Hand Gesture Detection based on EMG Signals using ML Techniques

Surbhi Garg[a] , Ayush Grover[a], Gurarpit Kaur[a], Sanya Devgon[a] and Jyoti Yadav[a]
[a]Department of Instrumentation and Control, Netaji Subhas University of Technology, Azad Hind Fauj Marg, Dwarka , Delhi, India.

ABSTRACT

Our work aims to develop a hand gestures detection model using EMG signals, with a vision to control prosthetic components and other human assisting manipulators and help aid people with limb disabilities. The raw EMG signals are used to obtain time and frequency domain features. A number of features are extracted in time and frequency domain. In order to enhance the performance of classifiers another dataset with windowing method applied on it is also created from which the same features are extracted. The extracted features are then scaled using Min-Max normalization method. Machine learning algorithms applied including Random Forest, K-Nearest Neighbours (KNN) and Decision tree are used for classification. The results demonstrate KNN when used with a feature set with windowing, as the best classifier (accuracy 0f 98.44%). Also, the results show that the overall feature set with windowing performs better than the one without.

1. INTRODUCTION

Electromyogram signals are the bio signals that evaluate the activity produced by the contraction of the skeletal muscles which is representative of the neuro-muscular activity within the human body controlled by the nervous system. They are suitable for hand movement classification problems since any movement has a particular signature in the produced EMG signal. EMG results are able to detect nerve and muscle damage or troubles associated with nerve to muscle signal transmission giving information about underneath problems like muscle disorders including but not limited to muscular dystrophy or polymyositis, disorders affecting nerve and muscle connection like myasthenia gravis, carpal tunnel syndrome and so on.

Surface electromyography (sEMG) signals are being employed in various examinations for the arrangement of hand motions and developments and effectively actualized in the position control of various prosthetic hands for amputees.

Many biomedical researchers now days are giving attention to the concept of myoelectric, due to its amplitude and motion relationship making EMG signal one of the major aspect in prosthesis control [1]. Hence, EMG-signal is a suitable option for classification of various hand movement since each movement produce a distinct signal [2]. Further, there has been much work done in the field of developing robust algorithms for EMG signal detection, subsequent conversion to kinematic variables and hence, control prosthetic devices and other human-assisting manipulators Presently, a number of human interfaces based on EMG have been introduced as a means for the handicapped and elderly people to control powered prosthetic limbs, wheelchairs, teleoperated robots [37]. Our motivation stems from this importance, application of EMG signals holds in the instrumentation domain already, along with the fields like using them to help people with limb disabilities/deficiency.

Feature extraction is used to gain important details from the signals [2][3]. It can be divided into Time, Frequency and Time-frequency domain.

Five EMG features for pattern recognition were introduced by Hudgins et al. [4] proposing that the features discriminate the EMG patterns better. Then, three new EMG features (Absolute value of the Summation of Square root (ASS), Mean value of the Square Root (MSR) and Absolute value of the Summation of the exp^{th} root of the data and its Mean (ASM)) for classification of movement of arms were proposed by Samuel et al. [5].He showed that the features performed better than other conventional features in EMG pattern recognition. Later on, Too et al. [6] introduced two new features (Enhanced MAV (EMAV) and Enhanced WL (EWL)) proposing that more information lies in the middle region of the EMG signal than considered earlier. In our work Time and Frequency domain features are considered.

Windowing technique is an important pre-processing as it increases the number of training samples which can help increase accuracy of classifiers. Apart from choosing the correct windowing, length of the window is also important so as to achieve better performance in parameters estimation. In the past, a number of window lengths were investigated by Smith.et al [7] to discriminate 7 motions which were wrist flexion, wrist extension, wrist pronation, wrist supination, hand open, hand close and a relaxed position using classifiers. Triwiyanto compared different window size to see its effect on the performance of estimation of elbow joint angle and to determine optimum window length [10]. Alam incorporated windowing along with feature extraction, feature reduction for classification of EMG signals using Linear Discriminant Analysis (LDA) classifier [20]. Recently, windowing, when applied on dataset overall gave better result with classifiers (jia.et.al) [8].

Previous works from many researchers showed that combination of normalization,feature reduction along with tuning of hyper parameters of the classifiers achieved high accuracies. Too [6] presented that KNN when used with 4 selected features achieved 96% accuracy. Al-Faiz [21] used KNN to recognize human arm movements based on EMG signals. A. Abdullah [22] used a combination of WPD (wavelet packet decomposition) and random forest for classification of surface EMG signal and achieved an accuracy of 92.1%. Gokgoz [23] compared different decision tree algorithms on EMG signal using DWT and evaluated that random forest performs best with an accuracy of 96.67%.

The aim of the paper is to investigate and compare the performance of two feature set extracted from publicly accessible EMG dataset, one in which features are extracted and the second in which the same features are extracted after applying windowing. Feature normalization and dimensionality reduction is used to remove redundancies from the dataset. For classification three classifiers have been used i.e. K-nearest neighbour (KNN), Random forest (RF) and Decision Tree (DT). Performance of both the feature sets have been evaluated using these classifiers and subsequently the results have been compared.

The paper is organized as follows: Methodology is described in section 2 including the windowing technique method and the features extracted followed by result and discussion in section 3 and conclusion in section 4.

2. METHODOLOGY

Fig 1. Represents the steps followed in the experiment. First, data consisting of EMG signals is acquired from a publicly accessible UCI repository. Two feature set are created. In the first set, features are extracted from the signals. A total of 25 features are extracted including 18-time domain and 7-frequency domains features. In the second set, windowing is applied on the EMG signals and then, features are extracted from the newly created samples. Normalization and dimensionality reduction (PCA) are performed on both the sets and are then fed into the classifier (Random Forest, KNN and Decision Tree) to classify the six hand movements and also to compare the accuracy along with other performance parameters of both the feature sets.

