Electromyogram Signal Analysis using Python

Khyati Tripathi
Reg. No: 20BEC0213
Mobile no: 8935938141
Electronics and Communication Engineering
Vellore Institute of Technology
Vellore, India
khyatitripathi511@gmail.com

Rittika Deb
Reg. No: 20BEC0257
Mobile no: 9007822828
Electronics and Communication Engineering
Vellore Institute of Technology
Vellore, India
rittika.deb86@gmail.com

Jagruti Mahapatra
Reg. No: 20BEC0395
Mobile no: 7848047089
Electronics and Communication Engineering
Vellore Institute of Technology
Vellore, India
jagrutirkm@gmail.com

Kokila Gulati
Reg. No: 20BEC0402
Mobile no: 9759776271
Electronics and Communication Engineering
Vellore Institute of Technology
Vellore, India
kokila.gulati.23@gmail.com

***Abstract*—The main objective of our project is to analyze the various aspects and features of EMG signal and obtain important statistical information from it. It can often be challenging to identify specific hand motions or even anomalies because they are frequently subtle or challenging to assess. For movement classification, different machine learning techniques are employed like SVM, LR, Neural Network, Naive Bayes, K-means and GMM. In this work, there are six criteria for classifying motion. EMG data from five different people with six hand signals were used in the project. The classification results were compared in terms of training time parameters and classification accuracy.**

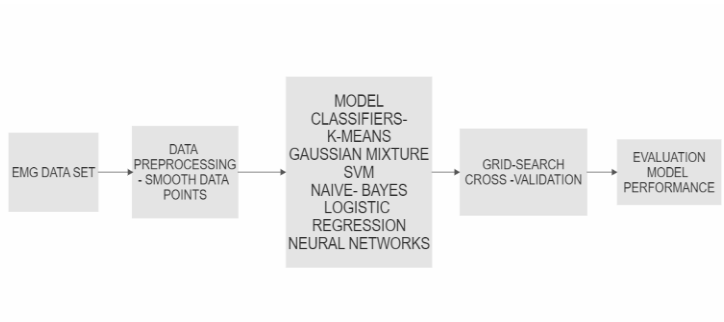
Keywords—Electromyography, FFT, power spectral analysis, ARV, feature extraction, myopathic,

I. INTRODUCTION

The EMG signal or the Electromyogram signals are a segment of signals which have found their application in the field of biomedical sciences wherein electrical signals generated in muscles are analyzed for neuromuscular activities of contraction and relaxation . The EMG signal is a weak signal as the amplitude lies in between 1-10 mV, making

For activated muscles a relative motion is observed between the muscle, skin and the electrode for decrease and increase in length of the muscle accordingly. At that time, the electrodes will show some movement artifacts that will generate some signals which we called earlier as the electromyogram signal or simply the EMG signal. After acquiring an EMG signal, the features will be obtained from the EMG data. Feature extraction is a technique to extract essential information from an EMG signal and eliminates the undesired signal. An EMG converts these signals into numbers or graphs to help doctors to make a diagnosis.

In this paper we are taking into account a dataset- which will be further divided into 2 databases of testing and training respectively- of EMG signals and passing it through six classifiers to detect various hand gestures and make an analysis on the accuracy of the classifier.



I.II. LITERATURE REVIEW

A. Frequency Analysis Of EMG Signals With Matlab Sptool

Abstract- Wavelet analysis, neural networks, and pattern recognition techniques are being developed in the field of biomedical digital signal processing (DSP) for the analysis of EMG signals (produced by the muscles) in neuromuscular diseases and CTG signals (the cardiotocogram) during childbirth. Since these signals have historically been particularly challenging to quantify, clinical quantification calls for novel ways to analysis. Real-time clinical applications are being developed for DSP hardware and software systems. By calculating the median and average frequencies and looking at how EMG signals behave in the frequency domain, the goal is to identify EMG signals. Fast Fourier transform and digital filters were key components in obtaining the results needed to establish these characteristics.

