Journal of Theoretical and Applied Information Technology
30th September 2024. Vol.102. No. 18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HAND GESTURE CLASSIFICATION USING TIME-FREQUENCY IMAGES AND VALIDATION APPROACHES

CHAPARRO MACIAS CAMILO ANDRES[1], CARDOZO SANCHEZ MAICOL ESTIVEN[1], JUAN PABLO GASCA CALDERÓN [1], DIEGO FERNANDO GONZÁLEZ [1] AND RUTHBER RODRÍGUEZ SERREZUELA[1]

[1] Mechatronics Engineering, Corporación Universitaria del Huila

ID 55179 Submission	Editorial Screening	Conditional Acceptance	Final Revision Acceptance
23-07-24	24-07-2024	31-07-2024	24-09-2024

ABSTRACT

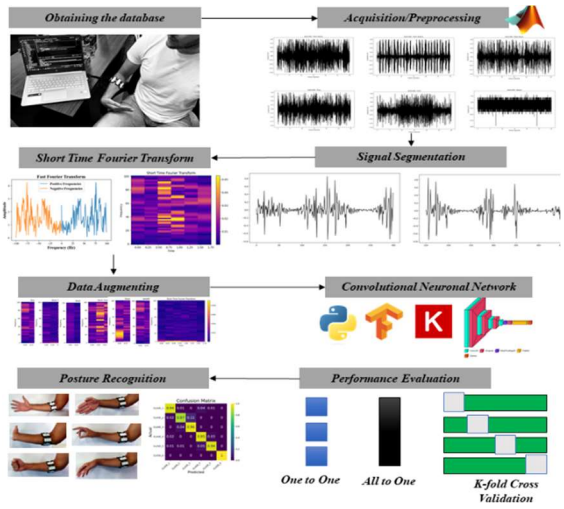
Surface electromyographic (sEMG) signals are a non-invasive method for acquiring signals that play a fundamental role in the monitoring of prosthetic devices by providing information about human motor functions. This leads to the need for accurate classification of sEMG signals, despite variations in signal stationarity, the presence of sensor noise, differences between the muscles involved, and the peculiarities of each patient. This study focuses on the classification of hand grip postures using sEMG signals acquired from amputee patients. Special emphasis is placed on the use of the time-frequency domain for feature extraction, using the spectral analysis of the reduced-time Fourier transform (STFT). To carry out this task, a classification model based on a convolutional neural network (CNN) is used. The classification method is adjusted, trained, and evaluated through three experiments. The first, called "One to One", yields accuracy percentages of 90.84%, 91.05%, and 91.13% for spectrograms of 32x32, 64x64, and 128x128 in size, respectively. In the second validation, called "All by One", an accuracy of 62.28% is achieved for spectrograms of 32x32 pixels. Finally, in the last K-fold cross-validation validation, an average accuracy of 86.73%, 86.77%, and 87.97% is obtained for spectrograms of 32x32, 64x64, and 128x128 in size, respectively.

Keywords: *Electromyography, Hand Gesture, Classification, STFT, CNN.*

1. INTRODUCTION

Advancements in robotics and artificial intelligence have propelled the creation of innovative technologies in fields such as robot-assisted rehabilitation, biofeedback, ergonomics, and neurophysiology. These technologies focus on understanding muscular coordination, defined as the process in which muscles activate and contract to generate force in the joints, whether consciously or automatically. Technological progress has stimulated the development of signal processing techniques, particularly those related to non-periodic and transient electromyographic (EMG) signals. These techniques find applications in the diagnosis of neuromuscular diseases, human-machine interaction, gaming experiences, sign language detection, virtual reality, exoskeleton design, and monitoring devices and systems for individuals with amputations, such as myoelectric prosthetics.

These innovations promise to enhance autonomy and the quality of life for individuals with upper limb disabilities, opening new perspectives in rehabilitation and healthcare. The acquisition and processing of EMG signals are crucial in the classification of hand grip postures, identifying relevant features in muscle contraction patterns. The EMG signal preprocessing to eliminate harmonics, feature extraction using time windows, and transforming the EMG signal into the frequency domain through the conventional Fourier transform



(FT), extracting features such as magnitude, phase, and power spectral density (PSD).

Figure 1: The schematic visualization of the proposed method.

In this context, the Short-Time Fourier Transform (STFT) is employed to represent temporal changes in the signal frequency. The STFT is generated by shifting a window along the signal and calculating the FT at time intervals defined by the width of the window, providing a time-frequency representation of the EMG signal. Recent research has revealed that the STFT offers frequency-related features, such as spectral moment, frequency centroid, PSD integration, frequency centroid variance, average spectral magnitude, autoregressive coefficients, and cepstral coefficients, which are useful in a multivariable regression model such as a Convolutional Neural Network (CNN).

The architecture of a CNN mimics the visual cortex, utilizing hierarchical and modular processing layers to learn more complex features from input signals. In this research, a comprehensive analysis of EMG signals was employed to develop an effective methodology for classifying six hand grip postures. Processed EMG signals in the time-frequency domain using STFT were represented as spectrograms, obtained from a database of ten amputee patients in Santiago de Cuba. Classification was performed using advanced machine learning techniques, optimized with Keras Tuner in Python.

2. MATERIALS AND METHODS.

2.1 Obtaining the database.

For data collection, a group consisting of ten amputees was formed, encompassing both men and women with ages ranging from 25 to 65 years. All participants provided their informed written consent voluntarily before engaging in the experimental procedures. Additionally, a thorough check was conducted to ensure the absence of neurological or psychiatric conditions in the case of older participants.

It is noteworthy that this study exclusively focused on patients with transradial amputations, meaning limb loss below the elbow and the upper portion of the wrist. Details about each participant involved in the research, such as their age, height, weight, gender, level of amputation, time elapsed since amputation, and Dash questionnaire results (assessing individual health before the procedure), are succinctly summarized in **Table 1**.

