Comparative Analysis between SVM & KNN Classifier for EMG Signal Classification on Elementary Time Domain Features

Yogesh Paul
Deptt.of Electrical Engineering
UIET, Kurukshetra University
Haryana-136119, India
Yogeshpaul@ymail.com

Vibha Goyal
Deptt.of Mathematics
Govt. College for women, Karnal,
Haryana-132001, India
Vibhagoyal82@gmail.com

Ram Avtar Jaswal
Deptt.of Electrical Engineering
UIET, Kurukshetra University
Haryana-136119, India
Ramavtar.jaswal@gmail.com

Abstract: - **The extraction of the feature is a significant method to extract the useful information which is hidden in the signal acquired form the types of different. These signals may be speech, EEG, EMG, ECG, EOG etc. Here, within this paper, we carry on further with EMG signal to discuss the comparative analysis in between linear SVM and KNN classifier using time domain features. For the purpose of successful classification of EMG signal, careful selection of feature is required. Within this paper, seven elementary time domain features are realized as they are frequently used for the same.**

Keywords:- **Electromyography Signal, Confusion Matrix, Feature Extraction, Human-Machine Interface, Pattern Classification.**

1. INTRODUCTION

SIGNAL is a single-valued information of some function dependent on time, distance, position etc. and carries some information of a phenomenon. In general term, one can define the signal as the output of some sensing or measuring device, or result of any mathematical operation. There are many signals that we encounter in our day to day life, for instances, light, sound, electrical, physiological signals etc. Signals are of four types: deterministic (stationary), stochastic, fractal and chaotic.

- Deterministic signals are those, which can be represented by some mathematical functions, for instance algebraic, trigonometrical etc. and its future value can be forecasted exactly for any cycle or time if the past value of the signal is known.
- Stochastic signals are those signals whose future value prediction is nearly impossible, as these signals show random nature, that's why these signals are also called random signals or non-deterministic signals. These signals cannot be represented by a mathematical function and are represented with the help of statistical properties or features. In the case of physiological signals non-stationary signals are:
  - Electromyography (EMG): Which is generated by our nerve fiber in the muscles while performing any physical activity. It is used to access the health of the muscle in medical and in Human-Machine Interface used as a control signal for prosthetics.
  - Electroencephalography (EEG): Which is generated by the firing of neurons in our brain due to ionic changes in our brain.
  - Electrooculography (EOG): Which measures the potential between the front and the back of the human eye.
- Fractal signals are included in the biomedical signal processing in very recent years. They possess very interesting features, that at every level of intensification they look very similar, this property is arbitrated as scale-invariance.
- Chaotic signals are kind of deterministic signals but its exact future prediction is not possible. This definition seems to contradict the definition of deterministic signal but that is so because of "sensitive dependence on the signal on initial conditions". The trajectory of these signals mimics the deterministic signals but these signals are so sensitive to their past value that determination of future value after a certain short period of time contains unignorable errors.

Now, the structure of this paper follows as; besides introduction in section 1, we have time domain feature description in section 2, then we have description about the EMG data in section 3, followed on by the experimental result of pattern classification in section 4 and finally, the paper lasted with the future scope and conclusion in section 5 and section 6 respectively.

2. Time Domain Based Feature Set

A statistical feature so discusses here within this section are regarded as time domain feature as they directly execute on a signal which is time based. Besides this time domain, there is a frequency or spectral domain features as well which are beyond the scope of this paper. Time domain features are quick and easily implementable, as there is no transformation required, which are mostly of Fourier type. In the case of bio signals, there is a major disadvantage of these features that the signals acquired are of the non-deterministic type, as a result, there is a change in statistical properties over time. The mathematical definitions related to the same here mentioned.

2.1 Mean Absolute Value (MAV)

One of the most popular features for EMG signal analysis is the absolute value, which is similar to IEMG (or integrated EMG) feature. It is an average of the absolute value of EMG signal in segment, which is defined as,

$$MAV = \frac{1}{N}$$

Where x_i is the EMG signal amplitude over segment i while N denotes the total length of the signal. It is also known by Average rectified value (ARV), averaged absolute value (AAV), integral of absolute value (IAV).