EMG Signal Analysis in the Frequency Domain By measuring and computing the parameters that characterise these signals' properties, EMG signals are analysed. Fast Fourier transform is frequently used to calculate the signals' power spectrum densities. Median frequency and average frequency are related as follows:

$$\int_0^{f_{med}} Sm(f)df = \int_{f_{med}}^{\infty} Sm(f)df = \frac{1}{2}\int_0^{\infty} Sm(f)df$$

where Sm(f) is the signal's power spectrum density. The median and average frequency are the most reliable variables in an EMG examination. Median frequency is less vulnerable to noise than average frequency. Low-level contractions with low signal-to-noise ratios are one symptom of this issue. The bandwidth, which also unambiguously defines time and variation, defines the spectrum. Additionally, it offers critical information on EMG signal filtering techniques. Applications of EMG Signal Analysis Methods: Detection, Processing, and Classification

For therapeutic and biomedical purposes, the development of Evolvable Hardware Chips (EHW), and contemporary human-computer interaction, electromyography (EMG) signals can be exploited. In order to identify, decompose, analyse, and classify EMG signals obtained from muscles, advanced techniques are required.

EMG: Background on Anatomy and Physiology Electromyography is also known as EMG. Muscle electrical impulses are the subject of this research. Myoelectric activity is a term used to describe EMG. The term "muscle action potential" refers to the electrical signals carried by muscle tissue, which function similarly to nerves in this regard. The data in these muscle action potentials can be captured via surface EMG. Two key problems that need to be addressed when detecting and recording the EMG signal affect the signal's fidelity. Signal-to-noise ratio is top on the list. in other words, the energy ratio of the EMG signals to the energy of the noise signal. Electrical signals that do not belong to the targeted EMG signal are generally referred to as noise. The distortion of the signal is the other problem, hence the relative importance of any frequency component in the EMG signal shouldn't be changed. Equation illustrative of a basic model of the EMG signal

$$x(n) = \sum_{r=0}^{N-1} h(r)e(n-r) + w(n)$$

where w(n), zero mean addictive white Gaussian noise, h(r), and e(n), point processed, represent the firing impulse, MUAP, and x(n), modelled EMG signal N represents the total number of motor unit firings. The electrode picks up the signal and amplifies it. Analysis and classification of surface EMG signals acquired for the use of a prosthetic limb.

B. Surface EMG Signal Acquisition Analysis and Classification for the operation of a Prosthetic Limb

ABSTRACT- Biomedical signal processing is a significant subfield in prosthetics.. Due to their compatibility with human body biomechanics, electromyogram (EMG) signals are used in prosthetic design. The purpose of this study is to evaluate Surface Electromyogram (SEMG) signal parameters for one test subject related to upper limb speed and flexion angle. With little ethical concerns, noninvasive SEMG signal capture was done for elbow flexion in the upper limb. Surface EMG signals were captured, amplified by an INA128 amplifier IC, and filtered into the 0Hz–500Hz band using a UAF42 filter IC. To categorise elbow flexion with respect to speed, offline SEMG signal speed was classified using Fast Fourier Transformation, Wavelet Transformation, and MATLAB R2015a software. The difference between a quick and slow elbow flexion could be seen graphically by displaying the amplitude fluctuations in each transition. By employing a goniometer to roughly calculate the flexion angle, data were collected using an Arduino ATMEGA 2560 microcontroller. Future studies will use the curve fitting algorithm to link SEMG signals with flexion angle. A generic method for speed classification is the ultimate objective.

Signal View and Storage: Because MATLAB R2015a is compatible with the Beagle bone, it was chosen to be used to visualise the processed EMG signals. A hardware support package needed to be installed in MATLAB in order to establish connection between the software and the Beagle bone. As a result, MATLAB issued specific operational instructions to connect to Beaglebone. For the signal visualisation and storage, we used the MATLAB Simulink package rather than the standard MATLAB command window. Each command on this platform is represented by a block. By doing this, we were able to design a graphical framework for visualising EMG signals. Before plotting the signal output, acquired EMG signals were processed through digital filters. The power line noise was reduced using 50Hz and 60Hz notch filters. Additional 30Hz notch filters were applied to get rid of baseline noise. The EMG signal was seen using a vector scope block, and the data was stored in a MATLAB workspace using a simout block. Analysis of EMG Signals Using Wavelets Signal domain wasn't altered during wavelet analysis. Due to the analysis being done in the time domain, this method is more efficient than Fast Fourier Transformation.

II.III. PROPOSED METHODOLOGY

Our categorization model construction process will consist of the following five high level steps:

1. Smoothing and removing noise from data points during preprocessing.

2. Dimensionality reduction using ISOMAP and principal component analysis.

3. Scale the data using conventional methods.

4. Constructing different model classifiers using SVMs, k-means, and gaussian mixture, Neural networks, logistic regression, and naive bayes.

