



Movement Identification in EMG Signals Using Machine Learning: A Comparative Study

Laura Lasso-Arciniegas^{1(✉)}, Andres Viveros-Melo¹, José A. Salazar-Castro^{1,2}, Miguel A. Becerra³, Andrés Eduardo Castro-Ospina³, E. Javier Revelo-Fuelagán¹, and Diego H. Peluffo-Ordóñez^{2,4}

¹ Universidad de Nariño, Pasto, Colombia
laurad.lasso.a@gmail.com

² Corporación Universitaria Autónoma de Nariño, Pasto, Colombia

³ Instituto Tecnológico Metropolitano (ITM), Medellín, Colombia

⁴ Yachay Tech, Urcuqui, Ecuador

Abstract. The analysis of electromyographic (EMG) signals enables the development of important technologies for industry and medical environments, due mainly to the design of EMG-based human-computer interfaces. There exists a wide range of applications encompassing: Wireless-computer controlling, rehabilitation, wheelchair guiding, and among others. The semantic interpretation of EMG analysis is typically conducted by machine learning algorithms, and mainly involves stages for signal characterization and classification. This work presents a methodology for comparing a set of state-of-the-art approaches of EMG signal characterization and classification within a movement identification framework. We compare the performance of three classifiers (KNN, Parzen-density-based classifier and ANN) using spectral (Wavelets) and time-domain-based (statistical and morphological descriptors) features. Also, a methodology for movement selection is proposed. Results are comparable with those reported in literature, reaching classification performance of $(90.89 \pm 1.12)\%$ (KNN), $(93.92 \pm 0.34)\%$ (ANN) and 91.09 ± 0.93 (Parzen-density-based classifier) with 12 movements.

Keywords: ANN · EMG signals · Feature extraction · KNN · Parzen

1 Introduction

Nowadays, technologies for human-computer interfaces (HCIs) are being widely used, as they are becoming indispensable for people life's activities. Thanks to the development of fields such as pattern identification, signal processing, among others, there have been proposed different tools and systems to control a machine through the use of signals, as it is the case of Electromyographic (EMG) signals. Some outstanding applications of EMG-based HCI are: Wireless-computer

controlling through hand movement, wheelchair directing/guiding with finger motions, and rehabilitation [1,2].

Once acquired and digitalized, and since being electrical-unidimensional-type signals, the ECG signals are analyzed within a process mostly involving stages for preprocessing, characterization, data representation, and classification [3]: Firstly, preprocessing is aimed at filtering (removing artifacts and noisy information) as well as adjusting the signal for subsequent tasks (some processing procedures are sensitive to length or amplitude). Secondly, a characterization task is performed for extracting descriptive features or attributes (especially is required when using directly the amplitude values is unfeasible for length, informative, or computational cost reasons). Thirdly, to select or best represent the afore obtained features in terms of separability, a data representation may be used. Finally, the automatic classification process is applied over the data to assign to relate the EMG signal with an instruction (for instance, a movement in a HCI-based device). All these stages have addressed for some studies [4–6], but a highly accurate system for EMG semantic interpretation is still a problem to be solved.

Based on state-of-the-art techniques, this work proposes a methodology for EMG signal characterization and classification techniques to identify movements. In particular, the spectral features (wavelet coefficients, average frequency, peak frequency, Curtosis, Shanon, Fischer asymmetry coefficient), temporal features and statistics are used (Area under the curve, absolute mean value, effective value, standard deviation, variance, median, Willison amplitude, Histogram, simple square integral entropy) [1,4,7–12]. The characterization of the signal, leaves a matrix of large dimensions so it is necessary to make a dimension reduction. Two processes are performed to achieve a good reduction in size, the first is the selection of movements, proposed methodology of comparison between movements, seeking for which have a greater differentiability and present a lower error in their classification. The second process, is a selection of features, consists of the calculation of contribution of each feature to the classification, is carried out through the WEKA program and its RELIEF algorithm [13,14]. Finally, the performance comparison of three machine learning techniques, k-nearest neighbors (KNN), Artificial neural network (ANN) and classifier based on Parzen density. Each of the stages is developed and explained in depth in the text.