The program in relation with the mean absolute value feature is given as,

```
function x= mav(signal)
sum=0;
n=length(signal);
for i=1:n
    sum= sum + abs(signal(i));
end
x = (sum/n);
```

2.2 Root Mean Square (RMS)

Root mean square (RMS) is the another popular feature for analysis of EMG signal that is similar to the standard deviation, which is given as,

The program in relation with the root mean square feature is given as,

```
function x= rms(signal)
sum=0;
y=0;
n=length(signal);
for i=1:n
    sum= sum + (signal(i)^2);
end
y = (sum/n);
x = sqrt(y);
```

2.3 Simple Square Integral (SSI)

It is simply the summation of square values of EMG signal amplitude. It is simply the measures of the energy of a signal.

The program in relation with the simple square integral feature is given as,

```
function x = ssi(signal)
sum=0;
for i=1:length(signal)
    a= (signal(i))^2;
    sum= sum + a;
end
x = sum;
```

2.4 Waveform Length (WL)

Waveform length is the measure of the complexity of the signal depends on amplitude, frequency and time of the signal. It is the cumulative length of the signal.

The program in relation with the Waveform length feature is given as,

```
function x = wl(signal)
sum=0;
for i=2:length(signal)
    sum= sum + abs(signal(i)-signal(i-1));
end
x = sum;
```

2.5 Average Amplitude Change (AAC)

It is nearly equivalent to the waveform length feature with the condition that wavelength is averaged, which is given as,

The program in relation to the Average Amplitude Change feature is given as,

```
function x = aac(signal)
n=length(signal)
sum=0;
for i=2:n
a=abs(signal(i)-signal(i-1));
sum=sum+a;
end
x= sum/n;
```

2.6 Difference Absolute Standard Deviation Value (DASDV)

It is similar to the RMS of a signal. It calculates the standard deviation of the waveform length.

The program in relation with the difference absolute standard deviation value feature is given as,

```
function x = dasdv(signal)
n=length(signal);
sum=0;
for i=2:n
a=(signal(i)-signal(i-1))^2;
sum=sum+a;
end
x= sqrt(sum/(n-1));
```

2.7 Log Detector (LOG)

It gives the strength in the signal (muscle contraction) and varies with the amplitude of the signal.

```
function x = log(signal)
n=length(signal);
x = exp((sum(log10(abs(signal))))/n);
```

3. EMG DATA SELECTION

Surface EMG signal data used in this work is taken from UCI machine learning repository. The data has been recorded from 2 forearm muscles Flexor Capri Ulnaris and Extensor Capri Radialis, Longus and Brevis using Delsys Bagnoli 2-channel EMG system. This data is collected from 5 healthy subjects 2 males and 3 females of approximately same age, out of which only one female subject is considered in this study.

The protocol used for the data acquisition is as follow: The Subject was asked to repeat the six basic hand movement out of which three has been considered in this study that is spherical for holding spherical tools, cylindrical for holding cylindrical tools and hook for supporting a heavy load, for each movement the data is recorded for 6 seconds. 30 trails were taken for each basic movement. Data is sampled at 500 Hertz. Preprocessing of the signal is done by passing it through a Butterworth band pass filter with a high and low cut off of 500Hz and 15Hz respectively. Also, power line interference has been removed by applying a notch filter at 50Hz.

4. EXPERIMENTAL RESULT

In connection with section 3, we have here the confusion matrix for each of the different time domain based feature mentioned in section 2 on an individual basis for the classes of different in collective using linear support vector machine and k-nearest neighbor classifier.

Support Vector Machine (SVM) is a classifier that performs classification tasks by constructing hyper planes in a high dimensional space that separates cases of different class labels. The cases are so mapped that the classes are divided by a clear gap that is as wide as possible. KNN classifier is the basic classifier and operates on the property that classification of unknown instances can be done by relating the unknown to the known according to some distance/similarity function. The unknown instance is labeled with the same class label as that of the known nearest neighbor.

4.1 Confusion Matrix for MAV feature.

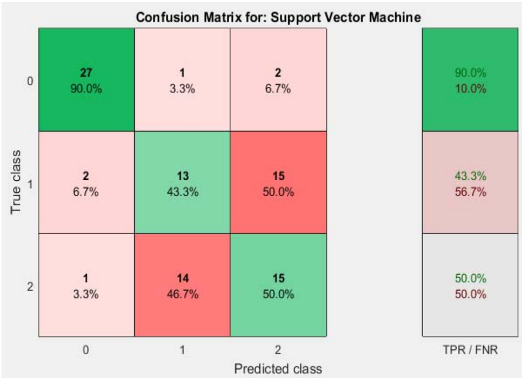


Fig.1 Confusion Matrix of MAV using SVM as a classifier.

