Research Journal of Applied Sciences 13 (9): 482-490, 2018
ISSN: 1815-932X

Hand Gesture Recognition Using Electromyographic Signals Throw a Deep Convolutional Neural Network

Javier O. Pinzon Arenas, Robinson Jimenez Moreno and Ruben D. Hernandez Beleno
Faculty of Engineering, Nueva Granada Military University, Bogota, Colombia

Abstract: This study presents the implementation of a convolutional neural network focused on the recognition of hand gestures for this case 3 specific types of gestures using the EMG signals as input which were acquired through the Myo armband device and processed by means of a characteristic map extraction technique which is the power spectral density. The development of this work is divided into 2 phases where the first consists of the acquisition and processing of the electromyographic signals of different users with different arm thickness from which 2 databases were built and the second phase describes the architecture of the convolutional neural network to be used and the training that was performed with each database independently, obtaining two trained networks. Finally, two types of tests are performed, a validation test in which the accuracy of the two trained networks is verified where a accuracy rate of 91.7 and 92.5% was achieved and a real-time behavioral test where the two networks responded adequately, meaning that the use of convolutional neural networks for the recognition of hand gestures by means of electromyographic signals can reach high ranges of accuracy, even greater than 90%.

Key words: Deep convolutional neural network, power spectral density, electromyographic signal, hand gesture recognition, Myo armband

INTRODUCTION

During the last decade, different advances have been made in the interaction between human and machine, generating a great variety of techniques and devices to set communication between them. One of the most investigated techniques is the use of Electromyographic (EMG) signals with which several developments have been carried out in the fields of medicine, health care and robotic control where these three have been synergistically joined to advance in the elements that have been implemented. An example of this is found in the work of Masson *et al.* where EMG signals are used to control a robotic prosthesis of an upper body limb using a device called Myo armband developed by Thalmic Labs™ or as the one presented by Noce *et al.* (2016) that use electromyographic signals for the control of a one-handed prosthesis for an amputated person.

As indicated by Singer and Goldin-Meadow (2005) the interaction with gestures made by hands is an important part of communication between people, so, research focused on the use of these gestures for human-machine interaction has been of great interest, even one of the main functions of Myo armband is the predetermined recognition of 5 different gestures performed by the hand. This function has allowed to implement the control of robotic arms (Murillo *et al.*, 2016), navigation in graphical user interfaces (Mulling and Sathiyanarayanan, 2015) or even adding more gestures as it was done by Abreu *et al.* (2016) where different gestures for the recognition of the alphabet were integrated by means of sign language using the EMG signals obtained by the Myo. The Myo has also, allowed to make an effort to extract EMG signal characteristics as proposed by Arief *et al.* (2015) where a comparison of 5 different extraction techniques is done in order to reduce the complexity of an unprocessed signal, due to the difficulty of processing such signals for the machine learning applications which at the same time, increase the computational cost required.

On the other hand, due to the development and effectiveness of techniques such as Deep Learning (DL), it has recently begun to combine this with the use of EMG signals. An example is shown byWand and Schultz, (2014) where DL is used through the implementation of a Deep Neural Network (DNN) to perform silent speech recognition through EMG signals, i.e., by means of muscular movements at the moment of speaking, it classifies the phonemes that are being used. In order to be able to use the electromyographic signals as input of the neural network in that case a feature extraction was made by means of the logarithmic power of the signal.

Corresponding Author: Javier O. Pinzon Arenas, Faculty of Engineering, Nueva Granada Military University, Bogota, Colombia

However, there are other types of techniques in deep learning such as Convolutional Neural Networks (CNN) which are specialized in pattern recognition. Although, they were initially implemented for the recognition of patterns in images as shown in the research of Simonyan and Zisserman (2014) and Krizhevsky *et al.* (2012) they have begun to strengthen in the recognition of patterns in signals for example in speech recognition (Abdel-Hamid *et al.*, 2014) by extracting features of audio signals. Considering this, it is possible to think that the functionality of the CNN can be implemented in the analysis of the electromyographic signals for the recognition of hand gestures.

The novelty of this research is the use of two techniques that have not been applied to the recognition of hand gestures by means of EMG signals which are the extraction of characteristics by power spectra and classification by convolutional neural networks, so that, its use can be focused on the human-machine interaction.

This research is focused on the recognition of 3 different gestures of the hand through convolutional neural networks which are "open", "closed for grip" and "closed for use" where these gestures are used to recognize if the user needs a tool when he receives it and when he accepts the tool to be used. For this, processed electromyographic signals obtained from the right forearm are used, to then be entered into a convolutional neural network and thus, to recognize what gesture the user is performing.

The development of this work was carried out in 2 phases. The first phase consisted in building the database of the electromyographic signals of each gesture, for which it was first necessary to perform the acquisition of the data of the signals and then to obtain their feature maps. The second phase corresponded to the implementationof the convolutional neural network architecture (Krizhevsky *et al.*, 2012) and its training. Finally, the validation and use tests in real-time are presented in order to reach the conclusions.

