engineering proceedings

MDPI

Proceeding Paper

Gesture Recognition Using Electromyography and Deep Learning †

Daniel Gómez-Verde, Sergio Esteban-Romero, Manuel Gil-Martín * and Rubén. San-Segundo

Speech Technology and Machine Learning Group, Information Processing and Telecommunications Center, E.T.S.I. Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain; d.gverde@alumnos.upm.es (D.G.-V.); sergio.estebanro@upm.es (S.E.-R.); ruben.sansegundo@upm.es (R.S.-S.)

* Correspondence: manuel.gilmartin@upm.es; Tel.: +34-910672500

† Presented at The 11th International Electronic Conference on Sensors and Applications (ECSA-11), 26–28 November 2024; Available online: https://sciforum.net/event/ecsa-11.

Abstract: Human gesture recognition using electromyography (EMG) signals holds high potential in enhancing the functionality of human-machine interfaces, prosthetic devices, and sports performance analysis. This work proposes a gesture classification system based on electromyography. This system has been designed to improve the accuracy of forearm gesture classification by leveraging advanced signal processing and deep learning techniques to optimize classification accuracy. The system is composed of two main modules: a signal processing module able to perform two main transforms (Short-Time Fourier Transform and Constant-Q-Transform) and a classification module based on Convolutional Neural Networks (CNNs). The dataset employed in this study "Latent Factors Limiting the Performance of sEMG-Interfaces" comprises EMG signals collected via a bracelet equipped with 8 distinct sensors, capable of capturing a wide range of forearm muscle activities. The experimental process is composed of two main phases. Firstly, we employed a k-fold cross-validation methodology to systematically assess and validate the model's performance across different subsets of the data for hyperparameter tunning. Secondly, the best system configuration was evaluated over a new subset reporting significant improvements. The baseline neural network architecture reported an accuracy of 85.0 ± 0.13% in classifying gestures. Through rigorous hyperparameter tuning and the application of various mathematical transformations to the EMG features, we managed to enhance the classification accuracy to 90.0 ± 0.12% (an absolute improvement of 5% compared to the baseline for a 5-class problem). When comparing to previous works, we significantly improved the F-score from 85.5%, to 89.3% for a 4-class problem (left, right, up, and down).

Keywords: human gesture recognition; electromyography (EMG) signals; human-machine interfaces; signal processing; deep learning; forearm gesture classification; neural network; hyperparameter tuning; k-fold cross-validation; confusion matrix; feature extraction; Convolutional Neural Networks (CNNs); MYO Thalmic Bracelet

Citation: Gómez-Verde, D.; Esteban-Romero, S.; Gil-Martín, M.; San-Segundo, R. Gesture Recognition Using Electromyography and Deep Learning. *Eng. Proc.* **2024**, *6*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Published: 26 November 2024

1. Introduction

Electromyography (EMG) is a technique used to evaluate and record the electrical activity generated by muscles. This activity is translated into an electrical signal using several electrodes situated in different positions of the muscles. There is a high variability of EMG signals within the same subject and across different subjects, due to the characteristics of the tissues and the recording electrodes, meaning that the area of signal decoding still has room for improvement. [1]

The classification of gestures based on EMG signals has attracted significant interest due to its potential applications in human-machine interfaces [2,3], rehabilitation [4], and sports performance analysis [5]. Surface electromyography (sEMG) is commonly used in exercise physiology, sports [6], and various applications where it is necessary to study

nerve activity. The electrical activity generated in the muscles can be used as good biomarkers of nerve activity. Muscles' activity can be used to control different devices such as games or exoskeletons, providing new possibilities of human computer interaction [7,8]. Accurately classifying gestures using EMG signals can greatly enhance the functionality of prosthetic devices, improve user experience in interactive systems, and offer valuable insights into muscle performance and fatigue [9,10].

Th application of deep learning for processing EMG signals has produced significant improvements in the development of classification and analysis systems focused on these signals [11]. Convolutional Neural Networks (CNNs) have proven to be effective in extracting relevant features from raw EMG data and classifying different types of muscle activities [12]. However, the variability in EMG signals across different users and the complexity of accurately interpreting these signals are important challenges in the field.

The main objective of this study is to develop a system for classifying forearm gestures EMG data. This objective involves the design of the deep neural network, the fine-tuning of hyperparameters, the assessment of various mathematical transformations for feature extraction and the analysis of the results to evaluate the best alternative.