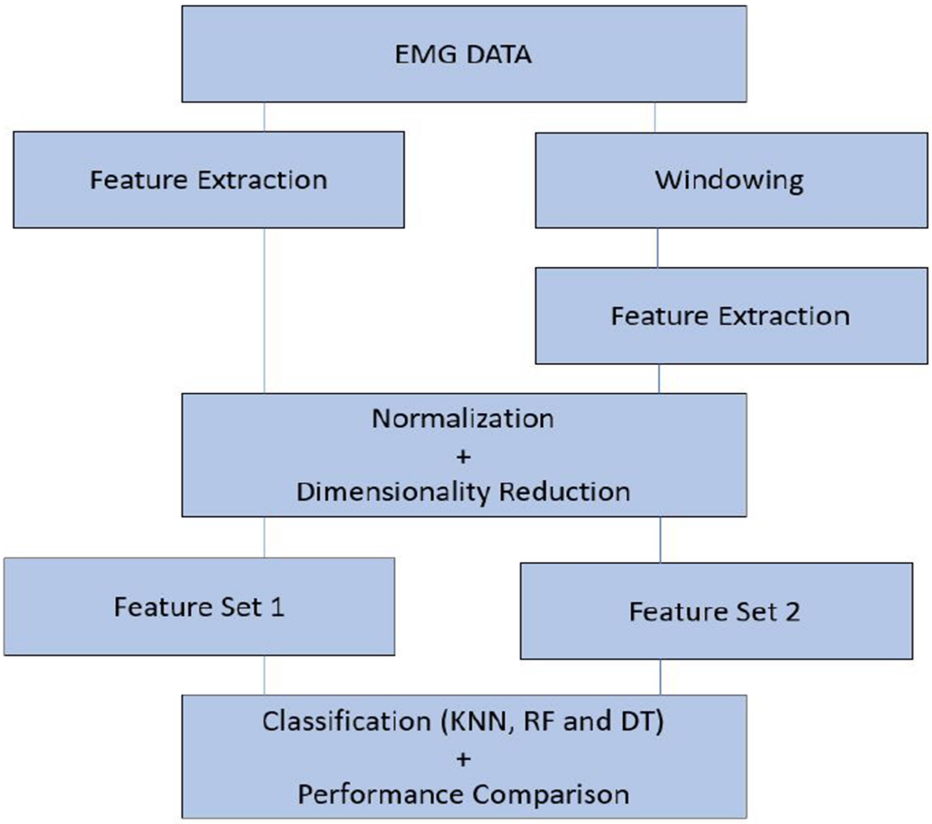

Fig 1. Methodology block diagram

2.1. EMG DATA

The EMG data is taken from UCI repository under the name "sEMG for Basic Hand movements" [9]. The dataset consists of two databases but only the first database is considered in the work. The database comprises segmented EMG signals corresponding to six types of hand movements recorded from five subjects of which two are males and three are females having their ages in the range of 20 to 22 years. The data was recorded using two channel electrodes at a sampling rate of 500Hz. Each subject conducted each movement for 6 seconds for 30 times each. The size of each sample is 3000.The hand movements are shown in fig 2 and data description in table 1.

Table 1. Data set description

SAMPLE RATE	SAMPLING TIME	NO. OF CHANNELS	SAMPLE'S SIZE	NO. OF SUBJECTS	NO. OF SAMPLES
500 Hz	6 seconds	2 channels	3000 x 2	5 subjects	900 (5 x 30 x 6)

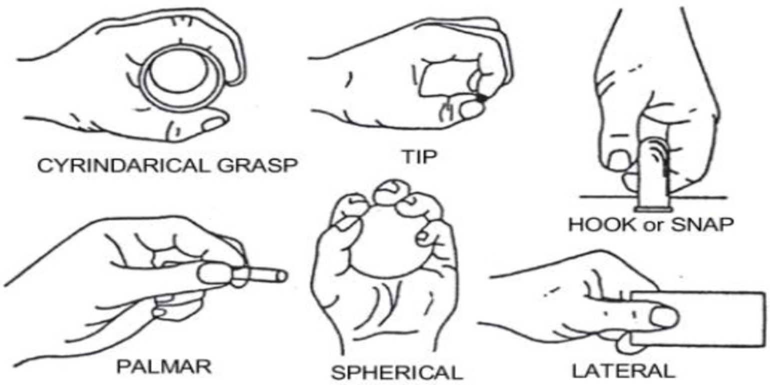

Fig 2. Hand movements Performed [9]

2.2. WINDOWING TECHNIQUE

Windowing method is a pre-processing technique that can help increase performance of classifiers. It's application increases the sample size (900 to 171900 in our work).As the amount of data increases, the features extracted from it have lower statistical variance which increases classification accuracy[39]. This method is divided into adjacent and overlap windowing methods [10]. In adjacent windowing the signal is segregated into windows with adjacent windows connected to each other at the end. In overlap window method part of the following window is overlapped with the previous window. Our work employs overlap window method because of its better performance on accuracy of classifiers as compared to adjacent windowing method[39]. The size of window is associated with data length of signal to size of training samples as[20]:

$$number\ of\ training\ samples = 1 + \frac{data\ length - wind\quad size}{window\ seperation} \tag{1}$$

The window size considered in our work is 150 samples and window separation to be 15 samples [12]. Hence windowing increases sample size by a factor of 191 (from 900 to 171900).

2.3. FEATURE EXTRACTION

Feature extraction is a technique that is used to obtain useful details from raw EMG signals with which classifiers can be trained. A total of 25 features are extracted from both the dataset (with and without windowing). 18 from time domain and 7 from frequency domain are extracted in our work

2.3.1. Time domain features

Time domain features are easily applied because they don't require any modifications. But limitation of these features is that they assume the signal to be at rest even if they are not (which is the case of EMG signals) [11]. They are calculated on the basis of amplitude of input signal. 18-time domain features are extracted in this work are as follows:

2.3.1.1. Integrated EMG (IEMG)

IEMG feature is defined as the total sum of mod values of amplitude of EMG signal [4], and is formulated as:

$$IEMG = \sum_{k=1}^{M} |x_k| \tag{2}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.2. Mean absolute value (MAV)

MAV feature is an average of absolute value of the EMG signal amplitude in a segment [4], and is formulated as:

$$MAV = \frac{1}{M}\sum_{k=1}^{M} |x_k| \tag{3}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.3. Simple square integral (SSI)

SSI is the total sum of square value of amplitude of an EMG signal. Generally, this is denoted as the energy index [4] and is defined as:

$$SSI = \sum_{k=1}^{M} x_k^2 \tag{4}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.4. Variance of EMG (VAR)

Variance is the mean of square values of the deviation of that variable power of EMG signal can be computed by this [4],[5] and is defined as:

$$VAR = \frac{1}{M-1}\sum_{k=1}^{M} x_k^2 \tag{5}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.5. Zero crossing (ZC)

Zero crossing feature computes signal's frequency information and is defined as the number of times zero amplitude level is crossed by the values of amplitude of the signal. Threshold is used to avoid background noise [4],[5]. ZC is expressed as:

$$ZC = \left(\sum_{k=1}^{M-1} [sgn(x_k \times x_{k+1}) \cap |x_k - x_{k+1}| \geq threshold] \right) \quad (6)$$

$$sgn(x) = \begin{cases} 1, x \geq threshold \\ 0 \quad ,otherwise \end{cases}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.6. Wilson amplitude (WA)