MATERIALS AND METHODS

Phase 1; Data acquisition and database build: In order to perform the training of a neural network in the first instance it is required to obtain the electromyographic signals that will be processed. For this, the Myo armband device of Thalmic Labs™ is used which is a bracelet that has 8 biosensors that measure the Electromyographic signals (EMG) of the muscles of the forearm according to the movements that are made by the hand, additionally, it contains an Inertial Measurement Unit (IMU) of 9 axes of which 3 axes belong to a gyroscope, 3 to an accelerometer and 3 to a magnetometer (Thalmic Labs Inc., 2017a). The sensors are distributed in eight segments joined by an elastic material which facilitates its use allowing to stretch or contract for the user's comfort (Thalmic Labs Inc., 2017b). The physical structure of Myo is illustrated in Fig. 1.

In order to collect the data in a similar way for all users, the Myo must be located in the right arm of the user as shown in Fig. 2, so that, the sensors take the electric signals approximately from the same muscle. However, since, the thickness of the arm is variable in everyone, the location of the muscles is different in each one as it is indicated by Murillo *et al.* (2016), so, it is necessary to observe which sensors have a similar behavior in each gesture and do not generate excessive

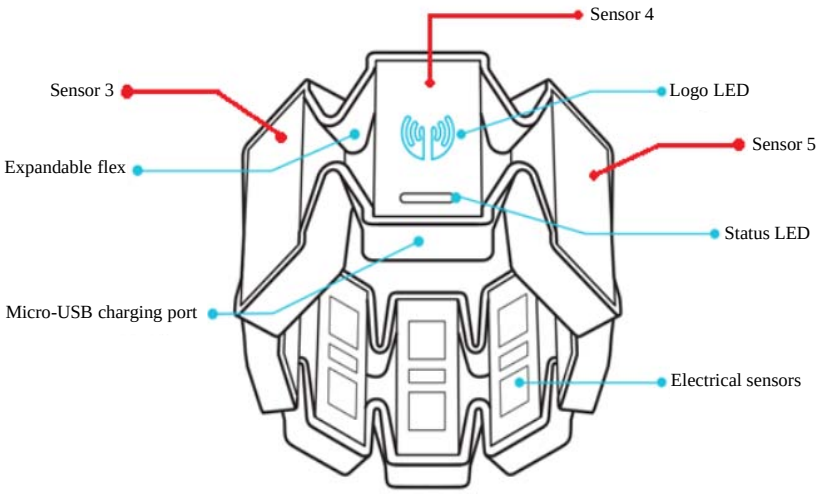

Fig. 1: Myo armband components by Thalmic Labs Inc., (2017)

noise which involves analyzing the movements that are required in conjunction with the signals acquired and thus, choose 3 sensors to build the database.

For the acquisition of the Myo armband signals, MATLAB® Software is used in conjunction with a toolbox created in previous works which allows the visualization of the signals of the sensors without any previous processing, called Myo data acquisition toolbox. The toolbox interface, shown in Fig. 3 has the options of constantly sampling the acquired signals or taking a sample in a specific time range. It should be noted that the signals are sampled at a rate of 200 Hz for EMGs and 50 Hz for IMUs.

In order to make the selection of the Myo EMG sensors to be used, the comparison of each of the sensors is made in arms of different thickness, making the 3 movements that are wanted to recognize. An example of this comparison is illustrated in Fig. 4 which shows different hand gestures tests to observe their behavior in three subjects. As can be seen, the sensors with the best behavior are 1, 2, 5 and 6, since, they generate less noise,

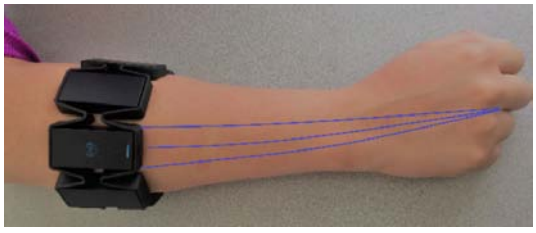

Fig. 2: Location of the Myo armband ih the right forearm

especially when the hand is open and the behavior of the signal is similar for each case in each type of gesture.The sensor 8 also had a good behavior for thick and medium-thickness users but for thin it did not react adequately and the signal did not significantly vary.

Once the sensors have been chosen, samples with different subjects are made by performing each gesture in a range of 2 sec per gesture. For this, the option of taking a sample in a specific time range of the toolbox is used, selecting the chosen sensors and writing the desired sampling time generating a matrix of 400×5 where 400 are the data obtained and 5 refers to 1 column of the sampling time and the 4 sensors chosen. With this, a database of each gesture is obtained, however, being a raw signal its features are not, so, evident to be the input of a CNN, since, relatively the signals would be seen almost as a random element by the network. Therefore, to obtain a more suitable input type, it is performed a feature extraction processing of the signals obtained from each user.

