Chapter 6
Classification of EMG Data to Assess the Fatigue Level of Different Subjects While Providing Traction Therapy

6.1 Introduction

The EMG signal is used for muscle fatigue analysis in clinical rehabilitation research. The analysis of these signals finds application in various areas such as medical diagnosis, sports biomechanics, myoelectric control, and ergonomics [Merletti et al., 2004]. The EMG signal is used in numerous studies for classification. The non-invasive EMG technique is widely used to analyze muscle fatigue [Cifrek et al., 2009]. The EMG signal feature extraction can be accomplished using different techniques such as signal amplitude and frequency domain analysis [Ainishet, et al., 1996].

Classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observations. Signal classification is focused on improving the accuracy in the discrimination of muscle activity [Scott, et al., 1988; Kuruganti, et al., 1995].

Types of classification are:

- Linear Classifier: Logistic Regression, Naive Bayes Classifier
- **Support Vector Machines**
- **Decision Trees**
- Boosted Trees
- Random Forest
- Neural Networks
- Nearest Neighbor

6.1.1 Support Vector Machine

The Support vector machine is a discriminative classifier defined by a separating hyperplane. In machine learning, support vector machine is supervised learning models which associated learning algorithms that analyze data used for classification and regression analysis. SVMs are a new technique suitable for binary classification tasks, which is related to and contains elements of non-parametric applied statistics, neural networks, and machine learning. A Support vector machine (SVM) has been shown to improve the performance in the classification task. SVM is founded in the framework of the statistical learning theory, which is appropriate for approaching classification and regression problems. [Cortes et al., 1995; Evgeniou et al., 2000; Gunn and Steve, 1998; Hsu et al., 2003].

SVM is a new computational technique based on the statistical learning theory [Vapnik, 2013]. In SVM, input data is mapped into a high dimensional dot product space called a feature space. Then an (n-1) dimensional hyperplane separates the space into two parts. Let n-dimensional input x_i (i=1, 2…, l) is labelled as Y_i=1 for class1 and as y_i=-1 for class 2 by y_i matrix. A hyperplane can be defined for linearly separable data.

$$F(x) = \omega . x + b = \sum_{i=1}^{n} \omega_i \, x_i + b = 0 \qquad \text{(i)}$$

Where ω an n-dimensional vector and b is a scalar. The vector ω and the scalar b determine the position of the separating hyperplane. Function sgn (f(x)) is also called the decision function. A distinctly separating hyperplane satisfies the constraints.

$$Y_i(x_i.\ \omega + b) -1 \geq 0 \leftrightarrow \begin{cases} f(x_i) = x_i . \omega + b \geq 1 & y_{i=} +1 \\ f(x_i) = x_i . \omega + b \leq 1 & y_{i=} -1 \end{cases} \qquad \text{(ii)}$$

The hyperplane that creates maximum limit is called an optimal hyperplane. In the equation below, ξ_i is the independent variable and C is the error penalty. The minimized solution of the hyperplane is as followed:

$\text{Ø}(\omega, \xi) = ½(\omega.\omega) + C\left(\sum_{i=1}^{l} \xi_i\right)$

(iii)

depending on

$y_i\,[(x_i.\omega)+b] \geq 1-\xi_i, \quad i=1,2.....l$ (iv)

ξ_i measures the distance between the limit and the samples x_i on the other side of the limit. The calculation can be simplified as followed:

$V(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{ij=1}^{l} \alpha_i\, \alpha_j\, y_i y_j\, Ker(x_i x_j)$ (v)

Depending on

$\sum_{i=1}^{l} y_i \alpha_i = 0, \quad C \geq \alpha \geq 0, i = 1,2 \ldots, l$ (vi)

The function Ker ($x_i x_j$) is called as kernel function returns the dot product of the feature space maps of the original data points [Wang et al., 2009].

6.1.2 Complex Tree

The decision tree builds classification or regression models in the form of a tree structure. A decision node has two or more branches. The basic idea involved in any multistage method is to break up a complex decision into a union of several simpler decisions, the final solution obtained this way would resemble the intended desired solution [Dattacharya et al., 1986].