Wilson amplitude is the number of times due to difference between the amplitude of signal concerning two adjacent segments exceeding a predefined value [4], and is expressed as:

$$WA = \sum_{i=1}^{M-1} [f(|x_k - x_{k-1}|)] \quad (7)$$

$$f(x) = \begin{cases} 1, x \geq threshold \\ 0 \quad ,otherwise \end{cases}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.7. Root mean square (RMS)

RMS is one of the most popular feature and is useful in providing muscle information [4],[5].It is fast and computationally efficient and is evaluated as:

$$RMS = \sqrt{\frac{1}{M} \sum_{k=1}^{M} x_k^2} \quad (8)$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.8. Waveform length (WL)

Waveform length is defined as the cumulative length of the EMG waveform over the segment [4] [5]. It can be formulated as:

$$WL = \sum_{k=1}^{M-1} |x_{k+1} - x_k| \quad (9)$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.9. Amplitude absolute change (AAC)

Amplitude absolute change is very similar to waveform length except it is the average of wavelength [4] and can be formulated as:

$$AAC = \frac{1}{M} \sum_{k=1}^{M-1} |x_{k+1} - x_k| \quad (10)$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.10. Log detector (LD)

Log detector gives estimation of contraction of muscle force [4]. It is evaluated as:

$$LD = e^{\frac{1}{M}\sum_{k=1}^{M} log(|x_k|)} \quad (11)$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.11. Slope sign change (SSC)

Slope sign change represents the number of times slope of signal changes sign [4] and is evaluated as:

$$SSC = \sum_{k=2}^{M-1} [f[(x_k - x_{k-1}) \times (x_k - x_{k+1})]]$$

$$f(x) = \begin{cases} 1, x \geq threshold \\ 0 \quad ,otherwise \end{cases} \quad (12)$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.12. Difference absolute standard deviation value (DASDV)

DASDV represents standard deviation value of the wavelength [4]:

$$DASDV = \sqrt{\frac{1}{M-1} \sum_{k=1}^{M-1} (x_{k+1} - x_k)^2} \quad (13)$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.13. *Modified mean absolute value 1(MAV1)*

Modified mean absolute value 1 (MAV1) is an enhancement of MAV feature. The weighted window function w_k is used to enhance the robustness of the MAV feature [4]. It can be formulated as:

$$MAV1 = \frac{1}{M}\sum_{k=1}^{M} w_k|x_k| \tag{14}$$

$$w_k = \begin{cases} 1\,, & 0.25M \leq k \leq 0.75M \\ 0.5\,, & otherwise \end{cases}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal.

2.3.1.14. *Modified mean absolute value 2(MAV2)*

Modified mean absolute value 2 (MAV2) like MAV1 feature is also an enhancement of MAV feature. However unlike in MAV1, w_k in MAV2 is a continuous function [4].It can be formulated as:

$$MAV2 = \frac{1}{M}\sum_{k=1}^{M} w_k|x_k| \tag{15}$$

$$w_k = \begin{cases} 1\,, & 0.25M \leq k \leq 0.75M \\ \frac{4k}{M}, & k < 0.25M \\ \frac{4(k-M)}{M}, & otherwise \end{cases}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal..

2.3.1.15. *Maximum fractal length (MFL)*

Maximum fractal length (MFL) is used for measuring the activation of low-level muscle contraction [6]. It is evaluated as:

$$MFL = log_{10}\left(\sqrt{\sum_{k=1}^{M-1}(x_{k+1} - x_k)^2}\ \right) \tag{16}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal..

2.3.1.16. *Myopulse percentage rate (MYOP)*

Myopulse percentage rate (MYOP) is the average of myopulse output which equates to one when the signal's absolute value is greater than a predefined value [4]. It is expressed as:

$$MYOP = \frac{1}{M}\sum_{k=1}^{M} f(x_k) \tag{17}$$

$$f(x) = \begin{cases} 1, x \geq threshold \\ 0 \quad , otherwise \end{cases}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal..

2.3.1.17. Enhanced mean absolute value (EMAV)

Enhanced MAV (EMAV) is an extension to MAV. It is expressed as [5]:

$$EMAV = \frac{1}{M}\sum_{k=1}^{M} |(x_k)^t| \tag{18}$$

$$t = \begin{cases} 0.75, & 0.2M \leq k \leq 0.8M \\ 0.5, & otherwise \end{cases}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal..

2.3.1.18. Enhanced waveform length (EWL)

Enhanced WL (EWL) is an extension to WL. It is expressed as [5]:

$$EWL = \frac{1}{M}\sum_{k=2}^{M} |(x_k - x_{k-1})^t| \tag{19}$$

$$t = \begin{cases} 0.75, & 0.2M \leq k \leq 0.8M \\ 0.5, & otherwise \end{cases}$$

Here,x_k denotes k^{th} segment of signal and *M* denotes length of the signal..

2.3.2. Frequency Domain Features

Frequency domain features are evaluated on their power spectral density (PSD) [4]. However, more computation time is required than time domain features. We have used 7 frequency domain features in our project.

2.3.2.1. Mean frequency (MNF)

Mean frequency (MNF) is defined as the mean frequency computed as the calculation as summation of product of the frequency and power spectrum of signal, divided by summation of the spectrum intensity [4] and is expressed as:

$$MNF = \sum_{k=1}^{L} f_k P_k \,/ \sum_{k=1}^{L} P_k \tag{20}$$

Here,$f_k$is the spectrum's frequency, P_k is the power spectrum and *L* is the length of frequencies

2.3.2.2. Total power (TTP)

Total power is accumulation of power spectrum of signal [4] and is expressed as:

$$TTP = \sum_{k=1}^{L} P_k \tag{21}$$

Here,$P_k$is the power spectrum and L is the length of frequencies

2.3.2.3. Median frequency (MDF)

Median frequency is defined as the frequency which divides the spectrum into two regions of same amplitude. [4]. and is evaluated as:

$$MDF = \frac{1}{2}\sum_{k=1}^{L} P_k \tag{22}$$

Here,$P_k$is the power spectrum and L is the length of frequencies

2.3.2.4. Mean power (MNP)

Mean power (MNP) is defined as mean of power spectrum of signal [4] and is expressed as:

$$MNP = \sum_{k=1}^{L} P_k /L \tag{23}$$

Here,$P_k$is the power spectrum and L is the length of frequencies

2.3.2.5. Frequency ration (FR)

Frequency ratio (FR) is calculated as the low frequency components of signal divided by the high frequency components [4] and is expressed as:

$$FR = \sum_{k=LLC}^{HLC} P_k / \sum_{k=LHC}^{HHC} P_k \tag{24}$$

Here, higher and lower cut-off frequency of the low frequency region are denoted by HLC and LLC and the higher and lower-cut off frequency of the high frequency region are denoted by HHC and LHC and P_k is power spectrum

2.3.2.6. Modified median frequency (MMDF)

Modified Median Frequency (MMDF) is the frequency which divides the spectrum into two regions with same amplitude [7] and is evaluated as:

$$MDF = \frac{1}{2}\sum_{k=1}^{L} A_k \tag{25}$$

Here,$A_k$is amplitude spectrum.