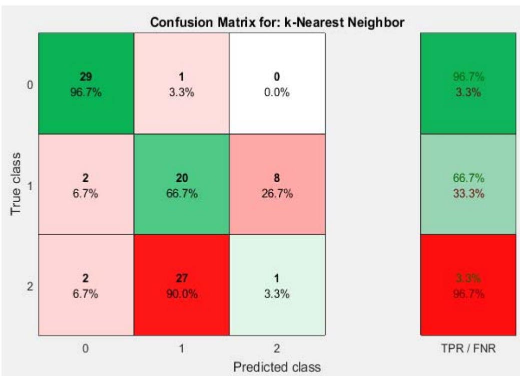


Fig.2 Confusion Matrix of MAV using KNN as a classifier.

4.2 Confusion Matrix for RMS feature.

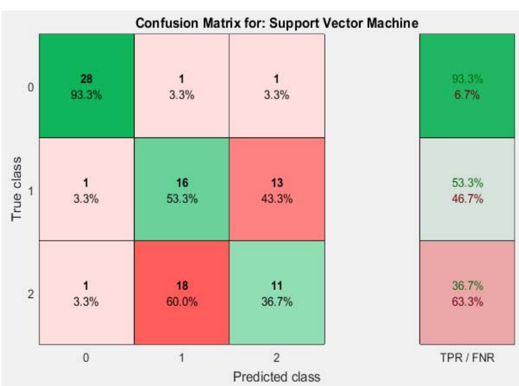


Fig.3 Confusion Matrix of RMS using SVM as a classifier.

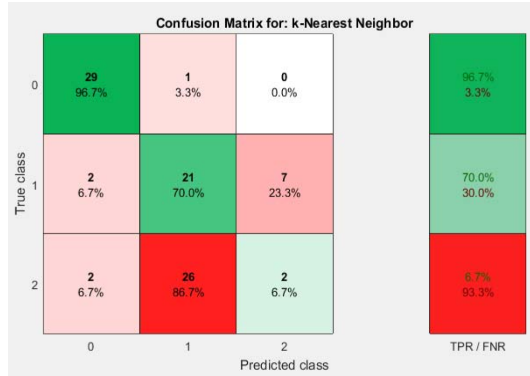


Fig.4 Confusion Matrix of RMS using KNN as a classifier.

4.3 Confusion Matrix for SSI feature.

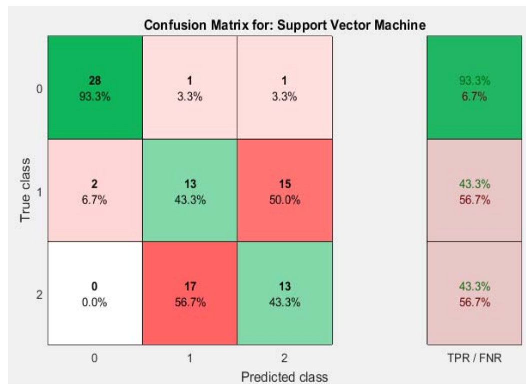


Fig.5 Confusion Matrix of SSI using SVM as a classifier.

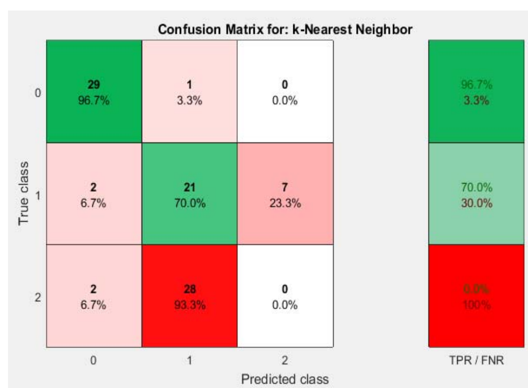


Fig.6 Confusion Matrix of SSI using KNN as a classifier.

4.4 Confusion Matrix for WL feature.

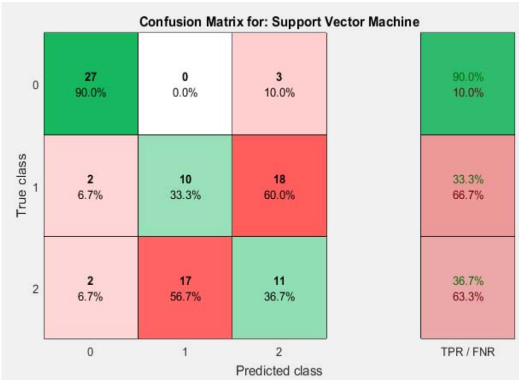


Fig.7 Confusion Matrix of WL using SVM as a classifier.