The decision tree algorithm recursively splits a data set of records using a depth-first greedy approach [Hunt et al., 1966] or the breadth-first approach [Shafer et al., 1996] until all the data items are attached to a particular class. Root internal and leaf nodes form the structure of a decision tree, and the structure of the tree is utilized to classify unknown data records, easy to interpret for small-sized trees, and accuracy is comparable to other classification techniques for many simple data sets. The decision tree classification in two stages: tree building and tree pruning. Tree building is performed in a top-down manner. As it is mentioned, during the first stage of classification, the tree is recursively divided until all the data items are related to the same class label [Hunt et al., 1966].

Many previous studies have been done for the classification of muscle fatigue, which is described below.

Alkan et al. (2012) proposed a surface EMG signal classification system which uses five discriminant function and an SVM classifier. This classifier gives a very good accuracy rate (99%) for four movements with the classification rate (1%). Al-Mulla et al. (2012) presented preliminary empirical evidence demonstrating that the developed features and methods for fatigue detection improve the current state of the art. The processing and classification of the EMG signal for muscle fatigue analysis.

Sharawardi et al. (2014) presented the implementation of the SVM technique (Support vector machine) for muscle fatigue analysis using single-channel EMG data. This paper concluded that the SVM method is significantly better than both the KNN and ANN.

Karthick et al. (2018) proposed time-frequency distributions are able to show the nonstationary variations of EMG signals. Most of the features a statistically significant difference in muscle fatigue and non-fatigue conditions. The combination of EMBD- polynomial kernel-based SVM is found to be the most accurate (91%) accuracy is classifying the conditions with the feature selected using (GA) genetic algorithm.

K.Uma Rani (2017) evaluated muscle strength of different disorders the time-dependent average values of RMS value, mean and median frequencies of myopathy, neuropathy, and healthy signals are compared, by comparing the time-dependent values one can know the energy of different disorder muscles.

This study aimed to classify EMG data to analyze the fatigue level of various neck pain patients while providing traction therapy over a specified period. Two classification techniques, namely, support vector machine and decision tree, were implemented using the time-frequency features of EMG data. The complex tree and support vector machine classifier was utilized to evaluate the classification performance.

6.2 Methodology

EMG data of five male and seven female patients suffering from neck pain were recorded under cervical traction treatment. In this traction treatment, the subject was sitting on an armchair and, the tension of 7 kg was applied to the subject for 15 min per day for a week in the BHU hospital. For the acquisition, a wireless EMG sensor, which has a range of 40 meters, was utilized to obtain real-time EMG data [Dong et al., 2014].

In EMG processing, the acquired EMG signal was band-pass filtered between 0.5Hz and 500 Hz. With a 50 Hz Notch filter to remove power line noise). Subsequently, various statistical and frequency features such as MAV, RMS, SD, MNF, and MDF were used to extract the main information from the EMG data. Afterward, using extracted features, the data of day 1, day 3, and day 7 is classified using a support vector machine (SVM) and decision tree. This is done to classify the different traction therapy stages. The day 1, day 3, and day 7 EMG data were considered as the start, mid, and end of the therapy, respectively. Accordingly, three classes are chosen for classification.

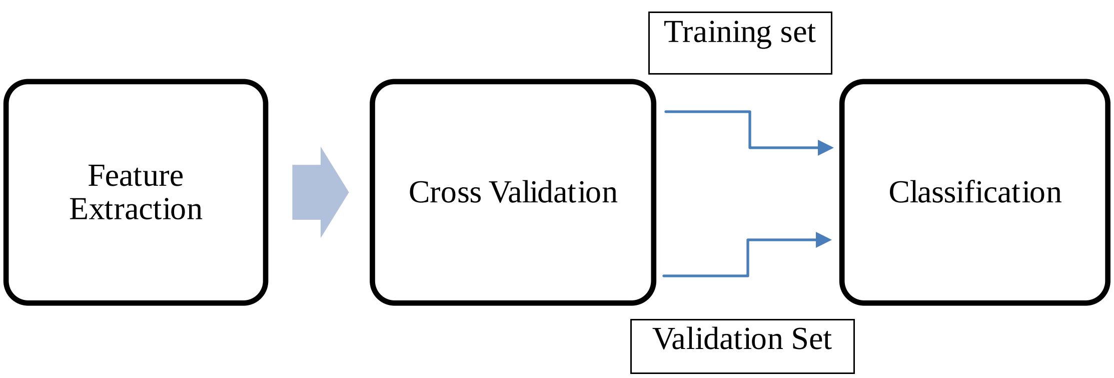

Fig 6.1 Framework elucidates the method to classify Fatigue level

6.2.1 Classification Evaluation Criteria

In the present study, two evolution criteria [Schlogl et al., 2007], namely the classification accuracy and the kappa coefficient, have been chosen to compare the effectiveness of the classification performance. The classification accuracy defined as