2.3.2.7. Modified mean frequency (MMNF)

Modified mean frequency (MMNF) is computed as the summation of the product of the frequency and amplitude spectrum with which summation of spectrum intensity is divided [7]. It is evaluated as:

$$MNF = \sum_{k=1}^{L} f_k A_k \, / \sum_{k=1}^{L} A_k \tag{26}$$

Here,$A_k$is amplitude spectrum and $f_k$is the spectrum's frequency

2.4. FEATURE NORMALIZATION

Feature Normalization is a method that rescales independent features present in the data to a predefined fixed range. It takes care of highly fluctuating magnitudes values [25],[26]. If not performed, then the algorithm regardless of the unit's values, considers bigger values to be higher and smaller values to lower values.

Most of the time, the dataset comprises highly varying features in magnitudes, units and range. But many of the algorithms use Euclidean distance to calculate distance between two data points (eg.KNN) in their computations, this is a problem. To eliminate this problem; all features need to be brought to the same level of magnitude which can be achieved by normalization.

In Min-max normalization we rescale the range of features to scale in the span of [0, 1] or [−1, 1]. Data's nature determines target range selection [25],[26],[27]. The general formula for a min-max of [0, 1] is given as:

$$x_{new} = \frac{x_i - min(x)}{max(x) - min(x)} \tag{27}$$

Here,x_i is the i^{th}data point, max is maximum of data x and min is minimum of data x

2.5. DIMENSIONALITY REDUCTION

Some of the features extracted might be correlated to each other therefore providing redundant information. Also, certain features might be irrelevant which can hinder capability of classifier [12]. Hence, dimensionality reduction technique is generally used. PCA was employed feature reduction as it reduces dimensionality of dataset containing interrelated variables [20] Güler and Koçer [24] used PCA to reduce the number of FFT coefficients which was used for EMG classification. Hence, in this work, we have used PCA.

2.5.1. Principal Component Analysis (PCA)

PCA is a method which allows visualization of large datasets by means of smaller datasets without much information loss. It is done to maximize variance due to formation of new uncorrelated variables. With the finding of such new variables, the principal components, minimizes into solving an eigenvalue/eigenvector problem, and the new variables are defined by the current dataset and not from the earlier, making PCA an "adaptive data analysis technique"[15].

In PCA covariance matrix is used to reduce dimension of the data. An EMG dataset consisting of a $k \times m$ matrix denoted by Y where k denotes the number of samples and m denotes the variables is first centred (i.e. column-wise average of Y is subtracted from the elements of the corresponding columns of Y) and covariance matrix Y denoted by COV ($m \times m$ dimensions) is calculated by [11]:

$$COV = Y^T Y/(k-1) \tag{28}$$

Here $(.)^T$ denotes transpose.

Eigenvalue decomposition of the data's covariance matrix diagonalizes COV to yield:

$$HVH^T \tag{29}$$

Here, columns of H and diagonal elements of V are eigenvectors and eigenvalues of covariance matrix (COV).

By selecting eigenvectors corresponding to a < features with largest eigenvalue, the data matrix gets converted into new space YH'(where $H'$is $k \times a$ matrix) hence, achieving dimensionality reduction.

The parameters considered for applying PCA using the sklearn library include:

Table 2: Parameters for PCA

N_components = None	Svd_solver = 'auto'

2.6. *CLASSIFICATION*

Classification techniques use supervised learning to train on the dataset and correctly identify target variables. The problem in consideration is a multiclass classification problem wherein the machine learning model is required to classify more than two target variables. In light of previous work done we have worked on two of the most renowned classification algorithms namely KNN (K-Nearest Neighbours), Decision trees and Random Forest algorithm. After applying the above-mentioned models, we conducted comparative analysis of their individual performances in order to find the most optimum model for our dataset. The implementations of these models were performed on scikit learn.

2.6.1. Decision Tree

Decision trees create subsets by splitting the parent nodes in order to reach the final prediction value [38]. As mentioned above we have implemented the models on scikit learn which uses an enhanced version of the classification and regression tree algorithm to create decision trees. This algorithm uses a binary tree and for the purpose of classification Gini impurity or Entropy can be used [28].

$$Gini\ Impurity = \sum_{k=1}^{c} f_k(1 - f_k) \tag{30}$$

f_k= frequency
c = the count of unique labels
Entropy is used to estimate how much information the node is providing [28],[39].

$$Entropy = \sum_{k=1}^{c} -f_k log(f_k) \tag{31}$$

Information Gain can be explained as the decrease in entropy after the creating a subset of the dataset to a parameter for splitting. [28]

$$Gain\ (L, Y) = Entropy\ (L) - Entropy\ (L, Y) \tag{32}$$

Here the variable L is the target or outcome variable, Y is the feature on the basis of which splitting occurs and Entropy (L, Y) is the calculated entropy after the splitting has occurred.

2.6.2. Random Forest Algorithm

The random forest algorithm constructs several single decision trees and pools the prediction results from these trees which is then subjected to majority voting[29]. The prediction with maximum votes is used to display the final decision. The random forest algorithm works better than individual decision trees in this case because a group of uncorrelated tree models performs better than individual tree models . Bagging of trees helps in randomly sampling the dataset and therefore resulting in varied tree segments. This helps the model to lower the correlation among trees and ensure variation. This diversification results in building robustness for the model and reducing biases in prediction results [30].