The rest of this paper is structured as follows: Sect. 2 describes the stages of the EMG signal classification procedure for movement identification purposes. Section 3 presents the proposed experimental setup. Results and discussion are gathered in Sect. 4. Finally, Conclusions and future work in Sect. 5.

2 Materials and Methods

This section describes the methodology used to carry out the comparative study and identify a group of features that allows get a high performance of classifiers.

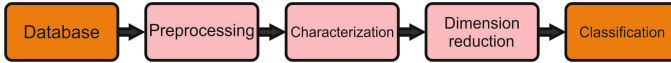


Fig. 1. Block diagram of the proposed methodology

In general, our research contains the following stages: preprocessing, segmentation, characterization, dimension reduction and finally classifiers as shown in the explaining block diagram from Fig. 1.

2.1 Database

For this research, we considered the database that is available *Ninaweb* repository from *Ninapro* project [7]. The *Ninapro* project describes the acquisition protocol and the subjects involved in the data acquisition. The muscle activity is recorded by 10 double differential electrodes, 8 electrodes are located uniformly around the forearm using an elastic band, at a constant distance, just below the elbow. In turn, two additional electrodes are placed in the long flexor and extensor of the forearm. The hand position is taken by a Dataglove and an inclinometer. The sample rate used is for 100 Hz and amplification factor is for 14000.

The *Ninapro* database include the electromyographic activity of upper-limb of 27 healthy people executing 52 movements, which are divided as: 12 movements of fingers, 8 isometric and isotonic configurations of the hand, 9 wrist movements and 23 functional grip movements [4].

2.2 Stages of System

Pre-processing. EMG signals can be affected by several factors, like Extrinsic and Intrinsic factors and others [15]. For the above, amplitude and frequency features are also affected, so it is necessary to compensate the effect of these factors with some procedure, in this research we applied a normalization and segmentation procedure. The normalization was applied to each electrode, we took an electrode and found the maximum value of the signal of this electrode, finally the whole signal is divided with this maximum value. The *Ninapro* database contains a tag vector, which we used for segmentation procedure. The tag vector lets identify when the patient was doing a movement, in addition what kind of movement. So this vector facilitating the trimming of the signals.

Characterization. The matrix of features is organized in the following way: In each row, the data corresponding to patient n is placed, which is performing a movement w and a repetition k and so on. For columns, a bibliographic review is made, obtaining 45 different features for the EMG signals, which are applied to each of the 10 electrodes. As mentioned above, the feature matrix has a size of 14040 per 450. This matrix is organized as shown forward in the next matrix.

$$X = [x_{i,j}] = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,d} \\ x_{2,1} & x_{2,2} & \dots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{M,1} & x_{M,2} & \dots & x_{M,d} \end{pmatrix}$$

where $X \in \mathbb{R}^{M \times d}$ is the matrix of features, M is the number of samples = 14040 corresponding to 270 repetitions $k \times 52$ classes w . d is the total number of features = 450 corresponding to 45 attributes \times 10 electrodes.

Among the features, there are two types, temporal and spectral features. For a better understanding, each one of them is explained below.

- *Temporal features*: Are the variables, which are in the time domain and quantified each T seconds of time. The features used are: Area under the curve, Mean absolute value, Root mean square value, Variance of EMG, Entropy, Simple square integral, Histogram, Willison Amplitude, Waveform length [1, 9, 10, 12, 16]
- *Spectral features*: It give us, the time-frequency representation of a signal where we can obtain a more complete description of the physical phenomenon. The most common techniques used in the extraction of spectral are Curtosis, Shanon, Fischer asymmetry coefficient, Peak frequency, Average frequency, Modified Average Frequency, Wavelet transform [5, 8, 10, 11, 16].