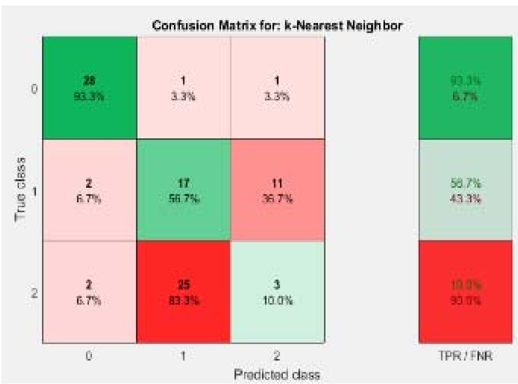


Fig.8 Confusion Matrix of WL using KNN as a classifier.

4.5 Confusion Matrix for AAC feature.

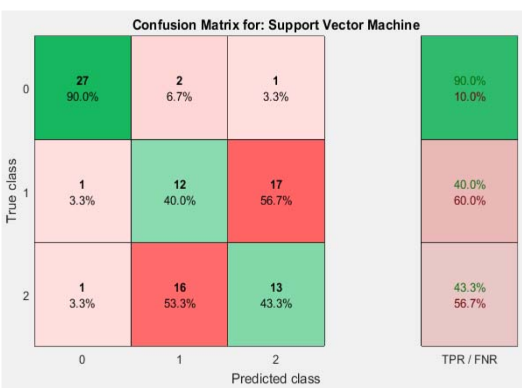


Fig.9 Confusion Matrix of AAC using SVM as a classifier.

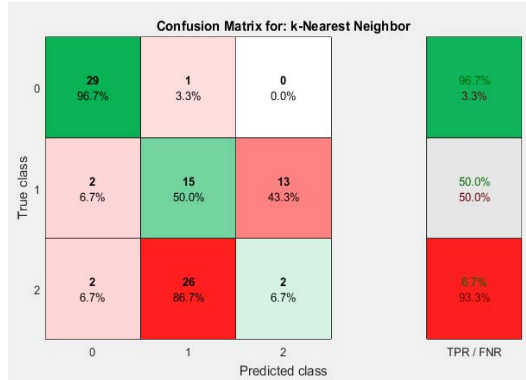


Fig.10 Confusion Matrix of AAC using KNN as a classifier.

4.6 Confusion Matrix for DASDV feature.

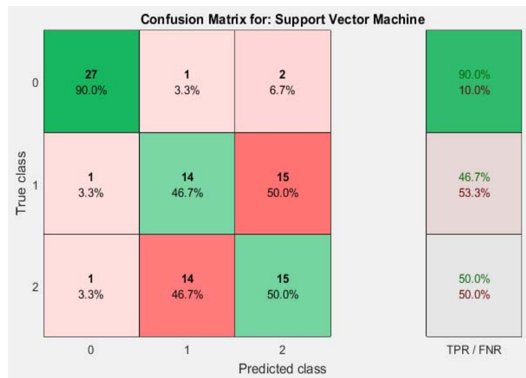


Fig.11 Confusion Matrix of DASDV using SVM as a classifier.

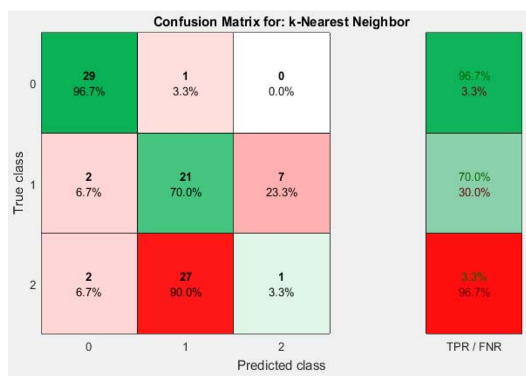


Fig.12 Confusion Matrix of DASDV using KNN as a classifier.

4.7 Confusion Matrix for LOG feature.

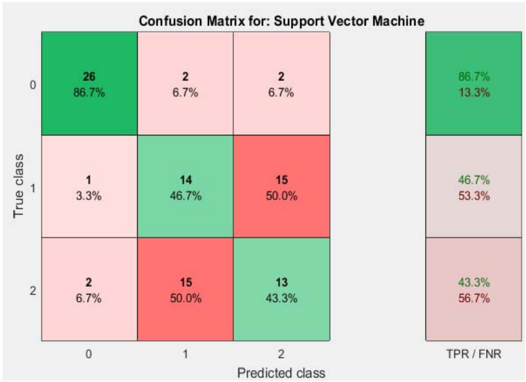


Fig.13 Confusion Matrix of LOG using SVM as a classifier.