2. Materials and Methods

In the next subsections, we will describe the main materials (datasets) and methods (different algorithms) used in this study.

2.1. Dataset

The dataset used in this study has been described in the article "Latent Factors Limiting the Performance of sEMG-Interfaces" [13]. This dataset is described as "files of raw EMG data recorded by MYO Thalmic Bracelet" [14]. The dataset includes EMG recordings at a sampling rate of 1 kHz.

For data acquisition, Lobov et al. developed a hardware-software system called MyoCursor [15]. The system includes a MYO Thalmic bracelet [16] placed on the user's forearm. This device is connected by Bluetooth to a computer that receives and records 8 signals simultaneously. These 8 signals are generated from the 8 sensors included in the bracelet.

The dataset includes recording from 36 different subjects who performed two series of several gestures (six or seven per sequence). The duration of a gesture is around three seconds, followed by a 3-s pause between consecutive gestures. The number of instances is about 40,000–50,000 samples per sensor, and it does not have any missing values. Additionally, new recordings were collected while playing Pacman with the bracelet along different days in several weeks.

The data was recorded at various time points, and the values for each channel were in scientific notation, reflecting the precise measurements of muscle activity. This dataset includes recordings of 8 different forearm movements (Unmarked data, Wrist flexion and extension, Hand at rest or in a fist, Radial and Ulnar deviations, Extended palm). The recording time for each movement is shown in the Table 1:

Table 1. Table of recordings of each gesture, without considering unmarked data.

Movements	Recording Time
Hand at rest or in a fist	208,600 ms and 200,100 ms
Wrist flexion and extension	206,600 ms and 209,200 ms
Radial and Ulnar deviations	209,400 ms and 209,600 ms
Extended palm	7100 ms

Figure 1 show the main gestures considered in this dataset.

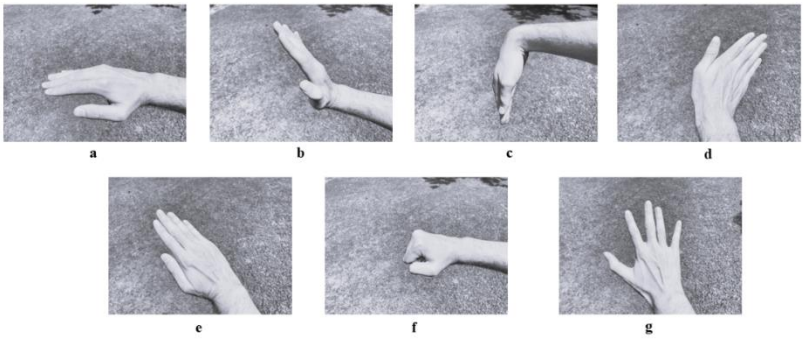

Figure 1. Hand movements: hand at rest (**a**), wrist extension (**b**) and flexion (**c**), radial (**d**) and ulnar (**e**) deviation, hand in a fist (**f**) and extended palm (**g**).

2.2. Data Pre-Processing and Feature Extraction

The data pre-processing module starts splitting the EMG signals into 200 ms windows with an overlap of 100 ms (considering a sampling rate of 1 kHz). With the aim of finding out which mathematical transform obtained the best features for classification, four different transformations were defined: Short-Time Fourier Transform (STFT), with and without the cube root over the module of the transformation, and the Constant-Q-Transform (CQT) including or not the cube root over the module of the transformation.

The Short-Time Fourier Transform (STFT) was used on the overlapped windows. Then, the magnitude of the resulting transform is used to determine the energy at each frequency in each time window.

The Constant-Q-Transform (CQT) is an alternative transform. This transform is a variation of the Fourier transform but with higher resolution in low frequencies and a constant difference between harmonics. This characteristic (constant difference between harmonics) is very interesting when using Convolutional Neural Networks with linear kernels.

2.3. Classification Using Deep Learning

A deep neural network was implemented to differentiate between several types of forearm movements (Figure 2). The architecture includes four convolutional layers, each one followed by a max-pooling layer, to learn features, and a second subnet including three fully connected layers for classifying the performed gesture. To prevent overfitting, Dropout layers were added between the layers.