The Random Forest algorithm uses feature importance as a metric to keep in check the stability of the algorithm. Feature importance can be termed as the decrease in node impurity which is then weighted with the probability of reaching that particular node. We have discussed the implementation of this model as used in scikit learn.
First, the importance of say node ‘j’ is calculated considering only two child nodes as follows:

$$nj_i = w_i c_i - W_{left(i)} c_{left(i)} - W_{right(i)} c_{right(i)} \tag{33}$$

nj_i= Importance of node i
w_i= Weighted samples that reach the node i
c_i= The impurity at node i

The other left and right variables suggest the respective left and right split of the child node. Next, we discuss how the importance of each feature is calculated in decision trees. After the calculation of feature importance, a normalized value of this feature importance is calculated. The normalized value lies in the range (0,1). This value is calculated by dividing the feature importance by the sum of all the feature importance values.

$$f_j = \frac{\sum_i nj_i}{\sum_k nj_k} \tag{34}$$

f_j= Feature importance
nj_i= Importance of node i

$$Normalized\, f_j = \frac{f_j}{\sum_i f_j} \tag{35}$$

In the case of random forest another level of feature importance is calculated which is nothing but the average of all the normalized feature importance values.

$$Rf_j = \frac{\sum_i Normalizedfeatureimportancevalues}{T} \tag{36}$$

Where, T = Total number of trees

2.6.3. K-Nearest Neighbours

KNN algorithm can be applied to classification problems as well as regression problems [31]. Here, KNN is used as a classifier. The foundational principle of the KNN algorithm is that similar data points lie close together. This assumption is based on the basic concept of calculation of distances between two data points [32]. There are different distance metrics used in KNN like the cosinedistance, Minkowski and Euclidean distances [31],[33]. Minwoski distances is used for obtaining our model results. It is a group of distances that include 3 distances depending on value of power parameter p.It is formulated as:

$$d(y, y') = \left(\sum_{j=1}^{n} |y_j - y_j'|^{1/p} \right)^p \tag{37}$$

when p=1,the metric is equivalent to manhattan distance .when p=2,it is equivalent to Euclidian distance and for p=∞,it equals to chebyshev distance[33].

Y is given in the input class x is the class label. With k, k-nearest neighbours finds k closest observations to test point y and evaluates probability pr of the point belonging to class j using the formula:

$$pr(x = j|Y = y) = \frac{1}{k} \sum_{i \in A} I(x^i = j) \tag{38}$$

Hyper tuning techniques were used to obtain optimal parameters for the above-mentioned models.

2.7. *GRID SEARCH METHOD*

We based the selection of these parameters on our observations of the published research on the same subject (For Optimizing Hyperparameters). The Grid Search method was used to hyper tune the model and obtain the most optimal hyperparameter values for our respective models. The method helped us enhance the robustness of our models by providing more accurate prediction values [34]. Often referred as an exhaustive search method, Grid Search sets a prediction value at first for each parameter. Then, it provides score for each parameter value to consider and which to choose [35]. The method is suitable only when the required maximum is known to be within definite area defined by upper bound and lower bound of each independent variable [34],[35].

Consider Ω^* to be a space of nuisance parameters $A=(A_1,A_2,A_3.....A_k)$ over which t-value have to be maximized. We can setup grid search by defining upper and lower bound vectors $x=(x_1,x_2....x_k)$ and $y=(y_1,y_2....y_k)$, m equally spaced points are taken in the interval $[x_j,y_j]$ creating m^k grid points to check. After computation of each pair of points, maximum is selected [36].

In our work, GridSearchCV library from sklearn is used for parameter tuning. It iterates through predefined hyperparameters to fit our classifier model and find the best parameters from given hyper parameters for us to select from. Here "CV" in GridSeachCV indicate the number of cross validation each hyper parameter selected has to go through. We have set CV=10 for our work.

2.8. *Other Performance Parameters*

Apart from accuracy, other performance parameters also considered to test the effectiveness of classifiers, precision, recall (sensitivity), specificity and f1-score are taken into consideration.

Precision is the proportion of positive observations predicted correctly to the total predicted positive observations [16]. Recall (sensitivity)is the proportion of the positive observations predicted correctly to the all observations in the actual class whereas specificity is the actual negative cases that have gotten predicted as negative by our model [17]. F1-score is the harmonic mean between precision and sensitivity [18].

The following are the formulas of performance parameters:

Table 3: Formulae of Precision, Recall, Specificity and F1-Score

PRECISION	RECALL(SENSITIVITY)	SPECIFICITY	F1-SCORE
TP/(TP+FP)	TP/(TP+FN)	TN/(TN+FP)	2TP/(2TP+FP+FN)

Here, TP is correctly predictive positive values, FP is falsely predictive positive values, TN is correctly predictive negative values and FN is falsely predictive negative values

2.9. ROC Curve

A receiver operating characteristic curve is a graph that indicates classifier's performance at all classification thresholds. To produce a ROC plot, threshold binarizes the classifier output for each class. The threshold iterates from 0 to 1 in n steps, and at every iteration a point is calculated with respect to two parameters i.e. “True positive rate (1-specificity)”and “False Positive Rate (sensitivity)” [19].

AUC is the area under the ROC curve. AUC value is closer to 1 indicates that the classifier's accuracy is high [19].

3. RESULTS AND DISCUSSION

Various time based and frequency-based features are extracted from signals in the considered dataset. A total of 50 features(25 feature × 2 channels) corresponding to each movement from each subject are obtained. Afterward, KNN, RF and DT classifiers are trained with the features obtained for the classification process. In our work, the Shuffle Split cross-validation (n_splits=10 and test_size=0.2) method is applied. In Split shuffle split manner, the data is worked upon iteratively. Finally, mean accuracy of all iterations is observed. For our project we have used MATLAB version R2018a for applying windowing technique and for feature extraction from the data set and for classification python 3.5 was used including pandas, numpy, Matplotlib, Seaborn and sci-kit learn libraries

Before feeding the classifiers with the feature sets,their hyper paramertes were tuned using grid search method. The following were the hyperparameters that were selected after applying grid search:

Table 4: Parameters of Classifiers

KNN	**RF**	**DT**
N_neighbours = 1	Max_depth = 61	Max_depth = 20
P = 1	Criterion = 'gini'	Criterion = 'gini'
Metric = Minwoski	Bootstrap = True	Splitter = 'best'

Table 4 shows all the parameters that were tuned .For knn,”N_neighbours=1” indicated that the algorithm will look for 1 nearest neighbour and classify the test data point using it.P is the power parameter of minwoski metric and “P=1”indicates that the minwoski metric converts to manhattan metric.In RF and DT “Max_depth” indicates maximum depth of the tree ,”criterion”defines the quality of split .Here,“gini” is short for gini impurity.