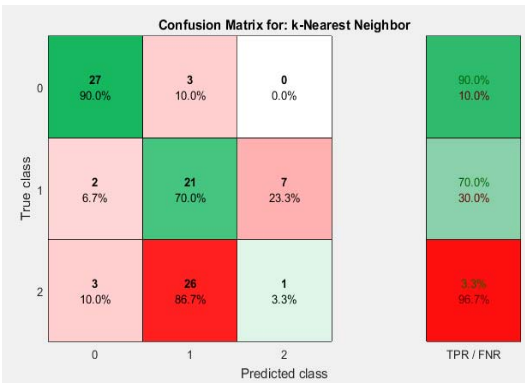


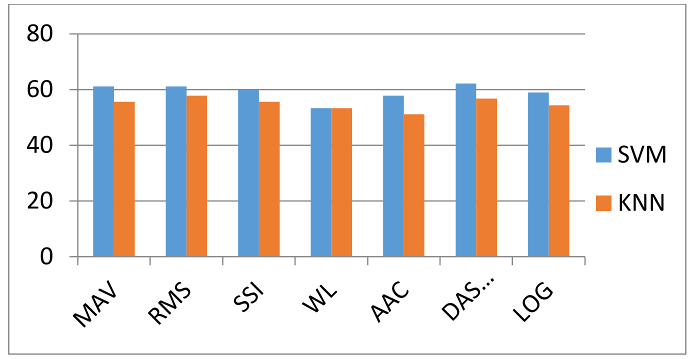
Fig.14 Confusion Matrix of LOG using KNN as a classifier.

5. FUTURE SCOPE

There are numerous classifiers present for the EMG data classification. In this paper, SVM and KNN classifier were compared for EMG classification (pattern recognition). Further, comparison of other classifiers can be done for EMG, EEG and other bio-signals which will give the idea about the best classifier for EMG, EEG data classification making classification of the data easier for the further use.

6. CONCLUSION

After employing the two classifiers, from the above section 4, we finally conclude that the linear support vector machine gives higher accuracy percentage other than a k-nearest neighbor, for the mentioned time domain feature of EMG signal. The plot is shown below depicting the results of the discussed features.



REFERENCES

[1] B. Hudgins, P. Parker, & R. N. Scott, “A new strategy for multifunction myoelectric control,” *IEEE Transactions on Biomedical Engineering,* vol. 40(1), pp. 82-94, 1993.

[2] M. Zardoshti-Kermani, B. C. Wheeler, K. Badie, & R. M.Hashemi, “EMG feature evaluation for movement control of upper extremity prostheses,” *IEEE Transactions on Rehabilitation Engineering,* vol. 3(4), pp. 324-333, 1995.

[3] R. Boostani, & M. H. Moradi, “Evaluation of the forearm EMG signal features for the control of a prosthetic hand,” *Physiological Measurement*, vol. 24(2), pp. 309-319, 2003.

[4] K. S. Kim, H. H. Choi, C. S. Moon, & C. W. Mun, “Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions,” *Current Applied Physics,* vol. 11(3), pp. 740-745, 2011.

[5] L. Philipson, “The electromyographic signal used for control of upper extremity prostheses and for quantification of motor blockade during epidural anaesthesia,” Ph.D. Thesis, *Linkoping University, Linkoping, Sweden*, 1987.

[6] D. Tkach, H. Huang, & T. A. Kuiken, “Study of stability of time-domain features for electromyographic pattern recognition,” *Journal of Neuro Engineering and Rehabilitation,* vol. 7(21), 2010.

[7] Sijiang Du, & M. Vuskovic, “Temporal vs. spectral approach to feature extraction from prehensile EMG signals,” In *Proceedings of IEEE International Conference on information Reuse and Integration,* pp. 344-350, 2004.

[8] M.A. Oskoei, & H. Hu, “Support vector machine based classification scheme for myoelectric control applied to upper limb,” *IEEE Transactions on Biomedical Engineering,* vol. 55(8), pp. 1956-1965, 2008.

[9] C. Sapsanis, G. Georgoulas, A. Tzes, D. Lymberopoulos, “Improving EMG based classification of basic hand movements

using EMD," *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 13* (EMBC 13), July 3-7, pp. 5754 - 5757, 2013.

[10] Eugene N. Bruce, *Biomedical Signal Processing and Signal Modeling*. Wiley, 2009.