Comparing Performance of Different Classifier Methods on HDEMG Hand Gesture Signals using Python

Pallavi Kulkarni, pkulkarni1@mail.sfsu.edu, SID : 921474174, Dept. Of EECS, SFSU

***Abstract* — Aim of the project is to evaluate the performance of the Machine learning methods to classify hand gestures from the high definition Electro-myo-graphic (EMG) data. Publicly available CSL-HDEMG data was chosen which is obtained using 8x24 grid electrode array strip sampled at a rate of 2048 Hz for an average of 3 seconds. Twenty-Seven gestures repeated ten times by five subjects for five sessions are chosen for this study. Four time-domain features namely mean absolute value (MAV), zero crossing (ZC), sign slope change (SSC) or turns count (TC) and Willison amplitude value (WAV) are computed from the raw EMG data for each channel. For classification, two categories of learning methods are chosen. The popular and still widely used shallow learning methods in the literature namely Linear Discriminant Analysis (LDA), Naive Bayes, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) are used for training and testing. The classical deep learning network, multi-layer perceptron (MLP) is used in this study. Finally, testing accuracy of the methods are computed. It can be concluded that the best results were achieved with the LDA classifier, while the rest of the classifiers achieved acceptable results.**

I. INTRODUCTION

The applications of wearable sensors in daily life are increasing every day. The wearable sensor technologies provide a wide range of information that can be utilized for human-machine interface (HMI). Among many HMI devices, the electromyography (EMG) is one of the most simple and effective wearable sensors, and is commonly employed in medical diagnosis. EMG provides a great amount of information and can be used to decipher human mechanical activities.

EMG signals represent neuromuscular activity and are effective biological signals for expressing movement that, among many other uses, can be used for external device control. EMG based HMIs have been widely applied in biomedicine and aerospace. In the field of rehabilitation engineering, EMG signals are primary neural based control sources for enabling powered upper limb prostheses, powered orthoses and exoskeletons, rehabilitation robots, robotic wheelchairs, and assistive computers. One way the muscular activity is interpreted is by using pattern recognition on EMG signals. Pattern recognition techniques are an advanced, intelligent signal processing technology and have shown to be reliable and consistent in classifications of user intent. Beyond signal magnitude, a typical pattern recognition algorithm extracts a set of features that characterize the EMG signals and then classifies the user's intended movement. Previous studies have evaluated the ability of various EMG features and classifiers to recognize user intent. The comparison of classification accuracies resulting from utilization of different types of classifiers and EMG features demonstrated that the type of classifier used does not significantly affect the classification performance, rather the choice of features has a significant impact on classification performance.

Among many user actions that can be interpreted from the EMG signals, hand gesture recognition has been studied extensively. Hand gestures can be interpreted by two broad categories of detection - visual based or internal sensor based. Visual-based gesture recognition perceives gestures by analyzing the images or videos obtained from a camera. Such detection methods typically suffer from various drawbacks including sensitivity to light, changing distance, hand motion modeling complexity, and position. On the other hand, hand gesture recognition systems based on internal sensor detection such as sEMG or HDEMG signal depend on the measurement of electrical signals exchanged between neurons that control motor muscles have been found to be more reliable and efficient.

In this study, the focus has been to compare the performance of pattern recognition methods and in particular the state-of-the-art machine learning algorithms, by training them on HDEMG data. In section 2, the experimental setup used to acquire the data and its specific attributes are detailed. In section 3, the details on pre-processing the data, extracting features and the machine learning methods used in this study are presented. Finally, in section 4, the results of the training, and testing of the methods along with their performance are compared and conclusions are derived.

II. Experimental Setup And Data Acquisition

There are two main designs of electrodes used to acquire EMG signals – in-situ (inside the skin) electrodes and surface electrodes. Among surface electrodes, there are two designs – gelled and dry. In this study, the data was acquired using dry surface grid electrodes. Wired electrodes, such as Myoware muscle sensor and wireless electrodes such as a Myo gesture control armband, are example of dry electrodes that are most preferred because of their simplicity and effectiveness. Furthermore, multiple dry electrodes can be used in the form of an array to collect data from different locations simultaneously. For this study, the publicly available CSL-HDEMG which is a HD sEMG database, published in [**1**] is used. In their setup, 192 electrodes organized in a grid of size 8 rows and 24 columns were utilized to record EMG signals from the forearm muscles of a subject while performing 27 different gestures. Five subjects were employed for data collection on five days. Each gesture was repeated 10 times.

