# Hand Gesture Classification from Surface Electromyography signals*

*Note: based on GrabMyo

1st Ismael Peña
*dept. Artificial Intelligence and Data Science (of Aff.)*
*Panamerican University (of Aff.)*
Aguascalientes, Mexico
0283027@up.edu.mx

***Abstract*—There are many works related to gesture recognition via electromyography; efficient detection of muscle inactivity remains a challenge. This document analyzes electromyographic signals of hand gestures to be categorized. Initially, it is necessary to understand correct electromyographic activity in the future to recognize its negative case, meaning inactivity, which often arises from communication problems in motor neurons.**

***Index Terms*—surface electromyography, hand gestures, muscle activity**

## I. Introduction

Electromyographic signals are electrical impulses generated by the depolarization of muscle fibers during muscle contraction. This signal is controlled by the motor neurons of the nervous system; since the human body is intrinsically electric, these signals have noise, which complicates their analysis. For this, there are two types of EMG detectors: intrusive and non-intrusive. Intrusive EMG is operated with needle-like electrodes inserted into muscle fibers near the nerve terminals, usually causing extreme pain in patients with demyelination. Non-intrusive EMG is known as surface EMG, which is what this document is interested in.

## II. Methods

### A. What data was used

For the classification of hand gestures, data from Grabmyo was used. This dataset was taken from 43 people, in 3 sessions, for 17 gestures across 7 attempts. A surface EMG was used, and the sample-taking configuration is visualized in the following image, where different electrodes were used; as a result, there are 32 channels at a frequency of 2048 hertz per movement in a 5-second interval.

## III. Classification

To classify the gestures, three types of neural networks were tested, namely ResNet, DenseNet, and TCN (a varying CNN). The algorithms were trained on an RTX 5070 ti mobile GPU.

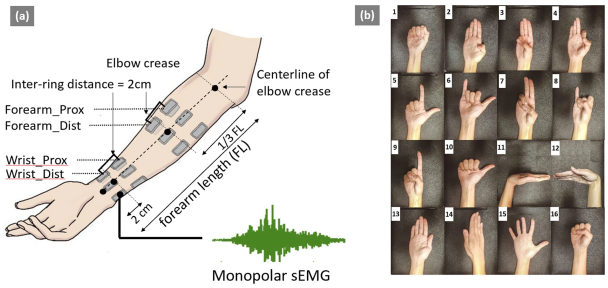


Fig. 1. Gestures made while sampling with a sEMG.

### A. ResNet

ResNet was tested first, and since the primary purpose of this type of network is image classification, the electromyographic signals were printed as images. This took considerable computational time, using up to 2 hours for conversion; a sample image is shown below.

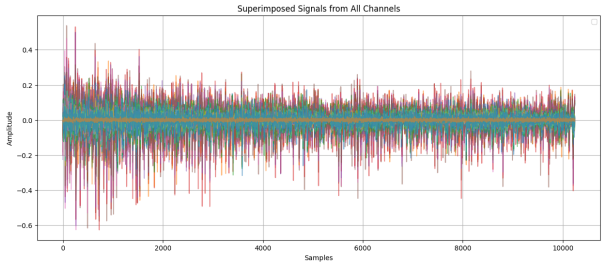


Fig. 2. A high res png image including 32 channels and full sampling.

During the training stage, it was impossible to use all samples with images that contained all the information of the signals because the image loading in memory exceeded available resources, so the image was separated by channels to see if training with fewer samples but across different channels was feasible.

Using this strategy, it was possible to obtain training and testing results; however, the outcome was genuinely disappointing, with an accuracy of 2 percent. This was even using