The feature extraction of the signal is done by means of the Power Spectral Density or PSD using the Welch's method, developed by Welch (1967) which represents the amount of energy that is in the signal in each frame. For the analysis, frames of 50 msec are used, due to the sampling condition of Myo signals (200 Hz) which obtains the signals every 5 msec, therefore, 10 samples are taken per frame to analyze their PSD. With this, it is obtained 101 coefficients per frame, representing the power every 1 Hz from 0-100 Hz, taking into account that the sampling

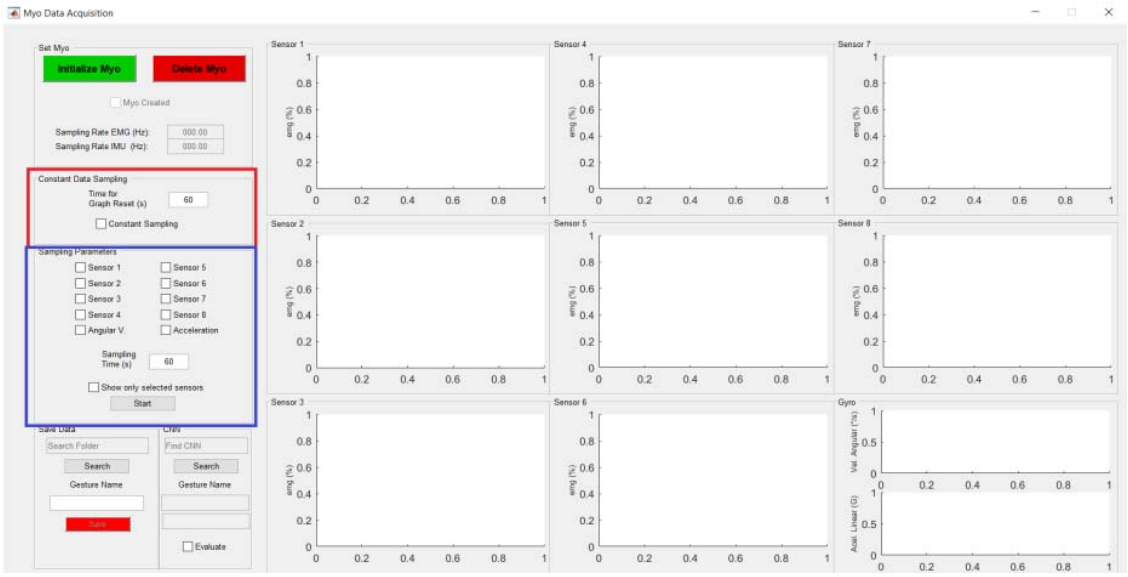

Fig. 3: Graphic interface of the Myo data acquisition toolbax

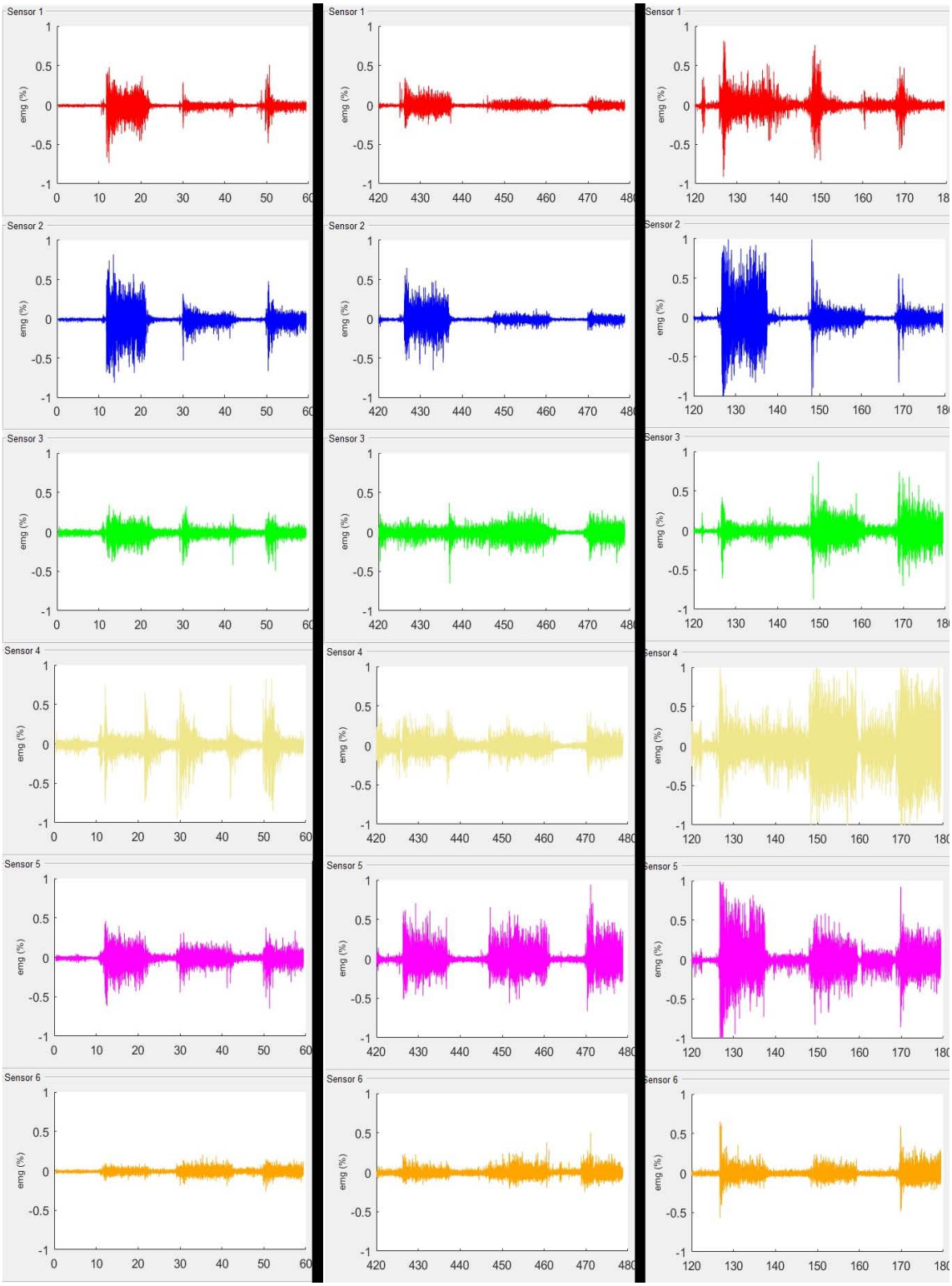

Fig. 4: Continue

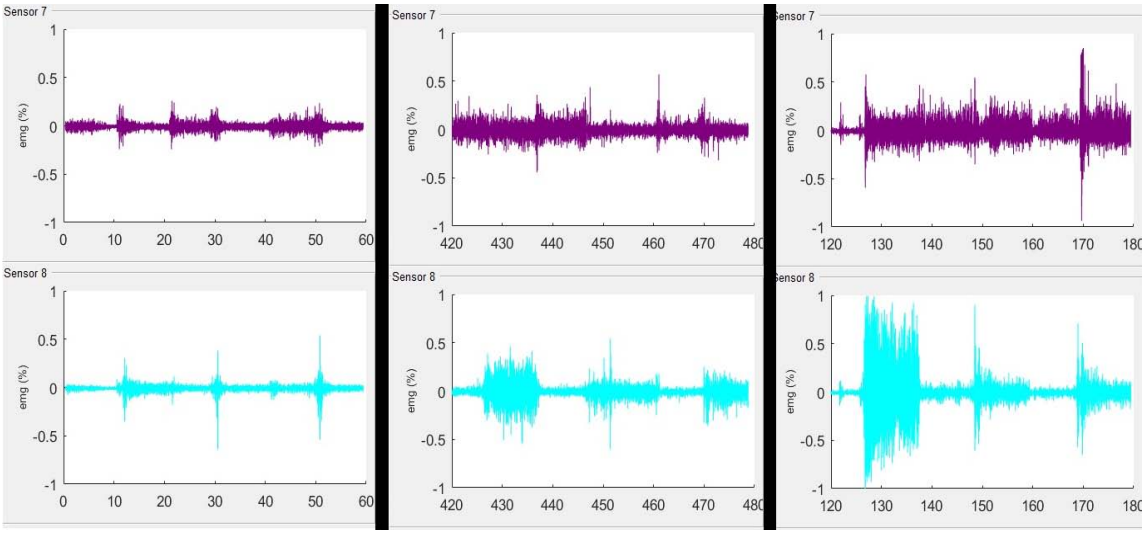

Fig. 4: Test for 3 kind of arm thickness (thin, medium and thick) where each user initially keep the hand open for approximately 10 sec then closes the hand as grasping a tool for 10 sec, following this opens the hand again and then closes the hand for tool use, repeating this last sequence one more time