Before the experimental procedures, participants received detailed training on the process that would be undertaken. Subsequently, sensors were calibrated on both the amputated and non-amputated limbs. This calibration aimed to measure the electrical activity of muscles and quantify the force and effort required to perform specific tasks in both arms. The information obtained from these measurements proved crucial in determining the resources needed for the classification of postures in patients with amputations.

Table 1: Features of Amputated Subjects.

Patients	Age Years	Height (cm)	Weight (kg)	Sex	Amputation level	Amputation time	DASH Index
P01	36	1,68	70	M	10 cm from the elbow	1 year	45
P02	51	1,82	80	M	Wrist	30 years	19
P03	62	1,78	90	M	Wrist	36 years	39,16
P04	26	1,79	68	M	10 cm from the elbow	12 years	20
P05	60	1,73	77	M	Wrist	41 years	26,6
P06	55	1,75	55	M	Wrist	5 years	16,66
P07	28	1,76	70	M	10 cm from the elbow	9 years	24,16
P08	48	1,75	72	F	Wrist	22 years	20,83
P09	65	1,65	78	M	Wrist	29 years	42,5
P10	35	1,66	69	M	10 cm from the elbow	2 years	47,5

2.2 Acquisition of EMG Signals.

Data capture was performed using the MYO ARMBAND EMG sensor **(Figure 2)**, developed by the Canadian company Thalmic Labs, and is available to consumers at an affordable cost. This device specializes in reading and recording the electrical activity of muscles. The Myo Armband is equipped with 8 EMG sensors, like surface electrodes, and can non-invasively acquire sEMG data at a sampling frequency of 200 Hz. What sets it apart from conventional electrodes is its ability to provide user comfort without causing muscle discomfort, and its placement is easily replicable.

Journal of Theoretical and Applied Information Technology
30th September 2024. Vol.102. No. 18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Additionally, the sensor has the capability to establish wireless communication with other electronic devices via Bluetooth. It is also equipped with a nine-axis Inertial Measurement Unit (IMU), enabling the detection of the position and movement of the user's arm. To obtain the myoelectric signals, the armband was positioned at 3-5 cm distal to the patient's elbow and was pre-calibrated for both limbs.

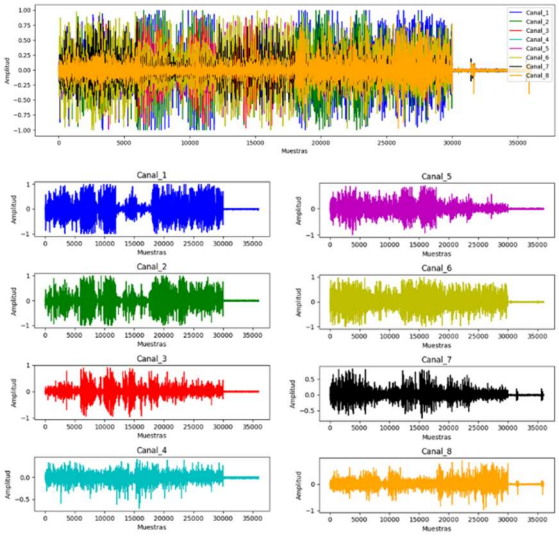


Figure 2: Acquisition of the EMG signal using the 8-channel MYO ARMBAND EMG sensor.

During the EMG signal acquisition process, patients were instructed to perform six different hand grip postures: open hand (class 1), wave in (class 2), wave out (class 3), pinch (class 4), fist (class 5), and rest position (class 6). The experiment was conducted in two sessions, and subjects were asked to sit in a comfortable chair with their elbows flexed at a 90-degree angle. They were positioned two meters away from a screen displaying visual cues indicating the posture they should adopt. Data were recorded for both the contralateral and amputated limbs. For each type of hand posture, data were acquired at a sampling frequency of 200 Hz for 30-second intervals. A one-minute break was included between each posture transition, following recommendations to reduce or avoid mental and physical fatigue [12, 22]. Consequently, a total of 6000 samples were collected for each posture per patient **(Figure 3)**. These data were stored in a matrix and subsequently processed on a computer for analysis.

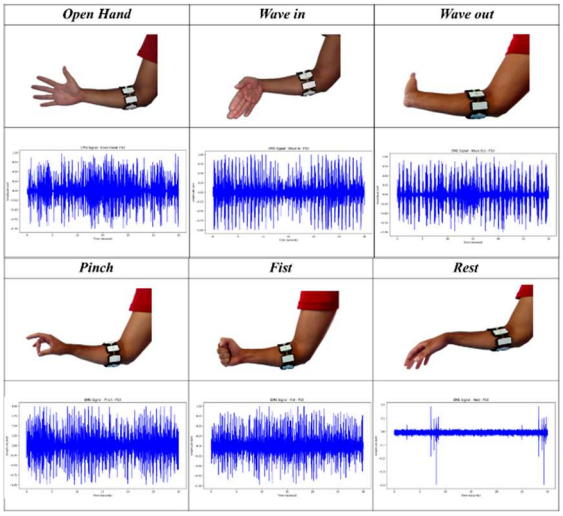


Figure 3: Six different hand gestures used in the study along with their corresponding EMG signals.

2.3 Signal Preprocessing.

During signal acquisition, it was observed that various harmonics overlapped with the signal, affecting its fidelity and subsequent feature extraction, and consequently, the diagnosis of the EMG signals. Chowdhury and colleagues mention that the most common noises present in EMG signals include electromagnetic interference from electronic devices, electrode movements during data capture, leading to unwanted artifacts in the signal, and finally, the intrinsic resistance of muscle motor units to movement, generating inherent instability in the signal. To counteract these unwanted effects, the original signals obtained underwent a filtering process using digital bandpass filters, with cutoff frequencies ranging from 20 to 500 Hz. This filtering process was performed using MATLAB software, allowing the selective removal of unwanted components and the preservation of relevant information in the EMG signals.