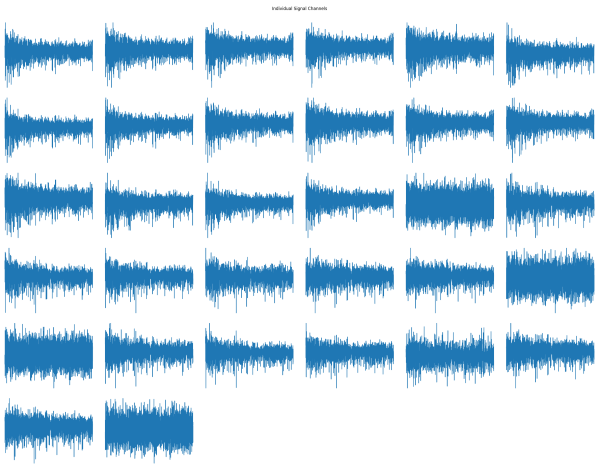

Fig. 3. A png image split by 32 channels and full sampling.

the data of one person over 5 gestures across 7 attempts with 32 channels, which was the maximum allowed due to memory resource limitations.

Another strategy used was to stretch the image, as it is visibly clear that it would be difficult to find a pattern. A sample image is shown below.

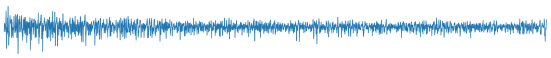

Fig. 4. A png image of 1 channel and full sampling.

In this experiment, there were no significant changes in accuracy or training time, which resulted in negligible differences.

As a last resort, the sampling of the signal was reduced, meaning reducing the sample of the 5 seconds of data extracted, from 10240 to 1000.

```
signal, fields = wfdb.rdsamp(record_path,
    sampto=1000)
```

In addition to the reduction, the image was also exported as jpg, as it was previously exported as png, resulting in 4 channels; the transparency channel is usually composed of a byte 255 or -1 depending on the desired representation. The 32 channels were also preserved. A sample image is shown below.

For this case, it was possible to use data from 20 participants, 17 gestures, and 7 attempts, with a total of 2380 data across 32 channels, meaning virtually 76160 samples of data.

However, the highest accuracy obtained was 5.88 percent. It is clear that an attempt is being made to fit a problem to an inadequate model. ResNet was pre-trained with the ImageNet dataset, which mainly contains animals. Being somewhat objective, the images that are classified using ResNet would mainly be those that a human could recognize at a glance, and since electromyographic signals are not easily categorizable at first sight, a different strategy will be required.

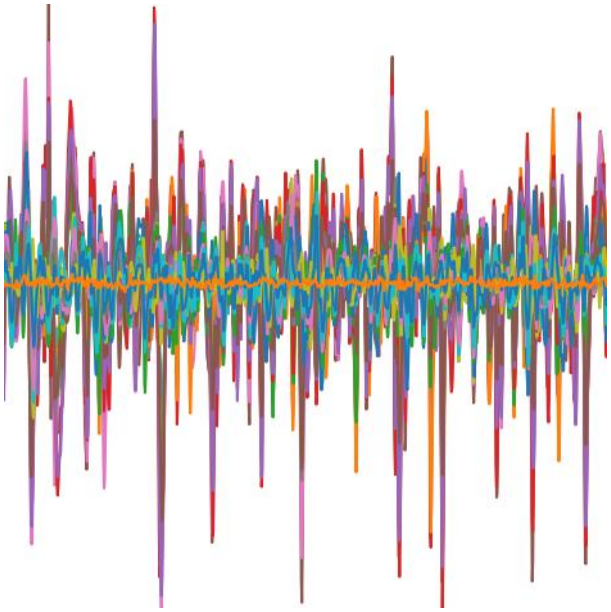

Fig. 5. A jpg image including 32 channels and 1000 of sampling.

TABLE I
MODEL PERFORMANCE COMPARISON

| Model | Val Accuracy | Test Accuracy | Num Parameters |
|---|---|---|---|
| ResNet | 5.88% | 5.92% | 272,378 |

### B. DenseNet

In this model, the use of CIFAR10 was eliminated, as instead of classifying images per se, EMG signal data will be used as a spectrogram, meaning that information such as pixels will not be used, but rather raw data. Below is a sample shown as a spectrogram.