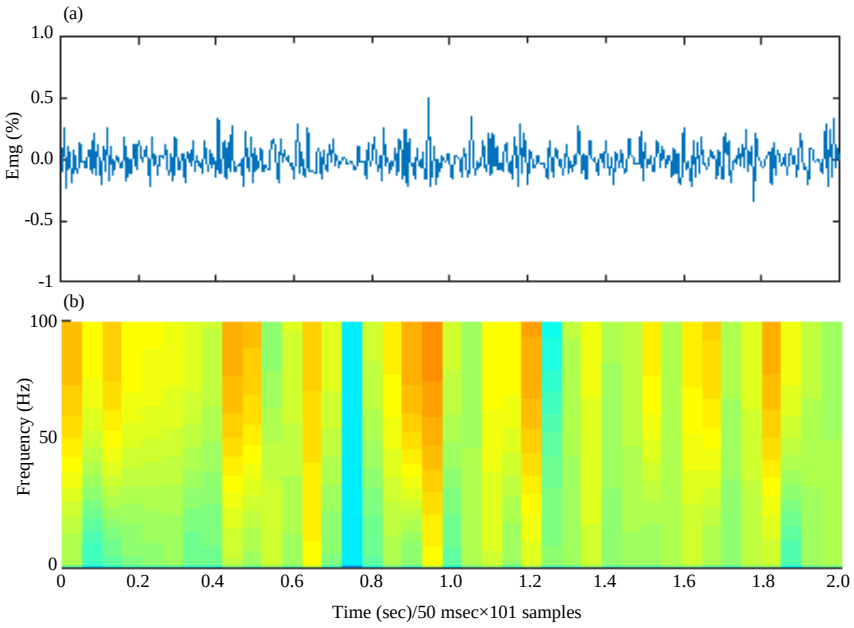

Fig. 5: Signal plot (upper graphic) and feature map of the signal (lower graphic) of the sensor 5 for the gesture "Closed for use"

frequency is at least double the signal frequency (sampling theorem), obtaining a feature map of 40 frames as shown in Fig. 5.

Performing the extraction of characteristics, it is obtained a matrix of 101×40 for each sensor, however, to feed the CNN, an emulation of a matrix composed by the RGB components of an image is made where by a matrix is created with the feature maps of 3 sensors of the 4 chosen, therefore, 2 databases are made, one with the data of the sensors 1, 2 and 5 (database 1) and another with the data of the sensors 6, 2 and 5 (database 2).

In total, two training database of 400 samples each was obtained where 200 samples are ofclosed for use (Closed_Use), 100 of closed for grasp (Closed_Grasp) and

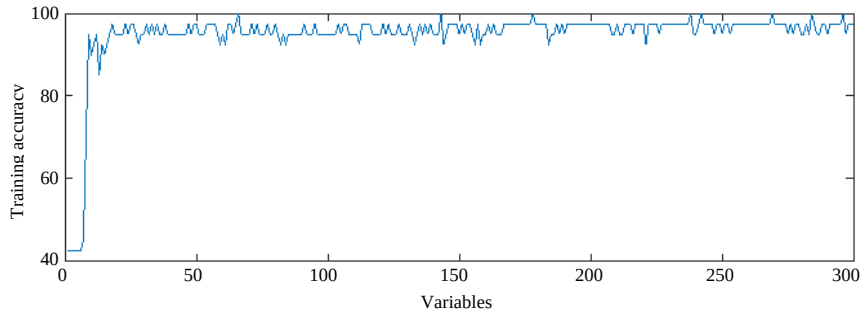

Fig. 6: Training behavior of the network trained with the database 1

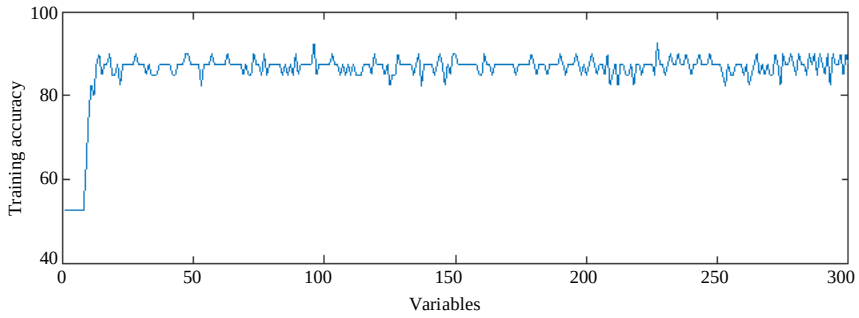

Fig. 7: Training behavior of the network trained with the database 2

100 of open hand (Open). The reason for using more samples in the "Closed_Use" category is because it may tend to create confusion between the two additional categories, depending on the arm thickness of the person and how the Myo armband is worn.

Phase 2; Architecture implemented for the convolutional neural network: The architecture of the convolutional neural network is trained with the two databases built to observe which one has the best behavior. This architecture is configured with two types of convolutional filters: frequency analysis filters which will extract the characteristics of the map in terms of frequency and combined analysis filters, i.e., a square filter which will analyze both frequency and time of the map, allowing to extract together behavioral characteristics of the frequency changes over time. The implemented architecture is found in Table 1 in which groups of convolution analysis combined plus convolution in frequency are made to obtain better extraction of characteristics, maintaining the size of the entrance volume by means of zero-padding, except in the last stage of the network where only a combined convolution is performed. Additionally, a downsampling is made after each group in order to reduce the input volume to the next group of convolutions for a more detailed feature analysis.