A. Block Diagram

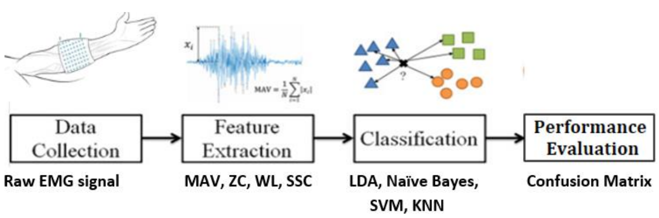


Fig 1. Block Diagram

The data acquisition is the first stage in studying hand gestures. The acquired data is processed to remove noise and unwanted parts of the signal. The feature extraction is the second stage that involves computing representative mathematical quantities. Choice of optimal features involves a thorough experimental research that can compactly represent the large amount of raw data. The computed features are subsequently used as inputs to classification methods that represent the backbone of sEMG signal analysis and processing. This is an area of active research and the recent trend in use of sophisticated machine and deep learning methods has improved the accuracy of predicting the gestures. For this reason, the focus of this study has been confined to evaluating the performance of such methods on the well-studied features computed from the acquired data . Each of the blocks from the block diagram of Figure. 1 are elaborated in the next three sub-sections.

B. Data Acquisition and Pre-processing

The experimental setup uses an electrode array with 192 electrodes arranged in a grid with 8 rows and 24 columns with an inter-electrode distance of 10 mm. Both electrodes and the amplifier are from OT-Bioelettronica[1]. Figure 2 shows a picture of the array. The array is connected via three flat cables, each one with 64 wires. A pre-amplifier is attached to every cable close to the electrodes to minimize noise caused by cable length. Bipolar recordings were made, which means the amplifiers takes the difference between consecutive electrodes resulting in 192 channels. In such a setup, every eighth channel does not contain meaningful data, because it represents the difference of the last channel in one column and the first channel in the next column. These channels are therefore ignored for the data analysis, leading to a total of 168 channels of usable data. As per the analysis performed in the original paper [**1**], it was shown that it is not necessary to consider data from all 168 channels for further processing but a subset of 20 to 80 channels gives a good trade-off between accuracy and complexity of the technical system. The is further realized in the results below.

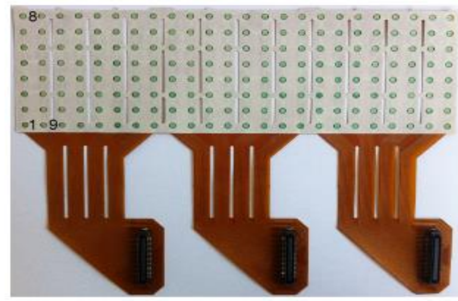


Fig 2. Picture of the electrode array used. The inter-electrode distance is 10 mm. The electrodes are arranged in an 8x24 grid. Electrode numbering starts at the lower left corner and works column wise.

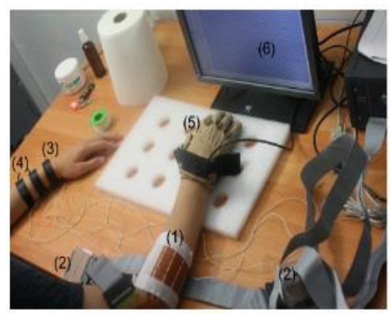


Fig 3. Picture of the actual setup used: (1) electrode array, (2) preamplifier, (3) Reference electrode, (4) DRL circuit, (5) data glove (not used), (6) real-time signal visualization,

Gesture set

A pre-defined set of 27 gestures were recorded, including an idle gesture. The set of gestures was chosen to cover extension and flexion of each individual finger and incorporate some typical gestures that might be used in a HCI context. The gestures were organized in three sets - a set of tapping gestures, a set of bending gestures and a set of gestures including more complex multi-finger movements. Figure 4 shows illustrations of the recorded gestures.