2.4 EMG Signal Segmentation.

After applying the signal filtering process **(Figure 4a)**, it was deemed essential to employ segmentation or windowing techniques to properly extract the required features. In this context, the decision was made to use the overlapped segmentation technique, which involves shifting a new segment over the current segment with an overlap smaller than the segment size. As a result, the signal was divided into segments of 300 ms **(Figure 4b)**, of which 150 ms **(Figure 4c)** overlapped, following the guidelines recommended by [2, 13]. This strategy allowed for the generation

of a more extensive dataset for the deep training of the method while providing additional information about the EMG signal at the exact moment of muscle contraction. The application of this segmentation technique led to the collection of a total of 312 samples for each posture from each of the 10 patients involved in the study.

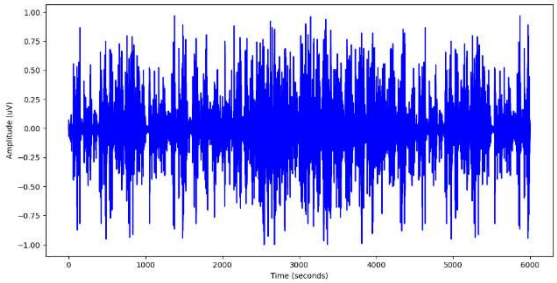


Figure 4a: Pure EMG signal in the time domain. Open hand of Patient 10.

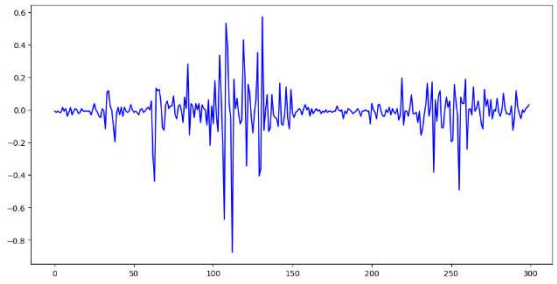


Figure 4b: Fig 4b. Segmentation N°1 - 300 ms EMG signal in the time domain. Channel 1 - Open Hand - Patient 10.

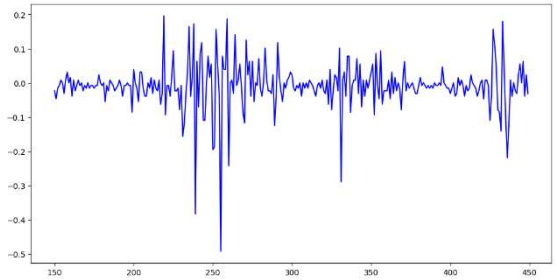


Figure 4c: Segmentation N°2 - 300 ms EMG signal in the time domain, with 150 ms overlap. Channel 1 - Open Hand – Patient 10.

2.5 Short Time Fourier Transform

The time-frequency representation (TFR) maps a one-dimensional time signal onto a two-dimensional frequency and time signal. TFR is widely used to analyze, synthesize, and accurately modify results of non-stationary signals, as it considers both frequency and time. The Short-Time Fourier Transform (STFT) is the basic form of TFR, generating narrow segments of long-distance signals (1). Sufficiently narrow segments appear as stationary, and the Fourier Transform (FT) is taken for each segment **(Figure 5)**. Each FT displays the spectral details of the signal at a different time cut, providing a simultaneous estimation of frequency and time [20,21]. The STFT of the EMG signal provides significant information about muscular activity during a task [15,16]. The generalized formula for STFT is:

$$STFT_x(t,\omega) = \int_{-\infty}^{\infty} x(\tau - t)e^{-2\pi f}\, d\tau \qquad (1)$$

Where $x(\tau)$ is the signal, $\omega(\tau - t)$ is the observation window, and the variable t slides the window over the signal, $x(\tau)$.

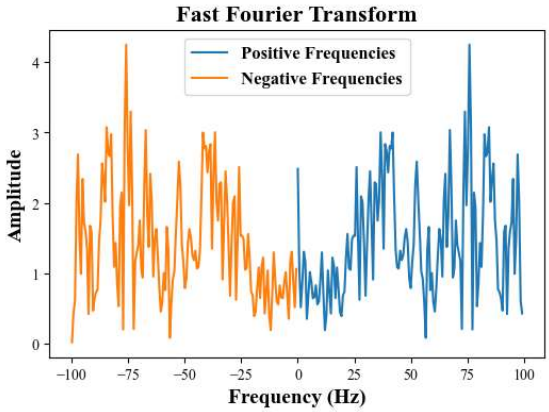


Figure 5: FT Patient 10 - Open Hand Segment 01. Frequency details of the signal at a different time cut.

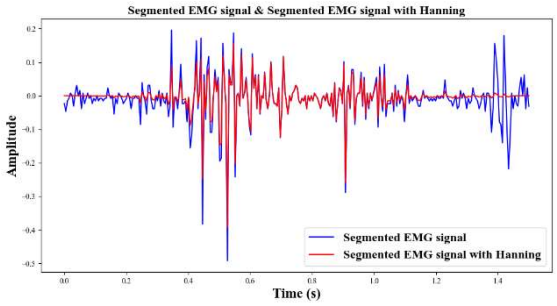
A spectrogram is essentially the magnitude squared of the Short-Time Fourier Transform (STFT), providing a representation of the power and energy distribution of a signal in terms of frequency at a specific moment (2). In both the case of the spectrogram and the STFT, a balance is struck between frequency-based and time-based perspectives of a signal. The accuracy of the representation in the time and frequency domain can be adjusted by the window size [8,9]. The mathematical expression for the spectrogram is defined as:

$$STFT_x(t,\omega) = \left| \int_{-\infty}^{\infty} x(\tau)\omega(\tau - t)e^{-2\pi f\tau} \right|^2 d\tau \quad (2)$$

However, it is important to note that the analysis window function plays a crucial role in the Short-Time Fourier Transform (STFT). If this window has a longer duration in time, it is equivalent to a narrow-bandpass filter in frequency, implying a more detailed sampling in the frequency domain. In this situation, the STFT could preserve subtle variations in frequency, although it tends to smooth out rapid changes in the time domain through averaging. On

the other hand, if the window is shorter, rapid variations in time are retained, but the detection of rapid changes in frequency is limited.