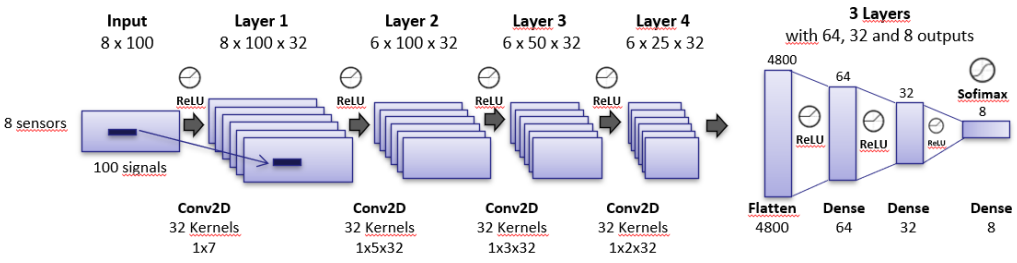

Figure 2. Neural network used, it should be noted that after each convolutional layer there is a dropout layer and after convolutional layer 2 and 3 there is a MaxPooling layer which we have not been able to illustrate.

In this architecture, the SoftMax activation function was used in the final layer to deal with the classification task, while the ReLU activation function was used in the intermediate layer. The optimizer employed was RMSprop. The following hyperparameters were

optimized using a validation subset (see the next section): learning rate, batch size, number of epochs, dropout rate, number of filters and size of kernels in feature extraction layers, and the number of neurons in the classification layers.

We managed to obtain these results after several experiments where we tested each neural network structure and each hyperparameter independently.

2.4. Validation Techniques

The experimental process consisted of two main phases. Firstly, we used the 15-Fold validation method with a subject-wise splitting strategy. This 15-Fold cross validation was chosen to ensure a robust evaluation of the classification system. Since we have 36 experimental subjects, we randomly select 30 subjects for performing the 15-Fold, leaving 6 of those subjects for the final test (16.7%). The 15-Fold cross validation considering 15 folds (2 subject in every fold) was focused on finetuning the main aspect of the system: type of transform, and hyperparameters of the deep learning network.

Secondly, in addition to the 15-Fold validation method, we did a final test evaluation with the 6 experimental subjects we separated at the beginning. Using an independent test subset is important to evaluate the generalization capability of the system. These 6 subjects weren't used during the training or validation process (system development). The purpose of using this test set is to get an independent measure of model performance on unseen data. This helps to get a realistic assessment of how the model will generalize with new data in practical real-world applications. This test set helps us to avoid overfitting and to learn not only the training set but also general patterns applicable to new subjects. This methodology ensures that the results obtained reflect the true performance of the model on unseen data.

2.5. Quality Measures

The quality measures we used during the experiments to ensure improvement were accuracy and f-score.

The accuracy was calculated as seen in Equation (1), where True Positives (TP) are the positive samples correctly detected, False Positives (FP) are negative samples classified as positive, True Negatives (TN) are the negative samples correctly detected and False Negatives (FN) is the positive samples incorrectly detected as the negative ones. Equation (1). Accuracy formula.

$$Accuracy = \frac{TP + TN}{TP + TN + \ FP + TN} \tag{1}$$

The F-score is the geometric mean of precision and recall, as seen in Equation (2). Where "Precision" is the TP divided by the detected positive samples ($Precision = \frac{TP}{TP+FP}$) and "Recall" is TP divided by the actual positive samples ($Recall = \frac{TP}{TP+FN}$). Equation (2). F1-score formula.

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{2}$$

3. Results

As commented previously, the experimentation was divided in two phases: system development using the 15-fold CV strategy and final testing of the best system configuration with new data. Our 15-fold experiments confirmed that the best configuration is the one shown in Table 2, with four Convolution Layers and two Max-Pooling layers with one Dropout layer after each Convolution layer.

Table 2. Best configuration of the Neural Network.

Conv (1,7)	Dropout	Conv (1,5)	MaxPooling	Dropout	Conv (1,3)	MaxPooling	Dropout	Conv (1,2)

The best values for the hyperparameters of this neural network configuration are the following: a batch size of 50, 10 epochs, a dropout rate of 0.3, 32 filters in the convolutional layers and a learning rate of 0.0005.

Making comparisons between the four different mathematical transformations for feature extraction, we concluded that the STFT transformation with cube root over the module of the transformation was the best features extraction strategy.

The next table shows the confusion matrix obtained with the best system in the 15-fold CV.

Analysing the confusion matrix on Table 3, we manage to take the decision of joining two of the seven movements. "Wrist flexion" with "Hand clenched in a fist" and "Radial deviations" with "Wrist extension". Additionally, we also notice that the "Guess 7" column is empty, this is because the movement "Extended Palm" was performed by very few users, which leads to the classification not being checked by much data, resulting in the values being assigned to the previous class, "Ulnar deviations".