Table 1: Architecture implemented each layer of the network architecture is composed by an amount of k filters with a specific kernel, i.e., a determined size N×M of the filter, a stride S which is the step that the filter is moved and a padding P that is the amount of files added at each border of the input volume

Layer	----------Kernel----------		
Input	----------101×40×3----------		Filters
Convolution	5×5	S =1/P = 2	32
Convolution	6×1	S = 1/P = 2×0	32
Max pooling	2×2	S = 2	-
Convolution	3×3	S = 1/P = 1	64
Convolution	5×1	S = 1/P = 1	64
Max pooling	2×2	S = 2	-
Convolution	3×3	S = 1/P = 1	256
Max pooling	2×2	S = 2	-
Fully-connected		1	512
Fully-connected		1	-
Softmax		12	-

With the architecture implemented, the training is done with the two databases obtained, training under the same learning parameters and the same number of epochs, obtaining the training graphs of Fig. 6 and 7. As can be observed, the network trained with database 1 had a better performance than the network of database 2 where the accuracy of training of the first managed to obtain more than 95% of accuracy while the second stabilized at approximately 87.5%, showing a lower recognition during the training process. Although, 300 epochs were used to perform network training, the two networks managed to achieve their stabilization in a high degree of training

accuracy in <25 epochs which also shows CNN's high ability to recognize signal patterns. However, a validation test should be performed with data that is not included in the training databases.

RESULTS AND DISCUSSION

Validation test: To validate each of the trained networks it is used two additional databases of 120 samples in total each, not belonging to database 1 and 2 where 60 samples are of closed for use (Closed_Use), 30 of closed for grasp (Closed_Grasp) and 30 of open hand (Open). With these validation databases, the confusion matrices of Fig. 8 are obtained where it can be observed that the two networks got an accuracy of recognition of the gestures in a percentage greater than 90%.

However, although, the network trained with database 2 had a less favorable behavior in its training, obtained a 0.8% more accuracy than the other network in the validation. What is more marked in the mistakes made by the network with database 1 are the mistakes with the category "Open" where it had 6 misclassifications which is due to the fact that in users with thick arm, the sensors, mainly sensor 1 and 5, tend to generate a sufficient amount of noise to raise the percentage of the electromyographic signal to values close to those registered in the "Closed_Use" category in users with thin arms which can be observed in the tests performed shown in Fig. 4. Something similar happens with errors in the "Closed_Grasp" category, since, a user with a very thin arm, can generate electromyographic readings close to the one generated by a medium shape arm in the "Closed_Use" category.

Real-time test: For the real-time testing it was used the implemented toolbox. In the test first, the user keeps their hand open for a tool, once it is delivered they close the hand handling the tool as a delivery signal then change the position of the hand to use the tool in this way, the 3 gestures categories are tested. These tests are mainly done to observe how the trained networks behave while a user is wearing the Myo and if when the gesture is made, the algorithm responds with the correct classification, clarifying that the network classifies every 2 sec the gesture made this because the input of the network is a feature map made for a sample of 2 sec of the EMG signal.

Figure 9 and 10 show two examples of tests performed with the network trained with database 1 for a user with medium-arm and with database 2 for a user of thick-arm, respectively.

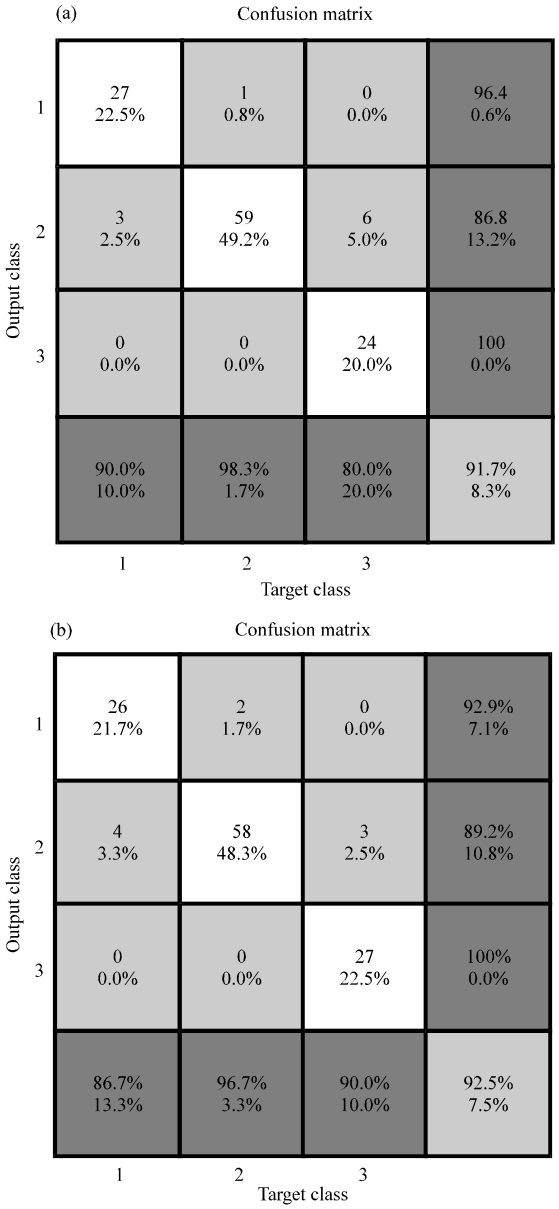

Fig. 8: Confusion matrix of the network trained with: a) database 1 and b) database 2 where target and output class 1-3 are Closed_Grasp, Closed_Use and open, respectively

As it can be seen in Fig. 9 and 10, although, the two subjects have different arm thickness and their signals on the 3 sensors chosen for each network are different in each user (mostly by comparing the signal strength of sensors 2 and 5), the two networks responded adequately, correctly classifying each type of hand gesture at the moment the user made the gesture.