Figure 6: Segmented EMG Signal vs Segmented EMG Signal with Hanning.



This balance between time and frequency is known as the uncertainty principle. The spectrogram allows visualizing the location in both time and frequency simultaneously, which is particularly useful in detecting sudden changes in rhythmic patterns present in biomedical signals (Figure 7). For the research, the decision was made to use a Hanning window with a size of 128, as this choice optimally suited the model in question **(Figure 6)**.

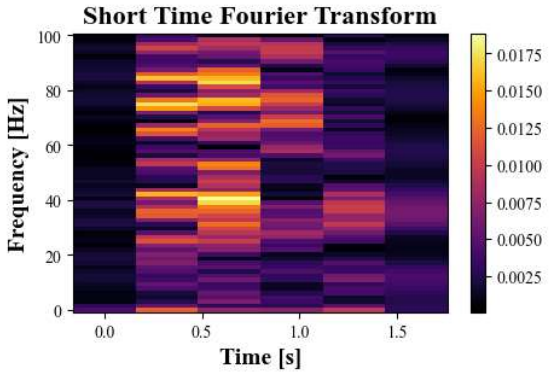


Figure 7: STFT Spectrogram - Patient 10 - Segment 01 Open Hand.

2.6 Dataset of Images

The original dataset comprised 312 spectrograms for each posture of each patient, with an image size of 434x343 pixels. This image size could incur high computational costs when training and validating the proposed Convolutional Neural Network (CNN) model. Given that the ultimate application of this methodology targets low-cost robotic prosthetics, using images with the original size would not be practical. Therefore, before proceeding with the training and validation of the model, the images were resized into three different groups (32x32, 64x64, 128x128). This allowed the creation of three different datasets with the aim of significantly reducing the number of network parameters, thereby achieving low latency and lower energy consumption.

2.7 Dataset of Augmentation

It is crucial that the dataset used to train, test, and validate a Convolutional Neural Network (CNN) is sufficiently diverse to avoid overfitting issues. Overfitting can lead to low generalization capacity and, consequently, low accuracy in predictions due to a lack of variability in the training data. Additionally, an insufficiently diverse dataset could make the model highly sensitive to the training data. To mitigate these issues, data augmentation techniques were implemented to enhance the generalization capacity of the CNN and reduce the risk of overfitting. As a result of these techniques, 1560 additional images were generated for each hand grip posture, in addition to the original 312 images. This means that 11,232 images were obtained per patient, constituting a reasonably sized dataset for evaluating the proposed model.

2.8 Convolutional Neural Network

The Convolutional Neural Network (CNN) stands out as one of the most effective deep learning methods for high-dimensional image classification. By using spectrograms as inputs, the manual feature extraction process is simplified, as this type of algorithm offers automatic feature extraction using the concept of a deep neural network [13]. Essentially, a CNN consists of three main layers. The convolutional layer is responsible for applying a set of filters to the input image. During training, the image convolves with each filter, generating a feature map that highlights regions of the image important for the classification task. Convolution occurs by matrix multiplication between the filter and the region of the input image, moving across it. This way, the convolutional layer can automatically learn representative features of the image [12], [13]. It then proceeds to the max-pooling layer, where the maximum values in a specific region of the input image are selected, creating a new feature map with reduced resolution. This operation is known as non-linear subsampling, where the input dimension is reduced, but the most important features of the image are retained [11].

Finally, there is the fully connected layer, whose main function is to classify the input images through a vector multiplication of the features extracted from the convolutional and pooling layers, stored in a weight matrix that the network learned during

training. Next, a non-linear activation function is applied, determining the final output of the network [7]. In this way, the convolutional and max-pooling layers are responsible for extracting features from the images so that the fully connected layer can handle the classification.

2.9 Structure of the proposed CNN

For the model creation, a hyperparameter tuning process was carried out, which is done before training the network as it plays a crucial role in determining the architecture and behavior of the model [14]. To perform this tuning, the Keras Tuner library in Python was implemented. This library trains and evaluates various models containing different combinations of hyperparameters, ultimately providing the model with the best performance. In this case, variations were made in kernel sizes, the number of convolutions, dropout layers, dense layers, and the learning rate.

After completing the tuning process, the best CNN model obtained consists of 4 convolutional layers with 64,56, 40, and 8 features respectively. These layers use a 3x3 kernel, and a dropout layer with a value of 0.2 is included to prevent overfitting. The activation function applied is ReLU (Rectified Linear Unit). Additionally, a 2x2 max-pooling layer is incorporated. Finally, there is the "flatten" layer responsible for transforming the previous layer into a one-dimensional dimension, allowing classification through the fully connected layer. The activation of the last layer is done using the SoftMax function **(Figure 8)**.

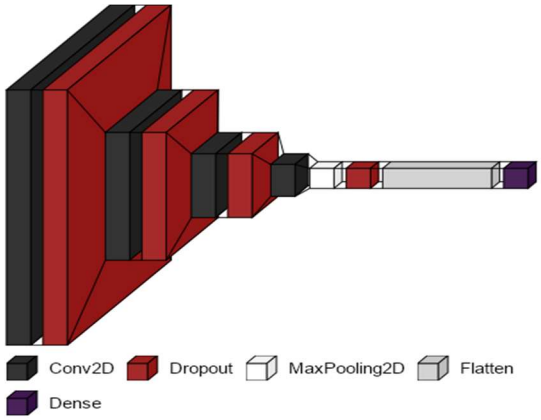


Figure 8: Structure of the proposed CNN.