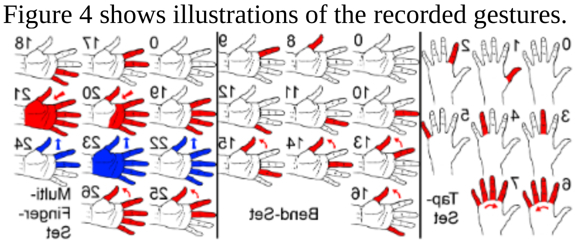


Fig 4. Illustrations of the performed gestures. The gestures were divided into three sets: the tap-set, the bend set and the multi-finger-set. The idle gesture (0) is contained in all three sets.

Using the electrode strip, electrical signal data of twenty-seven gestures from five healthy subjects were recorded using the 192-electrode grid (also referred to as channel) at a data-rate of 2048 Hz. The data was acquired for an average of 3 seconds resulting in 6144 samples from each channel or pair of electrodes. Data of all the five subjects is used for the study. For each gesture, there are 10 sets of data representing the number of times it was repeated.

C. Feature Extraction and Normalization

This step involves computing the features from the raw data and normalizing them so each feature has same range and thus same weightage. The characteristics of sEMG signals can be categorized by Time Domain (TD), Frequency Domain (FD), and Time-Frequency domain (TF). In this study, we chose to compute and use TD features because they are easy to implement and do not incur high computational cost.

The five time-domain features MAV, Standard Deviation, ZC, SSC (also referred to as TC) and WAV have been selected due to their ease of implementation, low computational cost, and robustness to represent the raw data that has been shown in several studies. Each of the five features are computed on the 6144 samples from each of the 168 valid channels resulting in a vector of length 168. The five features will have different range and scale of values. One of the critical steps when dealing with this type of diversified data is to perform normalization. Each feature of the channel data is normalized individually by linearly mapping them to a range of 0 and 1. From the matrix data of a gesture, the five normalized feature vectors each of length 168 are concatenated resulting in another vector of length 840. Further processing is performed on the normalized feature vector of length 840 obtained from the data collected for each gesture.

D. Classification

The shallow learning methods are shown [**2**] [**3**] to be among the best performing for the gesture recognition using EMG data. The use of deep learning to predict gestures is a very recent and active area of research. Four shallow learning methods namely LDA, Naïve Bayes, SVM, and KNN along with MLP of deep learning method are used in this study. Each of the methods classify the features to map them to the known gestures during training, and predict the gestures during testing step. The classifiers are applied to distinguish different sets of features.

E. Performance Evaluation

The output of the classifier methods is the prediction of the gestures. The predicted gestures are compared with the known ground truth gestures during the testing step to determine confusion matrix as the performance metric. Confusion matrix is used as the performance metric because of its robustness and simplicity to evaluate the performance of the classification methods.

Python programming with SciKit Learn library for shallow learning models and TensorFlow with Keras library for deep learning model was used for implementation along with computing the features, normalize them, and train and test the data sets, and finally obtain the confusion matrices and accuracies.

III. Theoretical Background

In this section, the mathematical formulas for the features are provided and the machine learning methods are briefly discussed.

Five TD features used are, MAV, SD, ZC, SSC or Turn Count, and WAV. The definition and their mathematical representations are as follows [**4**]:

a) Mean Absolute Value (mAV)

This feature is the mean absolute value of signal x in an analysis time window with N samples. x_k is the kth sample in this analysis window.

$$mAV = \frac{1}{N}\sum_{k=1}^{N} |x_k|$$

b) Zero Crossings (ZC)

