 healthy subjects and 8 predicted healthy subjects. The percentage of good prediction is thus 100%. The first line indicates that the classifier predicts 8 + 4 + 1 healthy subjects but 4 are myopathic patient and 1 is a neuropathic patient. Percentage of good prediction is thus 61.4%.

		Truth			Per
		Heal	Myo	Neuro	
Pred	Heal	8	4	1	61.4%
	Myo	0	6	0	100%
	Neuro	0	0	3	100%
Per		100%	60%	75%	77.27%

Table 1.4: Confusion Matrix for one test of the SVM (Pred = Prediction, Heal = Healthy, Myo = Myopathy, Neuro = Neuropathy, Per = Percentage of good prediction)

		Truth			Per
		Heal	Myo	Neuro	
Pred	Heal	5	3	1	55.56%
	Myo	0	7	0	100%
	Neuro	3	0	3	50%
Per		62.5%	70%	75%	68.18%

Table 1.5: Confusion Matrix for one test of the Bagging Trees (Pred = Prediction, Heal = Healthy, Myo = Myopathy, Neuro = Neuropathy, Per = Percentage of good prediction)

In the Confusion Matrix, the columns are the real pathologies and the rows are the predictions. The last column gives the percentage of good predictions for a pathology on all predictions of this pathology. For the last row, this is the percentage of good predictions for a pathology on all real signal for this pathology. The number in the last cell is the percentage of the total of good predictions on all signals. The accuracy is 77.3% for the SVM and 68.2% for the Bagging Trees.

1.4 Discussion

The percentage of errors for the two models is not very high, 23% for the linear SVM and 32% for the Bagging Trees. For our study, the SVM seems a better model. The errors of the linear SVM come from the mixed data between the pathologies. For the Bagging Trees, the errors can come from overfitting, but as for the linear SVM, it can also be the mixed signals.

The SVM classifier has better test results than validation from the study of k -fold cross-validation. The opposite is true for Bagging Trees. We have to be careful with these results because it depends on the training set and the test set.

The Confusion Matrix of the Tables 1.4 and 1.5 gives information about the localization of the errors. The models usually make the difference between the myopathic and the neuropathic signals. The most important part of the errors is between the healthy and the myopathy. They are more present between the healthy

subjects and the myopathy patients than between the healthy subjects and neuropathy patients.

The last column of the Confusion Matrix shows the good prediction for the myopathic and neuropathic signals. It is possible to trust them, but it is more difficult to believe the models when they say the signal is from a healthy subject.

The last line indicates a patient with neuropathy will usually have the right diagnostic. It is not the case for the healthy subjects and myopathic patients.

Both models can have better results. For example, with a higher number of signals, it will be possible to have models whose the train has signals with a lot of differences to take into account all cases and to avoid errors in the test.

Another way to improve the classification is to use more features. It can be by finding new temporal features or by using frequency ones. The force can also be used. Indeed, the measure can be done with a specific strength to have every features without depending of the force. Moreover, the signal can be measured to different values of the strength to take into consideration the evolution of the features depending to the force.

The SVM can be improved by using a non-linear method. For example, it can be a parabolic, a cubic or a Gaussian SVM.

Bagging Trees can have better results with a larger number of trees. The trees can also have a different stop parameter, but it is important to deal with the overfitting.

1.5 Conclusion

This new methodology automatically classifies the intramuscular EMG signal into three classes: healthy, myopathy and neuropathy. Two machine learning algorithms, the linear SVM and the Bagging Trees, have been studied. For our study SVM classifier has better performance. The classifier tools can give information to the doctor to help for the diagnosis of the pathology. The methodology needs a decomposition of the signal. The obtained MUAPs give features, that must be normalized to be used by the machine learning models. In this work the decomposition was semi-automatic. In future with an automatic decomposition, this methodology can be integrated on-line in a decision support system to help the doctors.

Acknowledgment

This work was supported University of Nantes, Transdisciplinary project: **Modélisation et Décomposition des EMG pour la Reconnaissance et l'Assistance aux Troubles neuromusculaires (Moderato)**.