3. RESULTS

3.1 "One to One" Approaches.

The initial validation of the model was carried out following the "One to one" methodology. This technique allows the network to learn from the spectrograms of each patient individually. In other words, a unique model is created for each patient, enabling the model to identify specific characteristics of each patient. This results in high accuracy in classifying hand grip postures. Following this methodology, the datasets for each patient contained a total of 11,232 images, of which 10% was used to evaluate the final model performance, another 10% to validate the model at each training epoch, and the remaining 80% was used for training. Table 2 shows the metrics accuracy percentage obtained by each patient in the classification of different postures, varying according to the size of the input images.

Table 2: Accuracy Score "One to One" Approaches.

	32x32				64x64				128x128			
Patient	Accuracy Score	F1 Score	Recall Score	Precision Score	Accuracy Score	F1 Score	Recall Score	Precision Score	Accuracy Score	F1 Score	Recall Score	Precision Score
P01	94,13	94,13	94,29	94,21	93,33	93,33	93,45	93,44	94,48	94,48	94,59	94,52
P02	84,34	84,34	84,34	84,57	85,14	85,14	85,18	85,23	87,46	87,46	87,42	87,34
P03	87,1	87,1	86,93	86,91	87,46	87,46	87,32	87,34	87,81	87,81	87,89	88,01
P04	95,02	95,02	95,04	95,06	92,88	92,88	92,86	92,99	92,44	92,44	92,3	92,6
P05	90,57	90,57	90,45	90,38	90,57	90,57	90,39	90,34	92,44	92,44	92,54	92,49
P06	92,79	92,79	92,76	92,74	91,55	91,55	91,71	91,6	91,1	91,1	91,11	91,35
P07	94,13	94,13	94,25	94,23	93,51	93,51	93,56	93,67	93,42	93,42	93,49	93,49
P08	90,57	90,57	90,64	90,53	91,37	91,37	91,43	91,31	90,48	90,48	90,46	90,55
P09	89,23	89,23	89,15	89,08	91,99	91,99	91,84	92	90,04	90,04	89,89	90,26
P10	90,57	90,57	90,65	90,7	92,7	92,7	92,73	92,79	91,64	91,64	91,55	91,56
Average	90,845	90,845	90,85	90,841	91,05	91,05	91,047	91,071	91,131	91,131	91,124	91,217

It is noteworthy that the models generated from the dataset of patient 01 **(Figure 9)** achieved a higher accuracy percentage. This is evident in their respective confusion matrices. These matrices display the class to which the images belong and the percentage of images that the model classified as belonging to that class. It is important to recall that class 1 corresponds to posture 1, which is "open hand," and so forth.

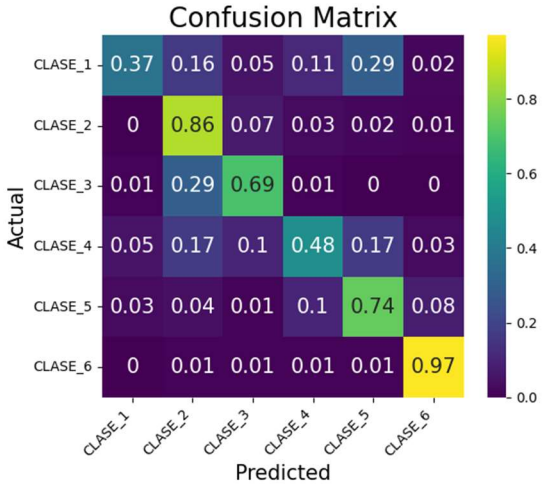


Figure 9: Confusion Matrix Patient 01. Spectrogram 32x32 pixels

3.2 "All to One" Approaches

The second training of the neural network was validated using the "All to one" methodology. To apply this methodology, it was necessary to train the model with the dataset that included all patients. Subsequently, the model was evaluated with each patient to obtain the average accuracy percentage. However, since the network is trained with data from all patients, it learns only the general characteristics and not the specific ones of each patient. This results in a decrease in its ability to correctly classify postures in each patient. To carry out this training, a dataset containing a total of 112,230 spectrograms from all patients was implemented, divided into their respective classes (postures). Twenty percent of the data was used to evaluate the accuracy of the final model, another 20% to validate the model at each training epoch, and the remaining 60% was used for the training process.

According to the data in Table 3, it can be observed that the best patient in this validation was patient 04 (Fig 10). Unlike the "One to One" validation, the proposed model faces difficulties in classifying different classes in the "All to One" validation. However, when evaluating the model trained with each patient's dataset, the confusion matrix shows a high degree of accuracy in Classes 2 (Open Hand), 5 (Fist), and 6 (Rest).

Table 3: Metrics "All to One" Approaches

Metrics - 32x32				
Patient	*Accuracy Score*	*F1 Score*	*Recall Score*	*Precision Score*
P01	*66,36*	*66,36*	*66,81*	*67,44*
P02	*47,49*	*47,49*	*47,73*	*48,68*
P03	*57,28*	*57,28*	*57,16*	*56,19*
P04	*68,71*	*68,71*	*68,46*	*70,5*
P05	*65,55*	*65,55*	*65,96*	*65,92*
P06	*59,81*	*59,81*	*59,38*	*61,86*
P07	*66,84*	*66,84*	*66,82*	*69,21*
P08	*60,97*	*60,97*	*61,32*	*63,08*
P09	*62,57*	*62,57*	*62,54*	*61,7*
P10	*67,25*	*67,25*	*67,56*	*68,61*
Average	*62,283*	*62,283*	*62,374*	*63,319*

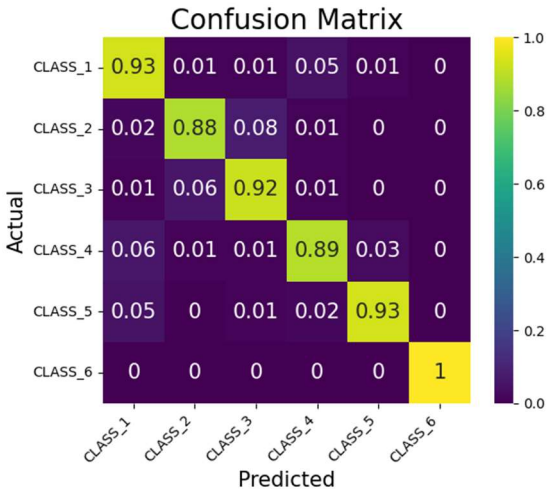


Figure 10: Confusion Matrix Patient 04. Spectrogram 32x32 pixels.