Movements Selection. Many movements can hinder the classification problem, so it is necessary to identify the largest number of classes, which can be classified without problems. By virtue of the above, we suggest a criterion based on the Euclidean distance (ED) and the standard deviation (SD). From the characteristics matrix its calculated the centroid \mathbf{c} of the points cloud (PC) for each movement, as the mean value of the repetitions k , belonging to each y^{th} class. Euclidean Distance of each class its calculated as the root of square differences between co-ordinates of a pair of centroids.

Euclidean distance between centroids allows evaluate how much far away are the classes, and the standard deviation says what so compact is a points cloud. To improve the performance of the classify its necessary that PC of each movement be far away and as compacts as possible. So a parameter that allows to evaluate these two features between two different classes is defined as:

$$P(y_1, y_2) = \frac{ED(\mathbf{c1}, \mathbf{c2})}{SD_1 + SD_2}, \quad (1)$$

where $ED(\mathbf{c1}, \mathbf{c2})$ is the euclidean distance between the centroids of the class 1 (y_1) and class 2 (y_2), SD_1 and SD_2 are the standard deviation of classes 1 and 2. SD must be inversely proportional to parameter in order to maximize P and improve the classification process. The movements for which the parameter P is greater are selected.

The movements that present the best separability are: contraction and extension of middle and ring fingers, wrist contraction and extension, writing tripod

grasp, tip pinch grasp, lateral grasp, power disc grasp and turn a screw. Several of these movements have great activity of the flexor digitorum superficiales and of the extensor digitorum superficiales.

Features Selection. A subset of the original data set is taken, to obtain this subset the Relief algorithm is used. Relief calculates a feature weight for each feature which can then be applied to rank and select top scoring features. RELIEF detects the interactions between attributes to determine those that are most relevant and those that work well in groups.

To select a suitable number of characteristics to train the classifiers, a test is carried out with the KNN classifier. We evaluated the performance of the classifier, varying the number of characteristics. From this test, it can be seen in Fig. 2, that 20 characteristics are enough to obtain a good performance. This 20 features are used for the next step, the comparison of each classifiers.

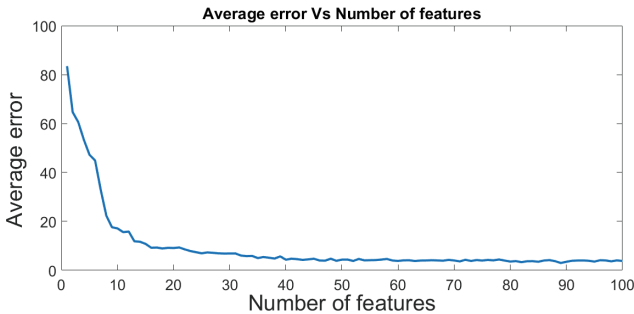


Fig. 2. Results of the selection of features, as the average error varies according to the number of features.

2.3 Classification

With the new feature matrix, the classification of these movements is carried out with the following techniques [1, 6, 8, 17, 18].

1. K-nearest neighbors (KNN): It's a method of non-parametric supervised classification, which utilizes a distance measure relative to the k-closest neighbors of a point to assign a class to a given data record. The number of the nearest neighbors is optimized with respect to the leave-one-out error on training set.
2. Artificial neural network (ANN): This heuristic classification technique emulates the behavior of a biological brain through a large number of artificial neurons that connect and they are activated by means of functions. The neurons are distributed in hidden layers, in this work a neural network is trained with a back-propagation algorithm with a hidden layer with 10 neurons. The weight initialization consists of setting all weights to be zero. Sigmoidal functions are used in the output. The dataset is used for training, validation and test.

3. Parzen-density-based classifier: This probabilistic-based classification method requires a smoothing parameter for the Gaussian distribution computation, which is optimized.

3 Experimental Setup

To carry out the different experiments of this stage, we used the toolbox of Matlab called PrTools. This toolbox has all necessary functions as classifiers and performance indicators. From the reduced feature matrix, we obtain two sets in a random way, each group is balanced in the number of samples. Classifiers are trained with 75% of the data and validated with the remaining 25%.