5. Using gridsearch and cross-validation to hypertune each model and choose the best one.

6. Examine model performance in comparison to a test set and contrast outcomes.

III.IV. EXPERIMENTAL SETUP

A. DATA SET

In the project we have taken data set from an open source dataset: https://archive.ics.uci.edu/ml/datasets/sEMG%20for%20Basic%20Hand%20movements

It contains two databases containing EMG data of participants doing a variety of hand movements. These hand motions can be divided into six categories-

- Spherical
- Tip
- Palmar
- Lateral
- Cylindrical
- Hook

Each row in the dataset corresponds to a participant's single trial of EMG data recordings.Each column will reflect a specific time point for each trial because the EMG data is recorded over time.We only use the first database, which comprises of 5 participants(3 female and 2 male) and 6 hand signals. We will use a matrix of 900 data points and 6000 characteristics as our dataset.

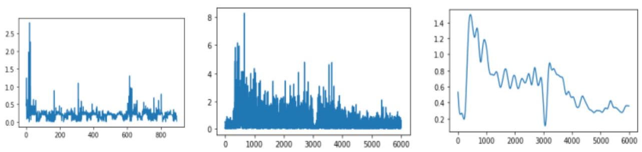
B. DATA PROCESSING AND TRANSFORMATION

We discovered that the EMG data were extremely noisy, leading to subpar model classification performance.
without modification.
We undertook two changes to raise the quality of the data:

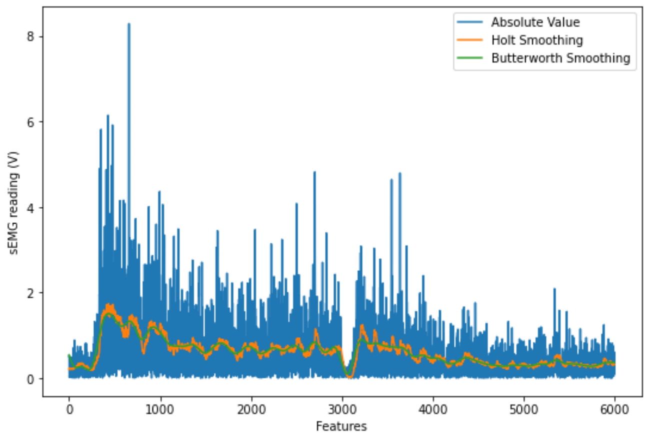
- Reduce the standard deviation and volatility by applying the absolute value function to each datapoint.
- Use Holt-Winters exponential smoothing to reduce the data's noise to a large extent.

From Left to right, the diagram below demonstrates the data processing steps in our preparation pipeline.

- The original EMG data from the raw data source is on the left.
- The EMG data after applying the absolute value is in the middle.
- The image on the right is a starting point for the remaining analyses for the project.

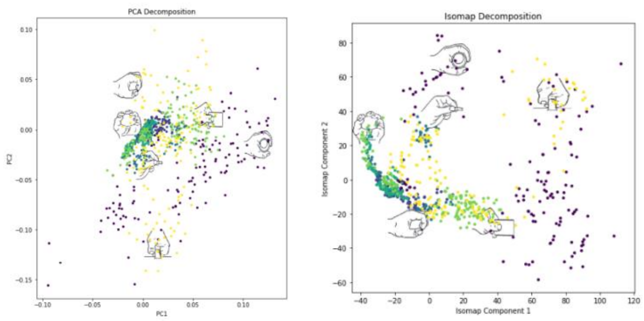


To make sure the smoothing settings were appropriate, they were adjusted and visually compared to the raw data.
Smoothed curve reduced noise while maintaining data patterns. The ideal degree of blending was 0.03, while 0.02 is the ideal smoothing slope.
Here is an illustration of a single datapoint:
After the output's smoothing and taking the absolute value:



Dimensionality Reduction with PCA and Isomap

Each data point comprises 6000 features, each of which denotes a specific moment in time.
We used principal component analysis (PCA)to determine the principal factors that best describes the variation of our model. In a similar vein, we can utilize Isomap, a nonlinear dimensionality reduction method. reducing strategy. The first two components are plotted using each of these methods for visual comparison.