ZC is the number of times signal x crosses zero within an analysis window; it is a simple measure associated with the frequency of the signal. To avoid signal crossing counts due to low-level noise, a threshold ε was included (ε = 0.015 V) [**6**]. The ZC count increased by one if

$$\{x_k > 0 \text{ and } x_{k+1} < 0\} \text{ or } \{x_k < 0 \text{ and } x_{k+1} > 0\})$$
$$\text{and } |x_k - x_{k+1}| \geq \varepsilon$$

c) Slope Sign Changes (SSC) or Turns

Slope sign change is related to signal frequency and is defined as the number of times that the slope of the EMG waveform changes sign within an analysis window. A count threshold ε was used to reduce noise-induced counts (ε = 0.015 V) [**6**]. The *slopeSign* count increased by one if

$$\{x_k > x_{k-1} \text{ and } x_k > x_{k+1}\} \text{ or } \{x_k < x_{k-1} \text{ and } x_k < x_{k+1}\}$$
$$\text{and } |x_k - x_{k+1}| \geq \varepsilon \text{ or } |x_k - x_{k-1}| \geq \varepsilon$$

d) Willison Amplitude (wAmp)

This feature is defined as the number of times that the change in EMG signal amplitude exceeds a threshold; it is an indicator of the firing of motor unit action potentials and is thus a surrogate metric for the level of muscle contraction [**5**]. A threshold between 50 and 100 mV has been reported in the literature [**5**]. In this study, the threshold ε was defined for each subject as the EMG signal value that had a 50% probability of occurrence as defined by a computed cumulative distribution function for each type of intended movement:

$$wamp = \sum_{k=1}^{N} f(|x_k - x_{k+1}|)$$

where $f(x) = \{1 \text{ if } x > \varepsilon; 0 \text{ otherwise}\}$.

e) Standard Deviation

This feature is given by the below formula;

$$\sigma = \sqrt{\frac{\sum_{k=1}^{N}(x_k - mAV)^2}{N}}$$

f) Classification

In this project four types of well-known shallow learning classifier algorithms have been used including KNN, SVM, Naive Bayes, and LDA. A simple deep learning technique the multi-layer-perceptron (MLP) is used.

The KNN method begins at the test point and increases its region until it includes K training samples and applies the majority vote of these samples to identify the test point. The research has shown that no optimal number of neighbors fit all kinds of datasets because each dataset has its requirements (Dougherty, 2013).

The SVM is a linear model used to implement non-linear classification boundaries. The support vector classification deduces a computationally efficient path of learning 'good' splitting hyperplanes in dimensional feature space, whereby 'good' hyperplanes could distinguish between new sample classes. In many real-life applications, there are non-separable cases of data when both classes are overlapping. In such a situation, it becomes impossible to split the data with linear separation. Therefore, the SVM maps the data by applying a nonlinear transformation by a suitable selection of basic functions into a higher-dimensional feature space, where the situation becomes linear. There are many kernel functions in the SVM algorithm such as the linear, Radial Basis Function (RBF), polynomial and sigmoid functions (Ali, 2013, Dougherty, 2013).

The feature distributions are modeled by kernel density estimation with Gaussian kernel function in Naive Bayes classifier. The independence assumption of the naive Bayes classifier is clearly violated in this case since multiple electrodes span across the same muscle and different muscles do not act independently. However, Naive Bayes classifiers are known to work well under such conditions [7].

The LDA is Statistical classifier where a new observation should be assigned to mutually exclusive categories. The objective of the LDA, like the SVM technique, is to find a hyperplane that can split the data points into different classes. This hyperplane can be obtained by finding a model which enlarges the distance between the mean of the classes and reduce the variance within the class under the assumption of normal data distribution. The key point of the successful system is how to choose the appropriate features to support the classifier.

IV. Results

Two major types of results of the machine learning methods are presented and compared. The accuracy and confusion matrix. The accuracy is computed as the ratio of number of correct predictions and the total number of predictions, measured in percentage (Ali, 2013).

The confusion matrix is used to understand how the selected classifier performed predicting each class. The confusion matrix also helps to understand the number of misclassifications for each class. In the confusion matrix, the rows indicate the true class and the columns indicate the predicted class. The diagonal cells indicate the number of correct classifications and the non-diagonal cells indicate the number of misclassifications for each gesture.

Each of the five methods are chosen along with the variables that contain the training data set and the training is performed. After training completes, the implementation reports accuracy of the training and confusion matrix along with the trained model that are noted and saved for further use. The training accuracies are tabulated in Table 1 and the confusion matrices are depicted in Figures 2, 3, 4 and 5.