3.3 "K-fold Cross Validation" Approaches

Cross-validation is a fundamental technique in the field of machine learning, playing a crucial role in model evaluation and hyperparameter selection. This methodology is based on the premise that a model should be able to generalize well to unseen data. To achieve this, cross-validation divides the dataset into multiple folds, and in each iteration, it trains the model on a subset of the data and evaluates it on another subset (for this research, it was performed with 5 folds). This strategy provides a more accurate estimate of the model's performance and helps prevent overfitting, allowing data scientists and machine learning engineers to assess the performance of their models more reliably.

Following this methodology, the datasets for each patient contained a total of 11,232 images. Of these, 10% was used to evaluate the final model performance, another 10% to validate the model at each training epoch, and the remaining 80% was used for the training process. It is relevant to mention that in each iteration, the model was trained with 5 subsets of data and evaluated using the same number of subsets. Table 4 presents the average accuracy percentage obtained by each patient in the classification of different postures, varying according to the size of the input images.

Table 4: Metrics "K-fold cross validation" Approaches

	32x32			64x64			128x128		
Patient	Average Accuracy Train	Average Accuracy Test	Average Accuracy Validation	Average Accuracy Train	Average Accuracy Test	Average Accuracy Validation	Average Accuracy Train	Average Accuracy Test	Average Accuracy Validation
P01	98,5	91,01	91,01	99,52	91,99	92,23	99,65	92,7	91,97
P02	94,07	78,82	78,15	98,64	82,11	83,03	99,06	82,56	82,05
P03	96,03	80,42	80,16	98,3	80,42	82,81	98,37	81,13	83,01
P04	98,06	90,21	89,83	98,78	90,3	90,07	99,5	89,67	90,66
P05	98,46	88,61	88,64	99,45	90,03	89,18	99,58	89,59	89,54
P06	98,41	89,85	88,72	99,44	90,65	87,99	99,47	87,72	88,4
P07	98,99	91,19	91,19	86,89	79,35	80,77	99,7	90,39	91,41
P08	98,21	88,07	86,44	99,33	87,54	86,17	99,65	86,56	86,62
P09	98,29	83,62	85,77	99,26	86,2	87,29	99,39	86,74	86,84
P10	99,13	89,05	87,46	99,21	90,03	88,18	99,54	87,54	89,29
Average	97,815	87,085	86,737	97,882	86,862	86,772	99,391	87,46	87,979

4. ANALYSIS

In the "One to One" validation, where individual models are created for each patient, the superior performance of patient 01 suggests that tailoring the model to the specific characteristics of each patient enhances its ability to classify handgrip postures accurately. This personalized approach allows the model to capture patient-specific nuances and variations in muscle activity patterns, resulting in a higher level of precision. The training process, focused exclusively on one patient's data, facilitates the model's adaptation to the unique features of that individual's muscular responses.

Conversely, in the "All to One" validation, where a single model is trained on a dataset combining samples from all patients, the model lacks the capacity to discern patient-specific intricacies. Training on a heterogeneous dataset, which amalgamates various patient characteristics, may lead to a generalized model that struggles to perform optimally for individual patients. The compromise in model specificity may be evident in the reduced accuracy observed, especially for patients with distinct muscle activation patterns.

The "K-fold Cross Validation" serves as a valuable augmentation to these methodologies by providing a more robust evaluation. By dividing the dataset into multiple folds and iteratively training and validating the model on different subsets, this approach ensures a comprehensive assessment across various scenarios. This technique guards against overfitting by assessing the model's performance on unseen data, enhancing its generalization capability **(Figure 11)**.

Figure 11: Confusion Matrix Patient 01. Spectrograms 128x128 pixels

The observed performance variations underscore the importance of a nuanced model evaluation strategy. While "One to One" may excel in capturing individual patient characteristics, it might lack the broader generalization seen in "All to One." The "K-fold Cross Validation" strikes a balance, offering a more complete understanding of the model's adaptability and performance across diverse scenarios. Further exploration and refinement of these methodologies could lead to enhanced model performance tailored to individual patient needs.

5. CONCLUSION

The study introduces the development of a methodology capable of classifying EMG signals captured by the MYO armband sensor. These signals were transformed into spectrograms using the Short-Time Fourier Transform (STFT). A convolutional neural network comprising 58,974 parameters was trained with the aim of future implementation on a microcontroller for a robotic prosthesis. After validating the methodology through the "One to One" method, an average accuracy percentage of 90.84%, 91.05%, and 91.13% was achieved for spectrograms of 32x32, 64x64, and 128x128, respectively.

In the second validation using the "All by One" method, a result of 62.28% accuracy was obtained for 32x32 spectrograms. This demonstrates that training a unique model for each patient allows the network to identify specific characteristics of each posture, leading to better classification. However, when generalizing the model to all patients, the network fails to capture these specific features, resulting in suboptimal classification.

Nevertheless, the K-Fold Cross-Validation was employed, proving useful in determining how the model would generalize to new and unseen data. This is crucial for preventing overfitting and measuring the model's generalization capacity in different situations, yielding an average accuracy percentage of 86.73%, 86.77%, and 87.97% for spectrograms of 32x32, 64x64, and 128x128, respectively.