The hand motion associated with each colour is displayed by superimposing the hand photos. As you can see, neither approach "wins" the dimensionality reduction in a blatant way. You can anticipate this mostly because we'll probably want to employ more than simply two factors for classification, which is It's difficult to picture. Since high-dimensional datasets are more frequently used for isomap, the Isomap data will be used for the categorization analysis below, with the first 20 features. To illustrate the decision boundaries of the models, we also construct models with only two components.

C. MODEL FORMULATION, CREATION & TUNING

IV.Support Vector Machine Classifier

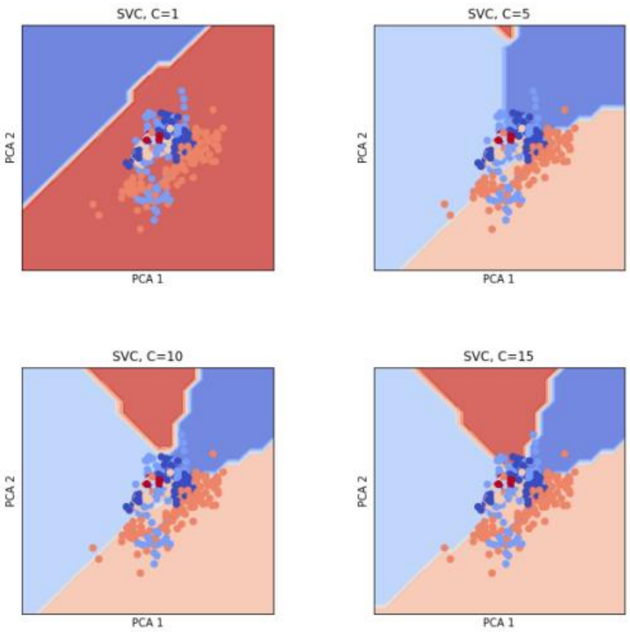
Using a variety of kernels, Support Vector Machines are a useful tool to distinguish data points.We reasoned that SVM would be a decent model to attempt for this issue given that we had six distinct classes with high dimensionality.The hand gesture classification problem for SVM is written as follows:

$$\min ||w| \; w,b \; | \; 2 \text{ s.t.} y \; (+ b) \geq 1, \forall$$

In layman's terms,to reduce total training error, the formula maximizes the soft margins among the six diverse hand movements.. The above formulation is for the linear kernel; however, we try different kernels because we suspect that many of the features between the different hand gestures may overlap. A different kernel gives these noisy boundaries more flexibility.

We experimented with a range of parameters and kernels using SVM from sklearn using the generalized SVM approach. By integrating an SVM model with two PCA components, the images were produced in order to visualize the decision boundaries.

The following diagram illustrates our model's decision boundaries for various parameter values:



We stayed with the ISOMAP's 20 components for the best-performing model and hypertuned the parameters. As a result, different kernels, gamma values, margin values, and degrees were investigated.

V.K-Means Clustering Classifier

One of the techniques we used to try to predict hand gestures was K-Means clustering classification. K-Means is an unsupervised learning algorithm that does not require response labels. Instead, we numerically encoded our labels (hand gestures from 1-6). The following is the outcome of building our models and using PCA=2:

We have six clusters that are colour coded differently. The white Xs represent the clusters' centres. The model does a good job of separating the training points. However, because this is an unsupervised model, one of the difficulties we encountered was associating the original hand gesture labels to the clusters.

VI.Gaussian-Mixture Model

We built a Gaussian-Mixture model with processed data because we believe that each class of hand gestures could be presented by a unimodal distribution using the top principal components. Our gaussian mixture model for this problem was formulated as the following:

We initialized , and to be the identity matrix. Then we run the expectation maximization algorithm below until we maximize the likelihood (convergence).

Expectation Step:

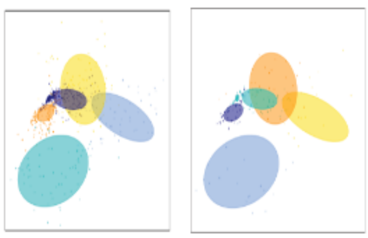
Maximization Step:

$$\pi_k = \frac{\sum_i \tau_k^i}{m}$$

$$\mu_k = \frac{\sum_i \tau_k^i x^i}{\sigma_i \tau_k^i}$$

$$\sum_k = \frac{\sum_i \tau_k^i (x^i - \mu_k)(x^i - \mu_k)}{\sum_i -\mu_k}$$

The result of formulating and training the model can be seen as:



Hand Gesture Gaussian Motion Classification   Hand Gesture Gaussian Motion Classification