The saved trained model for each of the machine learning methods is used to predict the gestures from the test data. The accuracy of the test results is tabulated in Table 3.

TABLE I. Testing Accuracy using FLOAT64

No. of Gestures	**Time Domain Features**				
	LDA	***Naive Bayes***	***SVM***	***K-NN***	***MLP***
22	95.76	90.10	90	91.62	88.89

The Confusion Matrices

Figure 1. LDA Confusion Matrix

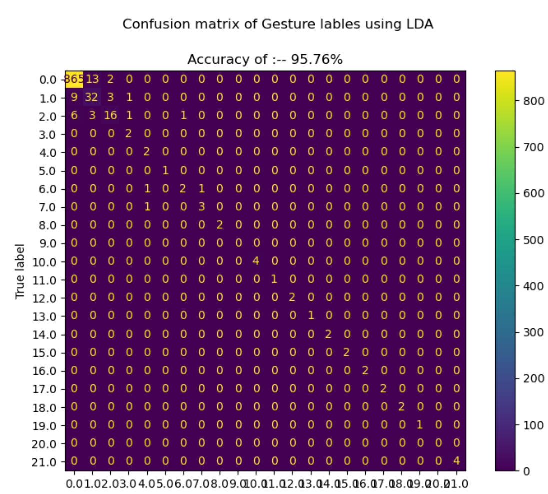


Figure 2. Naive Bayes Confusion Matrix

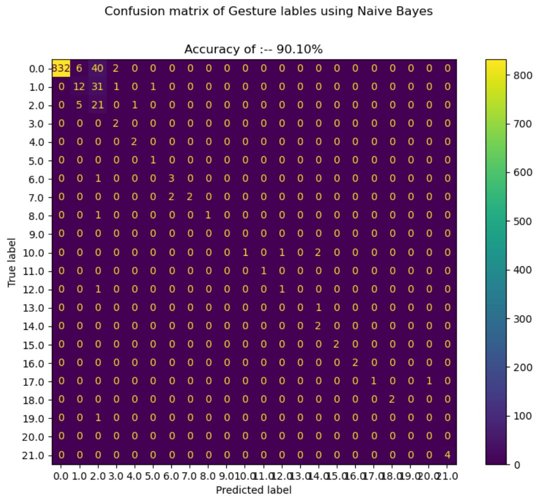


Figure 3. SVM Confusion Matrix

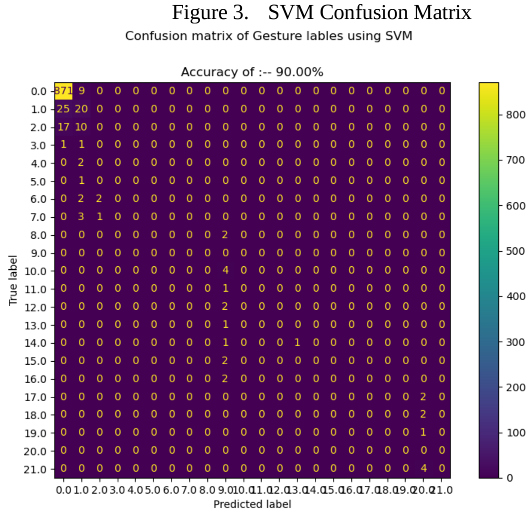


Figure 4. K-NN Confusion Matrix

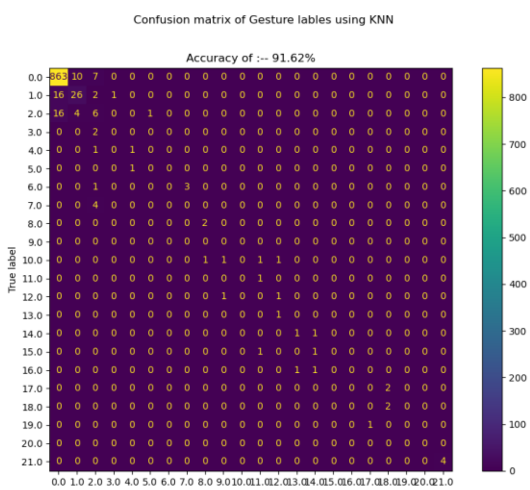
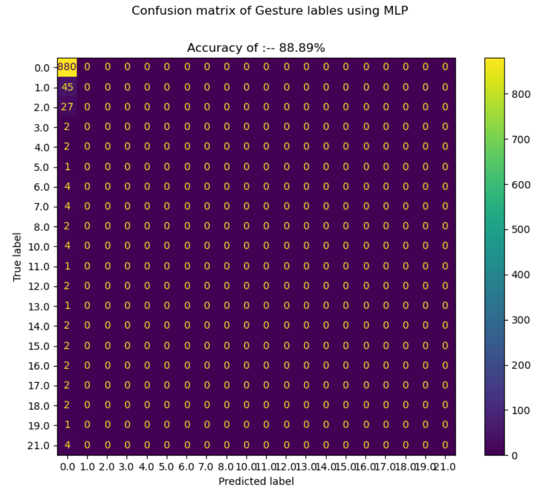


Figure 5. MLP Confusion Matrix



V. Conclusion

From the open source HDEMG dataset called CSL-HDEMG was used in our study. Five features were computed from each of the 168 channels and normalized to a range 0 to 1. The normalized vector for each gesture is used as input to the five classification algorithms to train them. The trained models were used to determine the accuracy of prediction on the test data that was left behind by splitting the original labeled data in the ratio 80 to 20. To evaluate the performance of the machine learning methods, the testing accuracy and confusion matrices were computed and noted.

I noticed that the original data is too large to read by the limited resources on my laptop when read in as a double precision floating point format also referred to as Float64. Only data from 22 gestures was able to be read and processed using Float64. Table I and Figures 1 to 5 depict the results using raw data being read in as Float64.

It was found that LDA provided the best accuracy of 95.76% during testing. It can be concluded that the best results were achieved with the LDA classifier, while the rest of the classifiers achieved acceptable results.