Table 3. Confusion matrix of the last 15-Fold in percentages.

	Prediction						
	Label 1	**Label 2**	**Label 3**	**Label 4**	**Label 5**	**Label 6**	**Label 7**
Label 1: Hand at rest	**99.9**	**0**	**0.1**	**0**	**0**	**0**	**0**
Label 2: Hand clenched in a fist	**1**	**76**	**10.3**	**1.1**	**5.9**	**5.8**	**0**
Label 3: Wrist flexion	**5.9**	**1.9**	**84**	**0**	**1.3**	**7**	**0**
Label 4: Wrist extension	**0.1**	**1.6**	**0.1**	**59.1**	**29.4**	**9.9**	**0**
Label 5: Radial deviations	**5.1**	**4.3**	**4.1**	**4.6**	**80**	**2.0**	**0**
Label 6: Ulnar deviations	**1.3**	**5.2**	**25.3**	**3.3**	**2.2**	**62.8**	**0**
Label 7: Extended palm	**0**	**0**	**0**	**0**	**0**	**100**	**0**

After joining the movement "Hand clenched into a fist" with "Wrist flexion" and "Radial deviations" with "Wrist extension", we got a better confusion matrix as we can see in Table 4:

Table 4. Confusion matrix of the last 15-Fold CV in percentages after joining the movement "Hand clenched into a fist" into "Wrist flexion" and "Radial deviations" into "Wrist extension".

	Prediction				
	Label 1	**Label 2**	**Label 3**	**Label 4**	**Label 5**
Label 1: Hand at rest	**99.8**	**0.1**	**0.1**	**0**	**0**
Label 2: Wrist flexion	**1.4**	**87.5**	**3.4**	**7.7**	**0**
Label 3: Wrist extension	**1.5**	**5.5**	**86.3**	**6.6**	**0**
Label 4: Ulnar deviations	**0.4**	**27.6**	**5.6**	**66.2**	**0.1**
Label 5: Extended palm	**0**	**0**	**0**	**100**	**0**

With this new selection of movements, the classification accuracy of the neural network increased from 85.0% to 90.0 ± 0.12% in the 15-fold CV.

Using the final testing set, we also reproduced the same experiment as the authors did in the reference study [13], using just the 4 principal movements, which are wrist extension (Up), wrist flexion (Down), radial deviation (Right) and ulnar deviation (Left). Compared to this previous work, we improved the F-score from 85.5% (in the previous work), to 89.3% for this 4-class classification problem (left, right, up and down).

4. Discussion and Conclusions

This work has proposed and evaluated a gesture recognition system based on deep learning for classifying forearm gestures using EMG signals. The pre-processing module divides the EMG signals into overlapping time windows of 200 ms with a 100 ms shift. For each window, the Short-Time Fourier Transform (STFT) with a cube root of its module is applied to extract relevant information in the frequency domain. The classification module is based on a Convolutional Neural Network. The analysis of the confusion matrix

revealed that merging certain gesture classes significantly improves the classification performance. Specifically, combining "Hand clenched into a fist" with "Wrist flexion" and "Radial deviations" with "Wrist extension" resulted in a more accurate confusion matrix. The developing phase led to an increase in classification accuracy from 85.0 ± 0.13%to 90.0 ± 0.12% using a 15-fold CV over 5 classes. When comparing to previous works with the testing subset, we improved the F-score from 85.5%, to 89.3% for a 4-class problem (left, right, up and down). The proposed preprocessing and classification modules developed in this study have the potential to significantly enhance the functionality of human-machine interfaces and prosthetic devices by providing more precise and reliable gesture recognition.

Author Contributions: Conceptualization, D.G.-V. and R.S.-S.; methodology, D.G.-V. and R.S.-S.; software, D.G.-V., S.E.-R., M.G.-M.; validation, D.G.-V. and R.S.-S.; formal analysis, D.G.-V., M.G.-M. and R.S.-S.; investigation, D.G.-V. and R.S.-S.; resources, R.S.-S.; data curation, D.G.-V. and S.E.-R.; writing—original draft preparation, D.G.-V.; writing—review and editing, S.E.-R., M.G.-M. and R.S.-S.; visualization, D.G.-V.; supervision, R.S.-S.; project administration, R.S.-S.; funding acquisition, R.S.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been funded by the European Union' EDF programme under grant agreement number 101103386 (FaRADAI project).