VII.Naive-Bayes Classifier

The predictors' independence is one of Naive-underlying Bayes's presumptions. Given that time series data is subject to autocorrelation and that each successive characteristic is hence loosely connected, we now know that this is obviously not the case with our dataset. This model's performance was expected to be the weakest. Using the general naive bayes classifier, we created the model:

1. The likelihood of each hand motion in the dataset, p(y), serves as the class priors.
2. Determine the posterior probability of the training set using the Bayes formula, paying particular attention to how frequently the features lead to each of the classes given the features:

$$P(y = i|x) = \frac{P(x|y)P(y)}{P(x)}$$

3. Increase the possibility that every data point represents the right action.

Given all of its fundamental data assumptions, the Naive-Bayes model is fundamentally straightforward. For NB, there was not much hypertuning required; it simply required a few builds to produce reliable results.

V. Multinomial Logistic Regression Classifier

A predictor function is created by linearly combining features in the probabilistic classification method known as logistic regression. In this function, regression is applied in a manner akin to straightforward linear regression. The logistic function changes this linear function to produce probabilistic categorization. The following describes the generalized logistic function:

$$p = \frac{1}{1 + e^x}$$

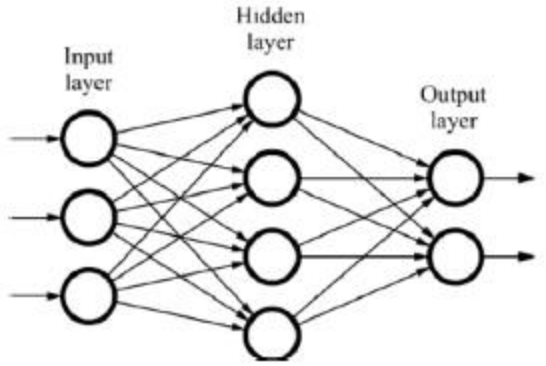
The Multinomial Logistic Regression equation becomes:

$$P(Y_i = K) = \frac{1}{1 + \sum_{k=1}^{K-1} e^{\beta_k X_i}}$$

when this is applied to the various categories and variables where K stands for the potential outcomes and must equal 1. This model's primary premise is that the data are case-specific, with a single value for each independent variable in each case. The independent variables do not have to be statistically independent from one another, unlike Naive-Bayes.

VI. Neural Network Classifier

The final model utilised to categorise the EMG data was a straightforward feed-forward neural network model. Although there are many various kinds of neural network models, we only covered the straightforward feed-forward model in detail. Example of this model in the picture below.



V. RESULTS

Gridsearch and cross validation were used to hypertune all of our models, and we selected the models
which was the most accurate. The results are as follows:

Model	Accuracy	Standard Deviation
SVM	76%	3.3%
K-means	62%	6.5%
Gaussian Mixture	78%	12.7%
Naive-Bayes	59%	3.4%
Neural Network	72%	4.4%
Logistic Regression	58%	2.6%

A. Gaussian Mixture
The gaussian mixture model had a high level of accuracy, averaging close to 78%. The standard deviation is 11.4%. This is most likely due to the model's performance being heavily dependent on initial starting points. Many of the EMG points appeared to be randomly distributed around the clusters.

B. Support Vector Machine
The SVM test accuracy was 76%, with a remarkably low standard deviation of 3.3%. Unlike most other models, the accuracy did not differ significantly when using raw data versus smoothed and transformed data. SVM performs well when there are more features than data points.

C. Neural Networks
The accuracy of the Neural Network model was almost identical to that of the SVM mode (72%), but with a slightly higher standard deviation. The complexity of this methodology may be greater than what is required to explain the data, which could explain why the Neural Network model does not return the best model.

D. Naive-Bayes
With 59% accuracy, Multinomial Naive-Bayes performed relatively well in comparison to some of the other models. Because the model contains less randomness, the standard deviation was low.

E. Logistic Regression
Logistic Regression showed poor accuracy at 55%. Since this model is actually a neural network we would anticipate that the logistic regression will not function on a network without any hidden layers. Logistic regression did not perform well because of nonlinear and highly correlated classes. This would prevent the coefficients of the regression to not explain each feature completely.

F. K-Means
K-Means clustering performed with accuracies averaging around 62%. K-Means couldn't explain the
variation in the response with additional feature information.The reason is because their is a robust separation between the data points features; something that the Gaussian-Mixture model was
able to deal with it.