Further, another validation method called Transfer Learning is contemplated. This involves utilizing a pre-trained neural network on a large and general dataset (such as ImageNet or ResNet) and fine-tuning it for a specific image classification task on a smaller and more specific dataset. This approach aims to demonstrate the feasibility of

implementing these methodologies in a myoelectric robotic prosthesis.

REFERENCES

[1] Oh, D.-C., & Jo, Y.-U. (2021). Classification of hand gestures based on multi-channel EMG by scale average wavelet transform and convolutional neural network. International Journal of Control, Automation, and Systems, 19(3), 1443–1450.

[2] Ozdemir, M. A., Kisa, D. H., Guren, O., Onan, A., & Akan, A. (2020). EMG based Hand Gesture Recognition using Deep Learning. 2020 Medical Technologies Congress (TIPTEKNO).

[3] Rabin, N., Kahlon, M., Malayev, S., & Ratnovsky, A. (2020). Classification of human hand movements based on EMG signals using nonlinear dimensionality reduction and data fusion techniques. Expert Systems with Applications, 149(113281), 113281.

[4] Roland, T., Baumgartner, W., Amsuess, S., & Russold, M. F. (2017, July). Capacitively coupled EMG detection via ultra-low-power microcontroller STFT. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 410-413). IEEE.

[5] Tsai, A. C., Hsieh, T. H., Luh, J. J., & Lin, T. T. (2014). A comparison of upper-limb motion pattern recognition using EMG signals during dynamic and isometric muscle contractions. Biomedical Signal Processing and Control, 11, 17-26.

[6] Satti, A. T., Kim, J., Yi, E., Cho, H. Y., & Cho, S. (2021). Microneedle array electrode-based wearable EMG system for detection of driver drowsiness through steering wheel grip. Sensors, 21(15), 5091.

[7] Too, J., Abdullah, A. R., Mohd Saad, N., Mohd Ali, N., & Tee, W. (2018). A new competitive binary grey wolf optimizer to solve the feature selection problem in EMG signals classification. Computers, 7(4), 58.

[8] Serrezuela, R. R., Cardozo, M. Á. T., Montiel, J. J. G., Zamora, R. S., & Reyes, E. M. (2019, August). Análisis comparativo entre de MAE y RNA en señales de EMG obtenidas para control de una prótesis mano robótica. In Memorias de Congresos UTP (pp. 107-112)

[9] Grandett, Y. D. C., Argote, N. S., & Serrezuela, R. R. (2020). ANALYSIS BETWEEN ELM AND ANN IN EMG SIGNALS OBTAINED FOR THE CONTROL OF A ROBOTIC HAND PROSTHESIS. ARPN Journal of Engineering and Applied Sciences. pp 2806-2811.

[10] Serrezuela, R. R., Zamora, R. S., & Reyes, E. M. (2020). Control Strategy for Underactuated Multi-Fingered Robot Hand Movement Using Electromyography Signal with Wearable Myo Armband. In Biosensors-Current and Novel Strategies for Biosensing. IntechOpen.

[11] Cardozo, M. Á. T., Serrezuela, R. R., Trujillo, J. L. A., Zamora, R. S., & Reyes, E. M. (2021). DESIGN AND IMPLEMENTATION OF A ROBOTIC HAND PROSTHESIS UNDER THE TENSEGRITY APPROACH FOR TRANSRADIAL AMPUTEES. ARPN Journal of Engineering and Applied Sciences. 16(4) pp 503-508

[12] Broche-Vázquez, L., Torres-Quezada, M., Milanés-Hermosilla, D., González-Romero, D., Rodríguez-Serrezuela, R., & Sagaró-Zamora, R. (2020). Exoesqueleto robótico para la rehabilitación del miembro superior del paciente hemipléjico. Ingeniería Mecánica, 23(3).

[13] Serrezuela, R. R., Quezada, M. T., Zayas, M. H., Pedrón, A. M., Hermosilla, D. M., & Zamora, R. S. (2020). Robotic therapy for the hemiplegic shoulder pain: a pilot study. Journal of neuroengineering and rehabilitation, 17(1), 1-12.

[14] Arjunan, S. P., Kumar, D. K., & Panigrahi, B. K. (2016). Recognition of finger/hand grip mechanism by computing s-transform features of surface Electromyogram signal from healthy and amputee. Journal of Mechanics in Medicine and Biology, 16(06), 1650076.

[15] D. Sueaseenak, T. Chanwimalueang, C. Pintavirooj y M. Sangworasil, "An accurate forearm EMG signal classification method using two-channel electrode", IEEJ Trans. Elect. Electron. Eng., vol. 8, n. º 4, pp. 328–338, mayo de 2013

[16] Shair, E.F.; Ahmad, S.A.; Marhaban, M.H.; Mohd Tamrin, S.B.; Abdullah, A.R. EMG Processing Based Measures of Fatigue Assessment during Manual Lifting. Biomed Res. Int. 2017, 2017, 1–12.

[17] J. P. G. Calderón, D. F. G. Ruiz and R. R. Serrezuela, "Time-frequency analysis of the EMG signal for the identification of hand grasping postures," 2022 V Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC), San Jose, Costa Rica, 2022, pp. 1-8, doi: 10.1109/AmITIC55733.2022.9941272.

[18] Chowdhury, R.H.; Reaz, M.B.I.; Ali, M.A.B.M.; Bakar, A.A.A.; Chellappan, K.; Chang, T.G. Surface Electromyography Signal Processing and Classification Techniques. Sensors 2013, 13, 12431-12466.

[19] T. Jing Wei, A. R. B. Abdullah, N. B. Mohd Saad, N. B. Mohd Ali y T. N. S. B. Tengku Zawawi, "Featureless EMG pattern recognition based on convolutional neural network", Indonesian J. Elect. Eng. Comput. Sci., vol. 14, n. º 3, p. 1291, junio de 2019.