References

- C. Heckman, R. M. Enoka, Motor unit, *Comprehensive Physiology* 2 (2012) 2629–2682.
- Y. Pereon, Neurogène ou myogène, *Revue Neurologique* 171S 1189 (2015).
[Http://dx.doi.org/10.1016/j.neurol.2015.01.462](http://dx.doi.org/10.1016/j.neurol.2015.01.462).
- C. A. Harwood, C. J. McDermott, P. J. Shaw, Clinical aspect of motor neurone disease, *Medicine* 40 (2012) 540 – 545.
- R. Gross, F. Leboeuf, M. Lempereur, T. Michel, B. Perrouin-Verbe, S. Vieilledent, O. Rémy-Noris, Modulation of lower limb muscle activity induced by curved walking in typically developing children, *Gait & Posture* 50 (2016) 34–41.
- L. J. Dorfman, K. C. McGILL, Aae minimonograph #29: Automatic quantitative electromyography, *Muscle & Nerve* 11 (1988) 804 – 818.
- C. S. Pattichis, C. N. Schizas, L. T. Middleton, Neural network models in emg diagnosis, *IEEE Transactions on Biomedical Engineering* 42 (1995) 486–496.
- C. D. Katsis, Y. Goletsis, A. Likas, D. I. Fotiadis, I. Sarmas, A novel method for automated emg decomposition and muap classification, *Artificial Intelligence in Medicine* 37 (2006) 55 – 64. *Intelligent Data Analysis in Medicine*.
- C. D. Katsis, T. P. Exarchos, C. Papaloukas, Y. Goletsis, D. I. Fotiadis, I. Sarmas, A two-stage method for muap classification based on emg decomposition, *Computers in Biology and Medicine* 37 (2007) 1232 – 1240.
- A. Holobar, D. Zazula, Multichannel blind source separation using convolution kernel compensation, *IEEE Trans. Signal Processing* 55 (2007) 4487 – 4496.
- A. Holobar, D. Farina, M. Gazzoni, R. Merletti, D. Zazula, Estimating motor unit discharge patterns from high-density surface electromyogram, *Clinical Neurophysiology* 120 (2009) 551 – 562.
- D. Ge, E. Le Carpentier, D. Farina, J. Idier, Spike Sorting by Stochastic Simulation, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 19 (2011) 249–259.
- J. Monsifrot, E. L. Carpentier, Y. Aoustin, D. Farina, Sequential decoding of intramuscular emg signals via estimation of a markov model, *IEEE Trans. on Neural Systems and Rehabilitation Engineering* 22 (2014) 1030–1040.
- D. Farina, R. Colombo, R. Merletti, H. B. Olsen, Evaluation of intra-muscular emg signal decomposition algorithms, *Journal of Electromyography and kinesiology* 11 (2001) 175 – 187.
- C. S. Pattichis, C. N. Schizas, Genetics-based machine learning for the assessment of certain neuromuscular disorders, *Journal of the Neurological Sciences* 7 (1996) 427 – 439.
- F. Sadikoglu, C. Kavalcioglu, B. Dagman, Electromyogram (EMG) signal detection, classification of EMG signals and diagnosis of neuropathy muscle disease, in: *Procedia Computer Science*, 120, Budapest, Hungary, 2017, pp. 422–429.
- J. Yousefi, A. Hamilton-Wright, Characterizing emg data using machine-learning tools, *Computers in Biology and Medicine* 51 (2014) 1 – 13.
- A. Subasi, Classification of emg signals using pso optimized svm for diagnosis of neuromuscular disorders, *Computers in Biology and Medicine* 43 (2013) 576 – 586.
- Emglab: A forum for sharing software, data, and information related to emg decomposition (<http://www.emglab.net>), *EMC - Neurologie* (2007).
- J. R. Florestal, P. A. Mathieu, A. Malanda, Automated decomposition of intramuscular electromyographic signals, *IEEE Transactions on Biomedical Engineering* 53 (2006) 832–839.
- L. Breiman, Support vector networks, *Machine learning* 20 (1995) 273 – 297.
- L. Breiman, Bagging predictors, *Machine learning* 24 (1996) 123 – 140.
- K. Machova, F. Barcak, P. Bednar, A bagging method using decision trees in the role of base classifiers, *Acta Polytechnica hungarica* 3 (2006) 121 – 132.
- E. Gokgoz, A. Subasi, Comparaison of decision tree algorithms for emg signal classification using dwt, *Biomedical signal processing and control* 18 (2015) 138 – 144.
- M. Nikolic, Detailed Analysis of Clinical Electromyography Signals EMG Decomposition, Findings and Firing Pattern Analysis in Controls and Patients with Myopathy and Amyotrophic Lateral Sclerosis, Ph.D. thesis, PhD Thesis, Faculty of Health Science, University of Copenhagen, 2001.

- V. Silani, A. Ciammola, A. Pizzuti, V. Cardin, G. Scarlato, Motor neurone metabolism, *Journal of the Neurological Sciences* 169 (1999) 161 – 169.
- D. Farina, F. Negro, S. Muceli, R. M. Enoka, Principles of motor unit physiology evolve with advances in technology, *Physiology* 31 (2016) 83–94.
- N. Kubis, P. Lozeron, Vitesses de conduction nerveuse, *EMC - Neurologie* 12 (2014) 1–13.
- J. A. Rice, *Mathematical Statistics and Data Analysis* (3rd ed.), Duxbury Advanced, 2006.
- L. Hakim, B. Sartono, A. Saefuddin, Bagging based ensemble classification method on imbalance datasets, *International Journal of Computer Science and Network* 6 (2017) 670 – 676.
- J. J. Sandvig, B. Mobasher, R. Burke, A survey of collaborative recommendation and the robustness of model-based algorithms, *IEEE Comput. Soc. Tech. Comm. Data Eng.* 31 (2008) 3 – 13.