VI. CONCLUSION

In this work, six different hand motions were categorized using EMG signals from 5 individuals. For classification, 6 fundamental machine learning techniques were applied. The acquired classification results were compared to the parameters for classification accuracy and standard deviation. The Gaussian Mixture Trees model, with an average classification accuracy of 77.77% among these models, has been found to have the highest classification performance according to the test findings. The raw EMG data was categorized by the Gaussian Mixture table with exceptional performance.

The algorithms that perform best based on the results will be used and evaluated in actual applications in the following study.

VIII.VII. REFERENCE

Artameeyanant, P., Sultornsanee, S., & Chamnongthai, K. (2016). An EMG-based feature extraction method using a normalized weight vertical visibility algorithm for myopathy and neuropathy detection. *SpringerPlus*, *5*(1), 1-26.

OMAR, S. N., Saad, N. M., Shair, E. F., & Zawawi, T. N. S. T. (2021). Application of digital signal processing and machine learning for Electromyography: A review. *Asian Journal Of Medical Technology*, *1*(1), 30-45.

Pauk, J. (2008). 419. Different techniques for EMG signal processing. *Journal of Vibroengineering*, *10*(4).

Seyidbayli, C., Salhi, F., & Akdogan, E. (2020). Comparison of machine learning algorithms for EMG signal classification. *Periodicals of Engineering and Natural Sciences (PEN)*, *8*(2), 1165-1176.

Singh, A., Dutta, M. K., & Travieso, C. M. (2017, October). Analysis of EMG signals for automated diagnosis of myopathy. In *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)* (pp. 628-631). IEEE.

Strzecha, K., Krakós, M., Więcek, B., Chudzik, P., Tatar, K., Lisowski, G., ... & Sankowski, D. (2021). Processing of EMG signals with high impact of power line and cardiac interferences. *Applied Sciences*, *11*(10), 4625.

Paul, Y., Goyal, V., & Jaswal, R. A. (2017, September). Comparative analysis between SVM & KNN classifier for EMG signal classification on elementary time domain features. In *2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)* (pp. 169-175). IEEE.

Gupta, A., Sayed, T., Garg, R., & Shreyam, R. (2017, August). EMG signal analysis of healthy and neuropathic individuals. In *IOP Conference Series: Materials Science and Engineering* (Vol. 225, No. 1, p. 012128). IOP Publishing.

Sadikoglu, F., Kavalcioglu, C., & Dagman, B. (2017). Electromyogram (EMG) signal detection, classification of EMG signals and diagnosis of neuropathy muscle disease. *Procedia computer science*, *120*, 422-429.

Fattah, S. A., Iqbal, M. A., Jumana, M. A., & Doulah, A. S. U. (2012). Identifying the motor neuron disease in EMG signal using time and frequency domain features with comparison. *Signal & Image Processing*, *3*(2), 99.

Saad, I., Bais, N. H., Seng, C. B., Zuhir, H. M., & Bolong, N. (2015, December). Electromyogram (emg) signal processing analysis for clinical rehabilitation application. In *2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)* (pp. 105-110). IEEE.

Abbaspour, S. (2019). *Electromyogram signal enhancement and upper-limb myoelectric pattern recognition* (Doctoral dissertation, Mälardalen University).

Singh, S., Mishra, B., & Wadhwani, A. K. Electromyography (EMG) based signal processing and their Feature Selection using Wavelet Transform.

Sattar, N. Y., Syed, U. A., Muhammad, S., & Kausar, Z. (2019, July). Real-time EMG signal processing with implementation of PID control for upper-limb prosthesis. In *2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)* (pp. 120-125). IEEE

Abbaspour, S., Lindén, M., Gholamhosseini, H., Naber, A., & Ortiz-Catalan, M. (2020). Evaluation of surface EMG-based recognition algorithms for decoding hand movements. *Medical & biological engineering & computing*, *58*(1), 83-100.

Zawawi, T. T., Abdullah, A. R., Shair, E. F., Halim, I., & Rawaida, O. (2013, December). Electromyography signal analysis using spectrogram. In *2013 IEEE Student Conference on Research and Developement* (pp. 319-324). IEEE.

Chowdhury, R. H., Reaz, M. B., Ali, M. A. B. M., Bakar, A. A., Chellappan, K., & Chang, T. G. (2013). Surface electromyography signal processing and classification techniques. *Sensors*, *13*(9), 12431-12466..