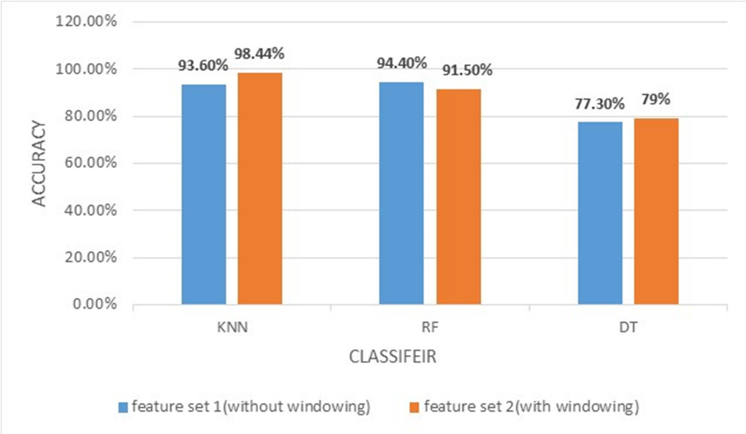

Fig 3: Accuracy of Classifiers on Both Feature Set

Table 5: Table of Accuracy of Classifiers

CLASSIFIER	**ACCURACY OF FEATURE SET 1 (SET WITHOUT WINDOWING)**	**ACCURACY OF FEATURE SET 2 (SET WITH WINDOWING)**
KNN	93.6%	98.44%
RF	94.4%	91.5%
DT	77.3%	79%

In the first part of the experiment, the performance considering all extracted features is calculated without applying any filters to the time series EMG data (feature set 1). The extracted features from EMG data are subjected to KNN, RF and DT classifiers and the performance of each classifier is measured. It is recorded in Table 5. From the results it can be inferred that RF classifier achieved much better results than its contemporaries we considered.

In the second part of the experiment, the dataset was further subjected to filtering using the windowing method. We decided to explore windowing of the dataset so as to be able to extract more information from the current EMG dataset. After the filtering, the same features were extracted (feature set 2) and were subjected to KNN, RF and DT classifiers. From Table 5 results, it is evident that KNN algorithm clearly and significantly outranks the other classifiers with 98.44% accuracy. The final results clearly indicate the superiority of KNN classifier on a dataset filtered with windowing method. It, by all means, is the best result that we have been able to achieve in the process of detection of hand gestures using EMG signals.

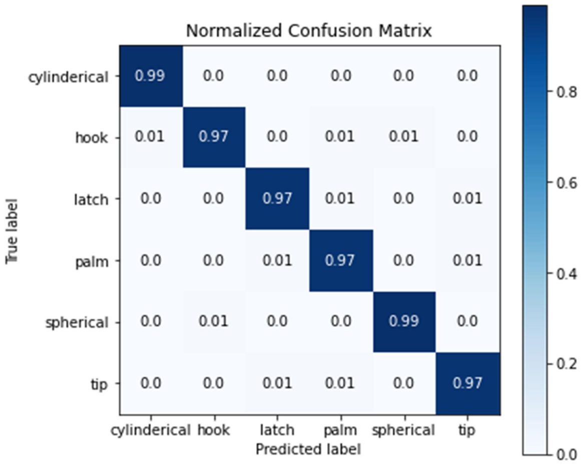

Fig 4. Normalized Confusion Matrix of Feature Set 2 (Dataset with Windowing) With KNN

From Fig 4, it can be seen that the spherical and cylindrical gesture is clearly recognized (both 99%) followed by other movements (all 97%).

Table 6: Performance Measures for KNN classifier

HAND GESTURE	KNN(Feature set 1)				KNN(Feature set 2)			
	Precision	**Recall**	**Specificity**	**F1-score**	**Precision**	**Recall**	**Specificity**	**F1-score**
LATCH	0.92	0.94	0.96	0.93	0.97	0.97	0.99	0.97
PALM	1.0	0.91	0.99	0.95	0.96	0.97	0.99	0.97
TIP	0.86	0.91	0.98	0.88	0.97	0.97	0.99	0.97
CYLINDERICAL	0.95	1.0	0.99	0.98	0.98	0.99	1.0	0.98
HOOK	0.96	0.93	1.0	0.95	0.99	0.98	1.0	0.98
SPHERICAL	0.97	0.97	0.99	0.97	0.99	0.98	1.0	0.98

Table 7: Performance Measures for Random Forest classifier

HAND GESTURE	RF(Feature set 1)				RF(Feature set 2)			
	Precision	**Recall**	**Specificity**	**F1-score**	**Precision**	**Recall**	**Specificity**	**F1-score**
LATCH	0.92	0.92	0.99	0.92	0.89	0.91	0.98	0.90
PALM	0.94	0.94	0.99	0.94	0.90	0.89	0.98	0.90
TIP	0.97	1.0	0.99	0.98	0.89	0.89	0.98	0.89
CYLINDERICAL	0.93	0.97	0.99	0.96	0.92	0.93	0.98	0.93
HOOK	0.97	0.94	0.99	0.95	0.93	0.94	0.99	0.93
SPHERICAL	0.96	0.92	0.99	0.94	0.97	0.93	0.99	0.95

Table 8: Performance Measures for Decision Tree classifier

HAND GESTURE	**DT(Feature set 1)**				**DT(Feature set 2)**			
	Precision	**Recall**	**Specificity**	**F1-score**	**Precision**	**Recall**	**Specificity**	**F1-score**
LATCH	0.89	0.78	0.92	0.83	0.74	0.75	0.95	0.75
PALM	0.72	0.74	0.97	0.73	0.75	0.75	0.95	0.75
TIP	0.79	0.79	0.92	0.79	0.73	0.74	0.95	0.73
CYLINDERICAL	0.68	0.77	0.96	0.72	0.82	0.8	0.97	0.81
HOOK	0.68	0.7	0.97	0.69	0.81	0.81	0.96	0.81
SPHERICAL	0.84	0.81	0.98	0.83	0.86	0.86	0.97	0.86

Table 9: Macro average performance Measures for both feature sets

FEATURE SET	**KNN**				**RF**				**DT**			
	Precision	**Recall**	**Specificity**	**F1-score**	**Precision**	**Recall**	**Specificity**	**F1-score**	**Precision**	**Recall**	**Specificity**	**F1-score**
Feature set 1	0.94	0.94	0.986	0.94	0.95	0.95	0.99	0.95	0.77	0.77	0.95	0.76
Feature set 2	**0.98**	**0.98**	**0.995**	**0.98**	0.92	0.92	0.98	0.93	0.79	0.79	0.96	0.79

Table 6,7 and 8 depict values for precision, recall and the f1 score for the three classifiers applied on both feature sets .From the parameter value comparison between the three tables ,it can be seen that KNN when applied on feature set 2 is able to identify different movements more accurately than other classifiers.Also,table 9 which contains the macro average of all performance measures indicates that the KNN performance with feature set 2 is best across all performance parameters considered other than accuracy.
Hence, it is evident that the process of windowing adds a significant value to the performance of our overall model, which was also evident from the model accuracy results.

To further deduce, it can be observed that KNN is the best performing model on our dataset when these other performance parameters are taken under consideration.