[20] Sengur A, Akbulut Y, Guo Y, Bajaj V. Classification of amyotrophic lateral sclerosis disease based on convolutional neural network and reinforcement sample learning algorithm. Health information science and systems. 2017; 5(1)

[21] Nawawi MA, Ismail FS, Selamat H. Comprehensive Pineapple Segmentation Techniques with Intelligent Convolutional Neural Network. Indonesian Journal of Electrical Engineering and Computer Science. 2018; 10(3): 1098-1195.

[22] Loshchilov, Ilya, and Frank Hutter. "Sgdr: Stochastic gradient descent with warm restarts. arXiv 2016." arXiv preprint arXiv:1608.03983 (2019).

[23] Buelvas, H. E. P., Montaña, J. D. T., & Serrezuela, R. R. (2023, October). EMG Signal Analysis for Hand Grip Posture Classification using Continuous Wavelet Transform. In 2023 VI Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC) (pp. 1-8). IEEE.

[24] Calderón, J. P. G., Ruiz, D. F. G., & Serrezuela, R. R. (2023, October). Deep learning of EMG signals in hand grip posture identification using time-frequency domain applying STFT. In 2023 VI Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC) (pp. 1-8). IEEE.

[25] Buelvas, H. E. P., Montaña, J. D. T., & Serrezuela, R. R. (2023, July). Hand Gesture Classification using Deep learning and CWT images based on multi-channel surface EMG signals. In 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) (pp. 1-7). IEEE.

[26] Cornejo, J., Barrera, S., Ruiz, C. H., Gutierrez, F., Casasnovas, M. O., Kot, L., ... & L'huillier, E. A. (2023). Industrial, collaborative and mobile robotics in Latin America: Review of mechatronic technologies for advanced automation. Emerging Science Journal, 7(4), 1430-1458.

[27] Serrezuela, R. R., Reyes, E. M., Zamora, R. S., & Leon, A. A. S. (2023). Perspective Chapter: Classification of Grasping Gestures for Robotic Hand Prostheses Using Deep Neural Networks. In Human-Robot Interaction-Perspectives and Applications. IntechOpen.

[28] Losada, C., Pastrana, B., & Rodriguez, R. (2022, September). Machine learning based feature extraction and classification of surface emg signals in the time domain. In 2022 V Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC) (pp. 1-8). IEEE.

Table 2: Accuracy Score "One to One" Approaches.

	32x32				64x64				128x128			
Patient	Accuracy Score	F1 Score	Recall Score	Precision Score	Accuracy Score	F1 Score	Recall Score	Precision Score	Accuracy Score	F1 Score	Recall Score	Precision Score
P01	94,13	94,13	94,29	94,21	93,33	93,33	93,45	93,44	94,48	94,48	94,59	94,52
P02	84,34	84,34	84,34	84,57	85,14	85,14	85,18	85,23	87,46	87,46	87,42	87,34
P03	87,1	87,1	86,93	86,91	87,46	87,46	87,32	87,34	87,81	87,81	87,89	88,01
P04	95,02	95,02	95,04	95,06	92,88	92,88	92,86	92,99	92,44	92,44	92,3	92,6
P05	90,57	90,57	90,45	90,38	90,57	90,57	90,39	90,34	92,44	92,44	92,54	92,49
P06	92,79	92,79	92,76	92,74	91,55	91,55	91,71	91,6	91,1	91,1	91,11	91,35
P07	94,13	94,13	94,25	94,23	93,51	93,51	93,56	93,67	93,42	93,42	93,49	93,49
P08	90,57	90,57	90,64	90,53	91,37	91,37	91,43	91,31	90,48	90,48	90,46	90,55
P09	89,23	89,23	89,15	89,08	91,99	91,99	91,84	92	90,04	90,04	89,89	90,26
P10	90,57	90,57	90,65	90,7	92,7	92,7	92,73	92,79	91,64	91,64	91,55	91,56
Average	90,845	90,845	90,85	90,841	91,05	91,05	91,047	91,071	91,131	91,131	91,124	91,217

Table 4: Metrics "K-fold cross validation" Approaches

	32x32			***64x64***			***128x128***		
Patient	*Average Accuracy Train*	*Average Accuracy Test*	*Average Accuracy Validation*	*Average Accuracy Train*	*Average Accuracy Test*	*Average Accuracy Validation*	*Average Accuracy Train*	*Average Accuracy Test*	*Average Accuracy Validation*
P01	*98,5*	*91,01*	*91,01*	*99,52*	*91,99*	*92,23*	*99,65*	*92,7*	*91,97*
P02	*94,07*	*78,82*	*78,15*	*98,64*	*82,11*	*83,03*	*99,06*	*82,56*	*82,05*
P03	*96,03*	*80,42*	*80,16*	*98,3*	*80,42*	*82,81*	*98,37*	*81,13*	*83,01*
P04	*98,06*	*90,21*	*89,83*	*98,78*	*90,3*	*90,07*	*99,5*	*89,67*	*90,66*
P05	*98,46*	*88,61*	*88,64*	*99,45*	*90,03*	*89,18*	*99,58*	*89,59*	*89,54*
P06	*98,41*	*89,85*	*88,72*	*99,44*	*90,65*	*87,99*	*99,47*	*87,72*	*88,4*
P07	*98,99*	*91,19*	*91,19*	*86,89*	*79,35*	*80,77*	*99,7*	*90,39*	*91,41*
P08	*98,21*	*88,07*	*86,44*	*99,33*	*87,54*	*86,17*	*99,65*	*86,56*	*86,62*
P09	*98,29*	*83,62*	*85,77*	*99,26*	*86,2*	*87,29*	*99,39*	*86,74*	*86,84*
P10	*99,13*	*89,05*	*87,46*	*99,21*	*90,03*	*88,18*	*99,54*	*87,54*	*89,29*
Average	*97,815*	*87,085*	*86,737*	*97,882*	*86,862*	*86,772*	*99,391*	*87,46*	*87,979*