Institutional Review Board Statement:

Informed Consent Statement:

Data Availability Statement:

Acknowledgments: Authors thank all the other members of the FaRADAI project for the continuous and fruitful discussion on these topics.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lv, B.; Sheng, X.; Zhu, X. Improving myoelectric pattern recognition robustness to electrode shift by autoencoder. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Honolulu, HI, USA, 18–21 July 2018; Institute of Electrical and Electronics Engineers Inc.: Los Alamitos, CA, USA, 2018; pp. 5652–5655.
2. Buongiorno, D.; Barsotti, M.; Barone, F.; Bevilacqua, V.; Frisoli, A. A linear approach to optimize an EMG-driven neuromusculoskeletal model for movement intention detection in myo-control: a case study on shoulder and elbow joints. *Front. Neurorobot.* **2018**, *12*, 74.
3. Buongiorno, D.; Barsotti, M.; Sotgiu, E.; Loconsole, C.; Solazzi, M.; Bevilacqua, V. A neuromusculoskeletal model of the human upper limb for a myoelectric exoskeleton control using a reduced number of muscles. In Proceedings of the IEEE World Haptics Conference, WHC 2015, Evanston, IL, USA, 22–26 June 2015.
4. Peppoloni, L.; Filippeschi, A.; Ruffaldi, E.; Avizzano, C.A. (WMSDs issue) a novel wearable system for the online assessment of risk for biomechanical load in repetitive efforts. *Int. J. Ind. Ergon.* **2016**, *52*, 1–11.
5. Bishop, M.D.; Pathare, N. Considerations for the use of surface electromyography. *Phys. Ther. Korea* **2004**, *11*, 61–69.
6. Besier, T.F.; Lloyd, D.G.; Ackland, T.R.; Cochrane, J.L. Anticipatory effects on knee joint loading during running and cutting maneuvers. *Med. Sci. Sports Exerc.* **2001**, *33*, 1176–1181.
7. Buongiorno, D.; Sotgiu, E.; Leonardis, D.; Marcheschi, S.; Solazzi, M.; Frisoli, A. WRES: a novel 3 DoF WRist ExoSkeleton with tendon-driven differential transmission for neurorehabilitation and teleoperation. *IEEE Robot. Autom. Lett.* **2018**, *3*, 2152–2159.
8. Stroppa, F.; Stroppa, M.S.; Marcheschi, S.; Loconsole, C.; Sotgiu, E.; Solazzi, M.; Buongiorno, D.; Frisoli, A. Real-time 3D tracker in robot-based neurorehabilitation. In *Computer Vision for Assistive Healthcare*; Elsevier: Amsterdam, The Netherlands, 2018.
9. Kiguchi, K.; Hayashi, Y. An EMG-based control for an upper-limb power-assist exoskeleton robot. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2012**, *42*, 1064–1071.
10. Fougner, A.; Stavdahl, O.; Kyberd, P.J.; Losier, Y.G.; Parker, P.A. Control of upper limb prostheses: Terminology and proportional myoelectric control—A review. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2012**, *20*, 663–667.
11. Wakeling, J.M. Spectral properties of the surface EMG can characterize motor unit recruitment strategies. *J. Appl. Physiol.* **2008**, *105*, 1676–1677.
12. Saturn Cloud. A Comprehensive Guide to Convolutional Neural Networks (the ELI5 way). Available online: https://saturncloud.io/blog/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way/ (accessed on 17 May 2024).

13. Lobov, S.; Krilova, N.; Kastalskiy, I.; Kazantsev, V.; Makarov, V.A. Latent Factors Limiting the Performance of sEMG-Interfaces. *Sensors* **2018**, *18*, 1122. https://doi.org/10.3390/s18041122.
14. Krilova, N.; Kastalskiy, I.; Kazantsev, V.; Makarov, V.A.; Lobov, S. *EMG Data for Gestures*; UCI Machine Learning Repository: Irvine, CA, USA, 2019. https://doi.org/10.24432/C5ZP5C.
15. Lobov, S.A.; Mironov, V.I.; Kastalskiy, I.A.; Kazantsev, V.B. A spiking neural network in sEMG feature extraction. Sensors 2015, 15, 27894–27904.
16. Amazon.in. Myo Gesture Control Armband—White. Available online: https://www.amazon.in/Myo-Gesture-Control-Armband-White/dp/B00O69U344 (accessed on 20 May 2024).