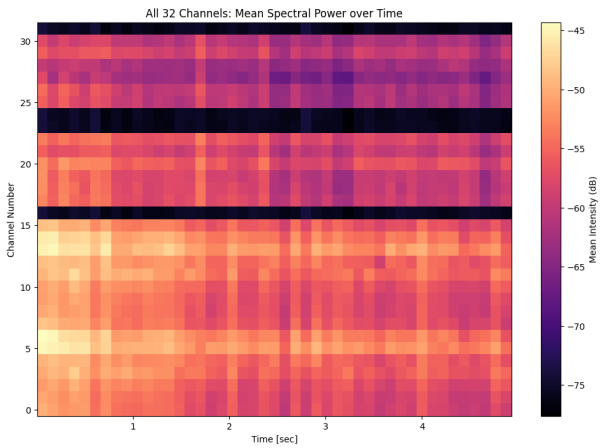


Fig. 6. A spectrogram png image including 32 channels and 10240 of sampling.

The spectrogram is displayed as a heat map; however, one can notice the vast difference between this and a wave graph.

Here the information is completely laid out on a 2D plane, which is good for a CNN.

This time it was possible to train the model with data from the 43 participants, 17 gestures, 7 attempts, and 32 channels, however reducing the sampling to 1000 as in the ResNet case due to memory resource issues.

The resulting accuracy rose significantly to 24.12 percent, indicating that using the spectrogram could be a better path rather than using a wave as an image. One can notice the loss of information if we generate an image for a sampling of 1000, which is shown below.

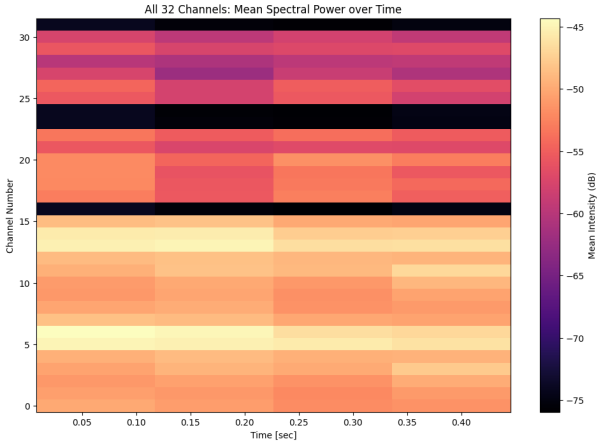


Fig. 7. A spectrogram png image including 32 channels and 1000 of sampling.

### C. Convolutional Neural Network

In the last case, a simple CNN was used to classify spectrograms, a network of two convolution blocks, first eliminating possible noise through a convolution of 64 filters, and then with one of 128 filters, the data is flattened so that it can be classified linearly.

### D. Mathematical Model

The neural network can be defined as follows: Given an input spectrogram $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, the output of each convolutional layer is defined by:

$$\mathbf{H}^{(l)} = \sigma\left(\mathbf{H}^{(l-1)} * \mathbf{K}^{(l)} + b^{(l)}\right) \tag{1}$$

where $*$ denotes the 2D convolution operation, $\mathbf{K}^{(l)}$ is the set of learnable filters, $b^{(l)}$ is the bias, and $\sigma$ is the ReLU activation function defined as $\sigma(z) = \max(0, z)$.

Subsequently, a subsampling operation (*Max Pooling*) is applied that reduces spatial resolution:

$$y_{i,j} = \max\{x_{m,n} : (m, n) \in \mathcal{R}_{i,j}\} \tag{2}$$

where $\mathcal{R}_{i,j}$ represents the local region of the pooling window. The final prediction of the model is obtained through a linear transformation after flattening the features:

$$\hat{\mathbf{y}} = \text{Softmax}(\mathbf{W} \cdot \text{flatten}(\mathbf{H}^{(2)}) + \mathbf{b}) \tag{3}$$

The results from this network were promising, as it was possible to use almost all data, meaning 3 sessions from 43 participants, 17 gestures, and 7 attempts across 32 channels, still working with 1000 samplings.

The accuracy obtained at this point was 34.58 percent.

Source code in gesture_classification repository in github

## REFERENCES


[1] Jun Kimura, Electrodiagnosis in Diseases of Nerve and Muscle: Principles and Practice.
[2] https://www.nature.com/articles/s41597-022-01836-y.
[3] https://physionet.org/content/grabmyo/1.1.0/