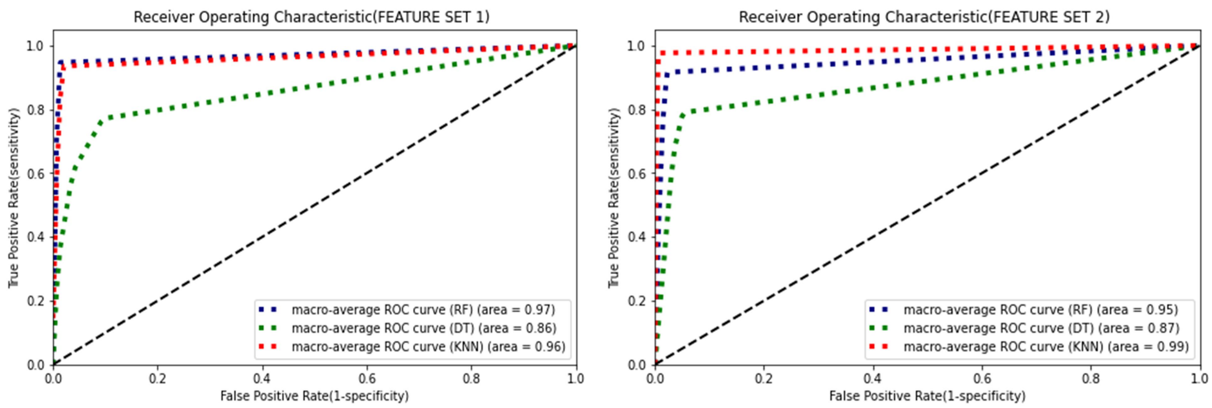

Fig 5: ROC Curve of all 3 classifiers for Feature Set 1(l) and Feature Set 2(r)

Table 8: Macro Average AUC for both Feature Sets

CLASSIFIER	AUC (FEATURE SET 1)	AUC (FEATURE SET 2)
KNN	0.96	0.99
RF	0.97	0.95
DT	0.86	0.87

The following image (FIG 5) displays ROC curves (macro average) and corresponding area of each classifier for both Features sets 1 & 2. It can be observed that the results are consistent with other parameter results. Considering the dataset without windowing, the area under the curve (AUC) is higher for RF compared to other models .However, as windowing is applied on our dataset there is a visible improvement in the performance of the KNN model. It outperforms the other models and at the same time, also has an AUC very close to 1(fig 5 and table 8).

Also, on further comparison with the other models used KNN has shown significantly enhanced prediction values as evaluated by the performance metrics in comparison to Random forest and Decision tree. The selective set of feature selection and optimized hyper parameter values have worked better for the KNN model in this case and hence the conclusion is skewed towards KNN performing better with enriched gesture detection prediction

3.1. COMPARISON WITH PAST LITERATURE

Based on the same dataset different experiments were conducted in the past by different researchers. [6] Used the same dataset but discrete wavelet transform (DWT) is applied on the dataset followed by feature extraction.14 features were extracted in total. The best accuracy of 97.57% was obtained considering 4 features (EWL + EMAV + MAV + WL) using LDA.
[40] Used two dataset among which one was the UCI hand gesture dataset (similar to the one used in this work) and another personally acquired. Waveform length Non-uniform filter bank were used to extract features along with PCA and LDA for dimensionality reduction. The highest accuracy achieved for UCI dataset was 84.66% using Euclidian distance (ED).
[41] Used singular value decomposition (SVD) to compute singular values and principal components of the sub-frame matrix obtained by dividing the EMG signal into overlapping sub-frame. The singular values along with along with first 5 principal components are used as features. The highest accuracy obtained is 86.71% with KNN classifier.

In comparison to the previous work described above, our work uses overlap windowing method after which 25 features are extracted. Normalization and PCA follows next which yields the best accuracy of 98.44% with KNN (considering all 25 features). Apart from accuracy other performance parameters are taken into consideration which indicate that KNN used with feature set 2(feature set with windowing) gives the best result

4. CONCLUSION

In our work, different techniques of machine learning were used to detect six hand gesture movements and were compared on standard performance metrics - accuracy, F1 scores and the AUC (area under ROC curve). Windowing as a processing tool was used on the feature set to explore the possibility of improved parameter performance. From the results, it can be observed that there has been an improvement of accuracy in predicting the hand gestures using the EMG signals which is imperative for proper control of prosthetic.

Random Forest Classification, Decision Tree algorithm, K – Nearest Neighbours were carried out on the dataset from EMG signals corresponding to six hand movements. Our results which remained consistent across parameter considerations, showed an improvement in the overall performance of KNN when applied to feature set 2 (with windowing) and that it outperforms every other classifier considered (98.44% accuracy). Apart from this, we observed that the feature set 2 (dataset with windowing) showed an overall better performance than feature set 1 (dataset without windowing).KNN classification with windowing applied feature sets has good absolute performance to a certain degree there should be further exploration to understand its potential.

There are various aspects that can be further explored to build up on our work, some of those aspects include using feature selection techniques to further better to model efficiency, exploring other available classifiers and inspecting the effect of various window sizes while applying the windowing technique on the dataset. Investigating these aspects can be instrumental in strengthening the understanding of using EMG signals to control prosthetic components and other hand assisted manipulators.

REFERENCES

[1] N. Wang, K. Lao, X. Zhang, J. Lin, and X. Zhang, “The recognition of grasping force using LDA,” Biomedical Signal Processing and Control, vol. 47, pp. 393–400, Jan. 2019.

[2] G. Gaudet, M. Raison, and S. Achiche, “Classification of Upper limb phantom movements in transhumeral amputees using electromyographic and kinematic features,” Engineering Applications of Artificial Intelligence, vol. 68, pp. 153–164, Feb. 2018.

[3] Y. Ning and Y. Zhang, “A new approach for multi-channel surface EMG signal simulation,” Biomedical Engineering Letters, vol. 7, no. 1, pp. 45–53, Jan. 2017.

[4] B. Hudgins, P. Parker, and R. N. Scott, “A new strategy for multifunction myoelectric control,” IEEE Transactions on Biomedical Engineering, vol. 40, no. 1, pp. 82–94, 1993.

[5] O. W. Samuel, H. Zhou, X. Li, H. Wang, H. Zhang, A. K. Sangaiah, and G. Li, “Pattern recognition of electromyography signals based on novel time domain features for amputees’ limb motion classification,” Computers & Electrical Engineering, vol. 67, pp. 646–655, Apr. 2018.

[6] **J. Too, A. Rahim, and N. Mohd, “Classification of Hand Movements based on Discrete Wavelet Transform and Enhanced Feature Extraction,” International Journal of Advanced Computer Science and Applications, vol. 10, no. 6, 2019.**

[7] L. H. Smith, L. J. Hargrove, B. A. Lock, and T. A. Kuiken, “Determining the Optimal Window Length for Pattern Recognition-Based Myoelectric Control: Balancing the Competing Effects of Classification Error and Controller Delay,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 19, no. 2, pp. 186–192, Apr. 2011.