To ensure the repeatability of the experiment, and that the results do not depend on a specific data group, the tests are repeated 30 times with the three classifiers. In each test the percentage of performance is obtained and at the end of the 30 repetitions an average of the tests is obtained. The results of each test are reported in the next section.

4 Results and Discussion

Based on the tests carried out and the performance results recorded in Fig. 3 it is clear that the ANN present a better performance than KNN and Parzen. Nevertheless all classifiers have error percentage under 10%. The performance of ANN is 93.92%, KNN is 90.89% and PARZEN is 91.09% using 20 features.

Figure 3 also reveals a greater uniformity of the ANN than KNN and Parzen classifiers in each of the tests, giving a standard deviation of 1.12% for KNN, 0.93% for Parzen-density-based classifier and 0.35% for the neural network. This results are comparable to [1], where with back-propagation neural networks a 98.21% performance was obtained but classifying only 5 movements what limits the control options. In [17] the results of pattern recognition EMG signal using wavelet transform and Artificial Neural Network (ANN) classification has an accuracy rate of 77.5% identifying 8 movements. For this reason it is emphasized that an adequate search of EMG signal characteristics in combination with the appropriate selection of movements allows obtaining a good performance of the ANN classifier. The recognition rates were 84.9% for the KNN in [19] where five wrist movements were classified. It is considerable that this result is due to the appropriate selection of movements based on Euclidean Distance criteria, also used by the KNN classifier.

The process of selection of attributes showed that the most relevant features were: Maximum concentration of histogram, standard deviation, Willison amplitude, modified average frequency, entropy of the signal in the time domain and entropy of the first level of decomposition of the wavelet transform. Most of the relevant information is concentrated in 4 electrodes, corresponding to the flexor and extensor muscles of the forearm.

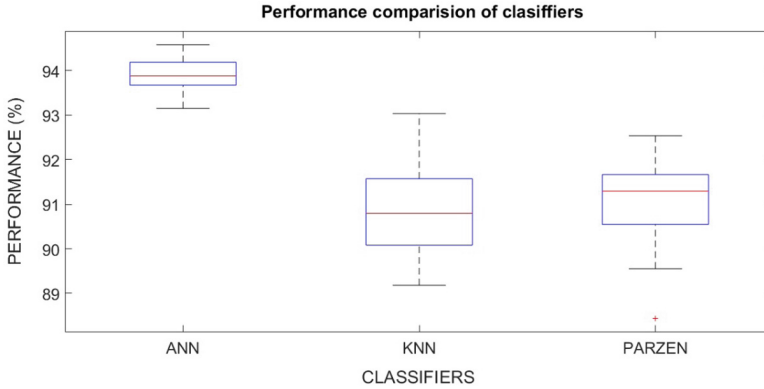


Fig. 3. Performance of the classifiers. In the following order, ANN, KNN and Parzen.

5 Conclusions and Future Work

This work presents a evaluation of different features of signal analysis, a dimension reduction method and a criterion to evaluate separability of classes. With the aim to improve movements classification with EMG signals. Experimentally we prove that the with the selected movements and the features used it is possible to implement a human - machine control interface with good performance. On the other hand a set of features that help in the movement recognition of upper limb is proposed. We identify 12 movements that with the characteristics described in Sect. 2 present a greater separability, it was also possible to identify the muscles that have the most intervention on these movements.

As a future work, we will explore the possibility to apply this knowledge in a prosthesis prototype with own signals of EMG.

Acknowledgments. This work is supported by the “Smart Data Analysis Systems - SDAS” group (<http://sdas-group.com>), as well as the “Grupo de Investigación en Ingeniería Eléctrica y Electrónica - GIIEE” from Universidad de Nariño. Also, the authors acknowledge to the research project supported by Agreement No. 095 November 20th, 2014 by VIPRI from Universidad de Nariño.