Although, the two trained networks responded with very good behavior and had a high degree of accuracy in

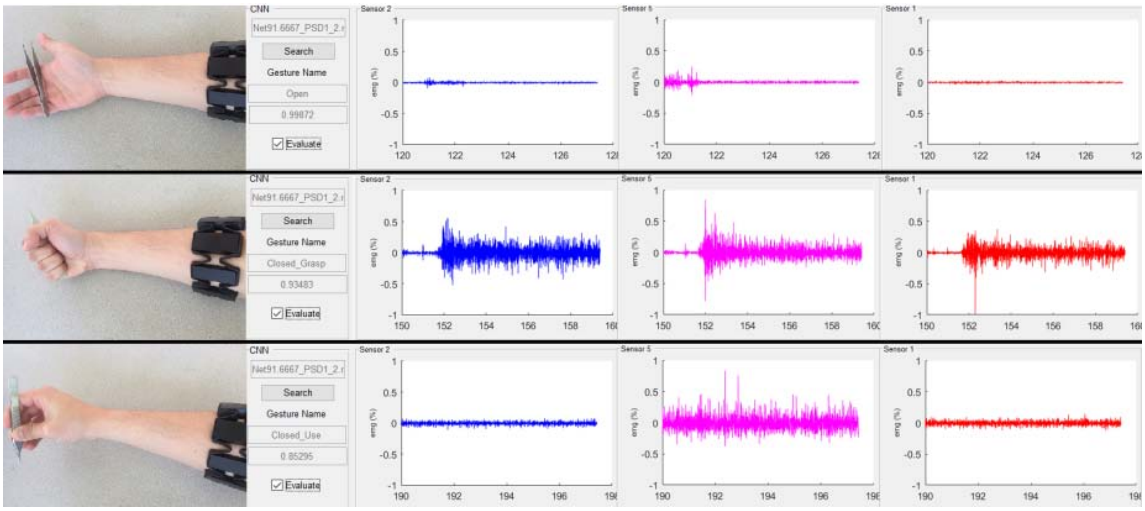

Fig. 9: Real-time test made for a medium-shape subject using the network trained with the database 1

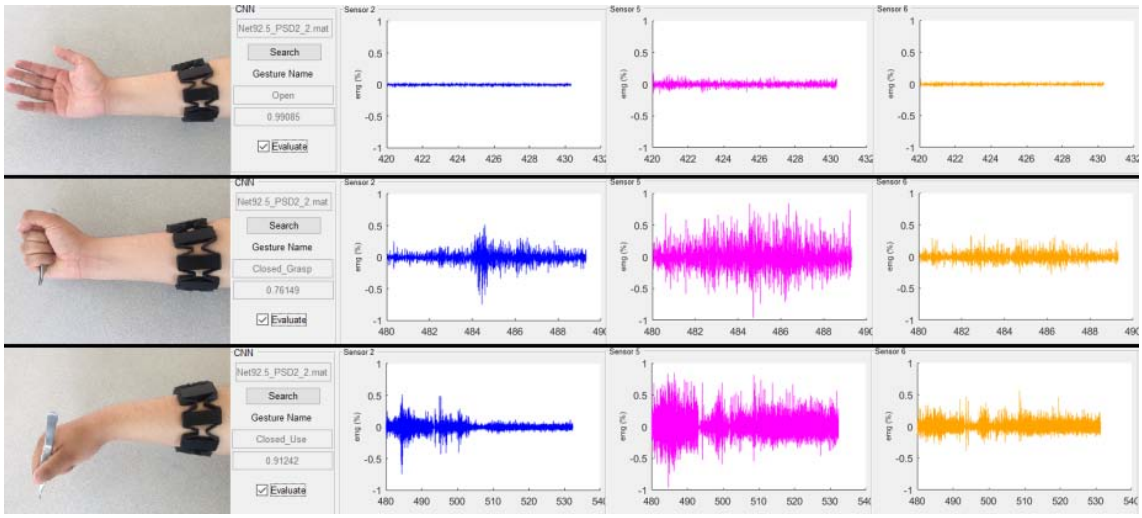

Fig. 10: Real-time test made for a medium-shape subject using the network trained with the database 2

validation, the 0.8% gap, although, small is a difference that can mean a faster correct classification, avoiding to confuse the gestures in more cases and a time reduction and effort made of the user when making the gesture. Therefore, in general, the best trained network is where database 2 was used for training.

CONCLUSION

It was implemented a new way of analyzing the electromyographic signals for the recognition of hand gestures, through the use of convolutional neural networks which in turn, extends the variety of applications where this type of neural network can be used.