Directions for future work would include formatting the HDEMG data as an image and use of convolutional neural network (CNN) to train the data. It has been shown that the CNN determine very rich features and thus could be an improvement over the hand-picked features used in this study.

References

[1] Amma, Christoph; Krings, Thomas; Böer, Jonas; Schultz, Tanja (2015). *[ACM Press the 33rd Annual ACM Conference - Seoul, Republic of Korea (2015.04.18-2015.04.23)] Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15 - Advancing Muscle-Computer Interfaces with High-Density Electromyography, (), 929–938.* https://dl.acm.org/doi/abs/10.1145/2702123.2702501

[2] C R. N. Khushaba and S. Kodagoda, “Electromyogram (emg) feature reduction using mutual components analysis for multifunction prosthetic fingers control,” in Control Automation Robotics & Vision (ICARCV), 2012 12th International Conference on. IEEE, 2012, pp. 1534–1539

[3] Chowdhury RH, Reaz MBI, Ali MABM, Bakar AAA, Chellappan K, Chang TG. Surface Electromyography Signal Processing and Classification Techniques. *Sensors*. 2013; 13(9):12431-12466. https://doi.org/10.3390/s130912431

[4] Tkach, Dennis et al. “Study of stability of time-domain features for electromyographic pattern recognition.” *Journal of neuroengineering and rehabilitation* vol. 7 21. 21 May. 2010, doi:10.1186/1743-0003-7-21

[5] Zardoshti-Kermani M, Wheeler B, Badie K, Hashemi R. EMG feature evaulation for movement control of upper extrmity prostheses. IEEE Trans Rehabil Eng. 1995; 3:324–333. doi: 10.1109/86.481972 [CrossRef] [Google Scholar]

[6] Hudgins B, Parker P, Scott RN. A new strategy for multifunction myoelectric control. IEEE Trans Biomed Eng. 1993;40:82–94. doi: 10.1109/10.204774. [PubMed] [CrossRef] [Google Scholar]

[7] Zhang, H. The optimality of naive bayes. AA 1, 2 (2004), 3.

[8] CSL-HDEMG -- https://www.uni-bremen.de/en/csl/research/motion-recognition