[8] G. Jia, H.-K. Lam, J. Liao, and R. Wang, “Classification of electromyographic hand gesture signals using machine learning techniques,” Neurocomputing, vol. 401, pp. 236–248, Aug. 2020.

[9] “UCI Machine Learning Repository.” Website: https://archive.ics.uci.edu/ml/index.php.

[10] Triwiyanto, O. Wahyunggoro, H. A. Nugroho, and Herianto, “Performance Analysis of the Windowing Technique on Elbow Joint Angle Estimation Using Electromyography Signal,” Journal of Physics: Conference Series, vol. 1108, p. 012004, Nov. 2018.

[11] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, “Feature reduction and selection for EMG signal classification,” Expert Systems with Applications, vol. 39, no. 8, pp. 7420–7431, Jun. 2012.

[12] C. Sapsanis, G. Georgoulas, and A. Tzes, “EMG based classification of basic hand movements based on time-frequency features,” 21st Mediterranean Conference on Control and Automation, Jun. 2013.

[13] J. Too, A. Abdullah, N. Mohd Saad, and W. Tee, “EMG Feature Selection and Classification Using a Pbest-Guide Binary Particle Swarm Optimization,” Computation, vol. 7, no. 1, p. 12, Feb. 2019.

[14] Phinyomark, A., Limsakul, C. and Phukpattaranont, P., 2009. A novel feature extraction for robust EMG pattern recognition. *arXiv preprint arXiv:0912.3973*.

[15] I. T. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments,” Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 374, no. 2065, p. 20150202, Apr. 2016.

[16] Lalkhen, A.G. and McCluskey, A., 2008. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia Critical Care & Pain, 8(6), pp.221-223.

[17] Powers, David Martin. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." (2011).

[18] C. Goutte and E. Gaussier, “A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation,” Advances in Information Retrieval, pp. 345–359, 2005.

[19] Ison, M.R. and Artemiadis, P., 2013, October. Beyond user-specificity for emg decoding using multiresolution muscle synergy analysis. In *Dynamic Systems and Control Conference* (Vol. 56123, p. V001T08A006). American Society of Mechanical Engineers.

[20] M. S. Alam and A. S. Arefin, "Real-Time Classification of Multi-Channel Forearm EMG to Recognize Hand Movements using Effective Feature Combination and LDA Classifier," Bangladesh Journal of Medical Physics, vol. 10, no. 1, pp. 25–39, Dec. 2018.

[21] M. Z. Al-Faiz, A. A.Ali, and A. H.Miry, "A k-Nearest Neighbor Based Algorithm for Human Arm Movements Recognition Using EMG Signals," Iraqi Journal for Electrical And Electronic Engineering, vol. 6, no. 2, pp. 158–166, Dec. 2010.

[22] A. A. Abdullah, A. Subasi, and S. M. Qaisar, "Surface EMG Signal Classification by Using WPD and Ensemble Tree Classifiers," CMBEBIH 2017, pp. 475–481, 2017.

[23] E. Gokgoz and A. Subasi, "Comparison of decision tree algorithms for EMG signal classification using DWT," Biomedical Signal Processing and Control, vol. 18, pp. 138–144, Apr. 2015.

[24] N. F. Güler and S. Koçer, "Classification of EMG Signals Using PCA and FFT," Journal of Medical Systems, vol. 29, no. 3, pp. 241–250, Jun. 2005.

[25] B. KumarSingh, K. Verma, and A. S. Thoke, "Investigations on Impact of Feature Normalization Techniques on Classifier's Performance in Breast Tumor Classification," International Journal of Computer Applications, vol. 116, no. 19, pp. 11–15, Apr. 2015.

[26] Patro, S. and Sahu, K.K., 2015. Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462*.

[27] Manikandan, G., Sairam, N., Sharmili, S. and Venkatakrishnan, S., 2013. Achieving privacy in data mining using normalization. *Indian Journal of Science and Technology*, *6*(4), pp.4268-4272.

[28] P. Gulati, A. Sharma, and M. Gupta, "Theoretical Study of Decision Tree Algorithms to Identify Pivotal Factors for Performance Improvement: A Review," International Journal of Computer Applications, vol. 141, no. 14, pp. 19–25, May 2016.

[29] Denil, M., Matheson, D. and De Freitas, N., 2014, January. Narrowing the gap: Random forests in theory and in practice. In *International conference on machine learning* (pp. 665-673).

[30] Ali, J., Khan, R., Ahmad, N. and Maqsood, I., 2012. Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, *9*(5), p.272.

[31] Imandoust, S.B. and Bolandraftar, M., 2013. Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *International Journal of Engineering Research and Applications*, *3*(5), pp.605-610.

[32] Deekshatulu, B.L. and Chandra, P., 2013. Classification of heart disease using k-nearest neighbor and genetic algorithm. *Procedia Technology*, *10*, pp.85-94.

[33] K. Chomboon, P. Chujai, P. Teerarassammee, K. Kerdprasop, and N. Kerdprasop, "An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm," The Proceedings of the 2nd International Conference on Industrial Application Engineering 2015, 2015.

[34] M. Ataei and M. Osanloo, "Using a Combination of Genetic Algorithm and the Grid Search Method to Determine Optimum Cutoff Grades of Multiple Metal Deposits," International Journal of Surface Mining, Reclamation and Environment, vol. 18, no. 1, pp. 60–78, Jan. 2004.

[35] M. M. Ramadhan, I. S. Sitanggang, F. R. Nasution, And A. Ghifari, "Parameter Tuning in Random Forest Based on Grid Search Method for Gender Classification Based on Voice Frequency," DEStech Transactions on Computer Science and Engineering, no. cece, Oct. 2017.

[36] J.-M. Dufour and J. Neves, "Finite-sample inference and nonstandard asymptotics with Monte Carlo tests and R," Conceptual Econometrics Using R, pp. 3–31, 2019.

[37] M. Z. and A. H., "Artificial Human Arm Driven by EMG Signal," MATLAB - A Fundamental Tool for Scientific Computing and Engineering Applications - Volume 1, Sep. 2012.

[38] Singh, S. and Gupta, P., 2014. Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey. *International Journal of Advanced Information Science and Technology (IJAIST)*, *27*(27), pp.97-103.

[39] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," IEEE Transactions on Biomedical Engineering, vol. 50, no. 7, pp. 848–854, Jul. 2003.

[40] Kim, J.S. and Pan, S.B., 2017. A Study on EMG-based Biometrics. *J. Internet Serv. Inf. Secur.*, *7*(2), pp.19-31.

[41] O. Iqbal, S. A. Fattah, and S. Zahin, "Hand movement recognition based on singular value decomposition of surface EMG signal," 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dec. 2017.