The use of an architecture based on frequency and time-frequency analysis allows a better extraction of patterns from the electromyographic signals used for a correct classification of the hand gesture, even using different sensors to train the same architecture, achieving a recognition accuracy of more than 90% in the two tests performed which is a degree of recognition suitable for applications of collection and use of tools. Given that the Myo base application recognizes the categories "Closed_Use" and "Closed_Grasp" as only "Closed", it can be observed that this neural network also, allows to classify similar gestures with each other only using 3 of the 8 sensors that the bracelet possesses.

For the acquisition of EMG through Myo armband, it is necessary to take into account the arm thickness of the different people who can use the bracelet in the application to be implemented, the position in which it is to be worn and the sensors to be used this is due to the fact that the reading of the signals varies because of these causes which can lead to an erroneous acquisition of data, either for the building of the database for the training of the network or a bad recognition due to the misleading data that is entered into the network, i.e. that for instance the network has been trained with sensors 1-3 and it is entered the data obtained from sensors 2-4.

Considering the results obtained, convolutional neural networks can also be used in the analysis of signals that may be viewed as random but using a suitable extraction technique, the network is able to recognize signal patterns, identifying it in a predefined category with a high percentage of accuracy. Also, gesture recognition through EMG using CNN can be used in a wide variety of applications, whether for rehabilitation, control or robotic assistance.

On the other hand, if it is desired to reduce the error in the recognition this technique can be used in combination with others such as with machine vision or correlation between different signals of the same Myo.

ACKNOWLEDGMENTS

The researcher are grateful to the Nueva Granada Military University which through its Vice chancellor for research, finances the present project with code IMP-ING-2290 and titled “Prototype of robot assistance for surgery” from which the present work is derived.

REFERENCES

Abdel-Hamid, O., A.R. Mohamed, H. Jiang, L. Deng and G. Penn et al., 2014. Convolutional neural networks for speech recognition. IEEE. ACM. Trans. Audio Speech lang. Process., 22: 1533-1545.

Abreu, J.G., J.M. Teixeira, L.S. Figueiredo and V. Teichrieb, 2016. Evaluating sign language recognition using the Myo armband. Proceedings of the 2016 18th International Symposium on Virtual and Augmented Reality (SVR), June 21-24, 2016, IEEE, Gramado, Brazil, ISBN:978-1-5090-4149-7, pp: 64-70.

Arief, Z., I.A. Sulistijono and R.A. Ardiansyah, 2015. Comparison of five time series EMG features extractions using Myo Armband. Proceedings of the 2015 International Symposium on Electronics (IES), September 29-30, 2015, IEEE, Surabaya, Indonesia, ISBN:978-1-4673-9345-4, pp: 11-14.

Krizhevsky, A., I. Sutskever and G.E. Hinton, 2012. Imagenet Classification with Deep Convolutional Neural Networks. In: Advances in Neural Information Processing Systems, Leen, T.K., G.D. Thomas and T. Volker (Eds.). MIT Press, Cambridge, Massachusetts, USA., ISBN:0-262-12241-3, pp: 1097-1105.

Mulling, T. and M. Sathiyanarayanan, 2015. Characteristics of hand gesture navigation: A case study using a wearable device (MYO). Proceedings of the 2015 International Conference on British HCI, July 13-17 2015, ACM, Lincoln, Lincolnshire, UK., ISBN:978-1-4503-3643-7, pp: 283-284.

Murillo, P.U., R.J. Moreno and O.F.A. Sanchez, 2016. Individual robotic arms manipulator control employing electromyographic signals acquired by Myo armbands. Int. J. Applied Eng. Res., 11: 11241-11249.

Noce, E., L. Zollo, A. Davalli, R. Sacchetti and E. Guglielmelli, 2016. Experimental analysis of the relationship between neural and muscular recordings during hand control. Proceedings of the 2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob), June 26-29, 2016, IEEE, Singapore, Singapore, ISBN:978-1-5090-3287-7, pp: 1104-1109.

Simonyan, K. and A. Zisserman, 2014. Very deep convolutional networks for large-scale image recognition. Master Thesis, Cornell University, Ithaca, New York.

Singer, M.A. and S. Goldin-Meadow, 2005. Children learn when their teacher's gestures and speech differ. Psychol. Sci., 16: 85-89.

Thalmic Labs Inc., 2017a. Myo SDK manual: Getting started. Thalmic Labs Inc., Kitchener, Ontario. https://developer.thalmic.com/docs/api_reference/platform/getting-started.html.

Thalmic Labs Inc., 2017b. Technical specifications. Thalmic Labs Inc., Kitchener, Ontario. https://www.myo.com/techspecs.

Wand, M. and T. Schultz, 2014. Pattern learning with deep neural networks in EMG-based speech recognition. Proceedings of the 2014 36th Annual International Conference on Engineering in Medicine and Biology Society (EMBC), August 26-30, 2014, IEEE, Chicago, Illinois, ISBN:978-1-4244-7929-0, pp: 4200-4203.

Welch, P.D., 1967. The use of fast Fourier transform for the estimation of power spectra: A method based on time-averaging over short, modified periodograms. IEEE Trans. Audio Electroacoust., 15: 70-73.