References

1. Phinyomark, A., Phukpattaranont, P., Limsakul, C.: A review of control methods for electric power wheelchairs based on electromyography signals with special emphasis on pattern recognition. *IETE Tech. Rev.* **28**(4), 316–326 (2011)
2. Aguiar, L.F., Bó, A.P.: Hand gestures recognition using electromyography for bilateral upper limb rehabilitation. In: 2017 IEEE Life Sciences Conference (LSC), pp. 63–66. IEEE (2017)
3. Rodriguez-Sotelo, J., Peluffo-Ordóez, D., Cuesta-Frau, D., Castellanos-Domnguez, G.: Unsupervised feature relevance analysis applied to improve ECG heartbeat clustering. *Comput. Methods Programs Biomed.* **108**(1), 250–261 (2012)

4. Atzori, M., et al.: Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Sci. Data* **1**, 140053 (2014)
5. Podrug, E., Subasi, A.: Surface EMG pattern recognition by using DWT feature extraction and SVM classifier. In: *The 1st Conference of Medical and Biological Engineering in Bosnia and Herzegovina (CMBEBIH 2015)*, March 2015, pp. 13–15 (2015)
6. Vicario Vazquez, S.A., Oubram, O., Ali, B.: Intelligent recognition system of myoelectric signals of human hand movement. In: Brito-Loeza, C., Espinosa-Romero, A. (eds.) *ISICS 2018. CCIS*, vol. 820, pp. 97–112. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-76261-6_8
7. Atzori, M., et al.: Characterization of a benchmark database for myoelectric movement classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **23**(1), 73–83 (2015)
8. Krishna, V.A., Thomas, P.: Classification of EMG signals using spectral features extracted from dominant motor unit action potential. *Int. J. Eng. Adv. Technol.* **4**(5), 196–200 (2015)
9. Negi, S., Kumar, Y., Mishra, V.: Feature extraction and classification for EMG signals using linear discriminant analysis. In: *International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Fall)*, pp. 1–6. IEEE (2016)
10. Phinyomark, A., Limsakul, C., Phukpattaranont, P.: A novel feature extraction for robust EMG pattern recognition. *CoRR* abs/0912.3973 (2009)
11. Ahlstrom, C., et al.: Feature extraction for systolic heart murmur classification. *Ann. Biomed. Eng.* **34**(11), 1666–1677 (2006)
12. Han, J.S., Song, W.K., Kim, J.S., Bang, W.C., Lee, H., Bien, Z.: New EMG pattern recognition based on soft computing techniques and its application to control of a rehabilitation robotic arm. In: *Proceedings of 6th International Conference on Soft Computing (IIZUKA2000)*, pp. 890–897 (2000)
13. Kononenko, I.: Estimating attributes: analysis and extensions of RELIEF. In: Bergadano, F., De Raedt, L. (eds.) *ECML 1994. LNCS*, vol. 784, pp. 171–182. Springer, Heidelberg (1994). https://doi.org/10.1007/3-540-57868-4_57
14. Kira, K., Rendell, L.A.: A practical approach to feature selection. In: *Machine Learning Proceedings 1992*, pp. 249–256. Elsevier (1992)
15. Halaki, M., Ginn, K.: Normalization of EMG signals: to normalize or not to normalize and what to normalize to? (2012)
16. Romo, H., Realpe, J., Jojoa, P., Cauca, U.: Surface EMG signals analysis and its applications in hand prosthesis control. *Revista Avances en Sistemas e Informática* **4**(1), 127–136 (2007)
17. Arozi, M., et al.: Electromyography (EMG) signal recognition using combined discrete wavelet transform based on artificial neural network (ANN). In: *International Conference of Industrial, Mechanical, Electrical, and Chemical Engineering (ICIMECE)*, pp. 95–99. IEEE (2016)
18. Shin, S., Tafreshi, R., Langari, R.: A performance comparison of hand motion EMG classification. In: *2014 Middle East Conference on Biomedical Engineering (MECBME)*, pp. 353–356. IEEE (2014)
19. Kim, K.S., Choi, H.H., Moon, C.S., Mun, C.W.: Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions. *Curr. Appl. Phys.* **11**(3), 740–745 (2011)