(A) Classification Accuracy

The classification accuracy (ACC) or the error rate (ERR = 1-ACC) is the most widely used evaluation method. The maximum accuracy can never exceed 100%. Sometimes, this could be a disadvantage, especially when two classification systems should be compared, and both provide a result close to 100%.

$$\mathit{Classification\ Accuracy} = P_o = \frac{\sum_{i=1}^{c} n_{ii}}{\sum_{i=1}^{c}\sum_{j=1}^{c} n_{ij}} \qquad \text{(vii)}$$

Where n_{ii} and n_{ij} represent the elements of the confusion matrix and indicate how many times class i has been predicted as class j. If i=j, then a true class is predicted by the classifier. C is the number of class, which is two in our case.

(B) Kappa Value

Kappa coefficient addresses several of the critiques on the accuracy measure. The calculation of the kappa (k) uses the overall agreement P_o =ACC, which is equal to the classification accuracy,

The kappa coefficient (k) is defined as

$$\mathit{kappa\ coefficient}\ (k) = \frac{P_0 - P_e}{1 - P_e} \qquad \text{(viii)}$$

P_o = Total accuracy

P_e = Random accuracy

6.3 Result & Discussion

This section elucidates the classification performance of the proposed method in comparison with the support vector machine and decision tree explained in this chapter. Algorithms used in this work were developed and implemented on a computer having 12 GB of RAM and an Intel core i5 @ 3.4 GHz processor using the 64-bit version of MATLAB R2018a software and were applied to the two different classifier support vector machine, and decision tree explained in this section.

The time and frequency features are utilized to classify the neck muscle fatigue condition. Table 6.1 shows that the acquired EMG data were used further to extract various features of neck muscle fatigue in the time and frequency domain features. Then the data analysis of day one, day three, and day seven were classified using extracted features. The neck muscle fatigue studies were classified in MATLAB.

Table 6.1 Time and frequency domain features between during traction in the sitting position

Subjects	Time Domain Features						Frequency Domain Features			
	MAV		RMS		SD		MF		MDF	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
1.1d)	2.182	2.170	2.235	2.195	5.190	3.296	0.055	0.066	2.533	7.356
3d)	1.784	1.706	2.111	1.878	1.296	8.367	0.057	0.031	1.153	8.969
7d)	1.676	1.662	1.877	1.783	9.043	6.964	0.024	0.036	8.484	9.363
2. 1d)	2.207	2.197	2.296	2.273	6.584	5.875	0.016	0.014	7.834	7.705
3d)	1.425	1.582	1.476	1.985	3.839	1.409	0.017	0.055	7.712	0.013
7d)	9.188	0.001	3.491	0.002	3.490	0.002	0.145	0.165	0.088	0.163
3.1d)	1.689	1.448	2.987	1.612	7.393	2.602	0.064	0.069	0.046	9.104
3d)	1.688	1.481	1.917	1.680	9.670	8.575	0.035	0.050	9.642	9.722
7d)	1.693	1.688	1.917	1.885	9.670	8.588	0.035	0.032	9.072	9.642
4.1d)	1.397	1.396	1.488	1.449	5.353	3.789	0.027	0.015	8.259	7.718
3d)	2.187	1.464	2.356	1.652	8.838	8.147	0.021	0.032	2.789	3.167
7d)	1.428	1.426	1.511	1.496	4.973	4.503	0.019	0.017	8.064	7.906

5.1d)	1.421	1.398	1.553	1.486	6.704	5.051	0.034	0.025	8.837	8.129
3d)	2.969	3.009	3.753	3.796	3.527	3.494	0.289	0.273	0.130	0.122
7d)	1.449	1.442	1.490	1.478	3.446	3.259	0.012	0.010	7.557	7.597
6.1d)	2.213	2.187	2.400	2.311	9.398	7.486	0.031	0.024	8.493	8.033
3d)	1.373	1.385	1.499	1.447	6.349	4.191	0.030	0.019	8.761	7.848
7d)	1.416	1.405	1.488	1.465	4.563	4.284	0.019	0.019	7.937	8.239
7. 1d)	3.760	2.829	4.841	3.506	4.311	2.720	0.069	0.089	0.044	0.059
3d)	3.379	2.551	4.524	3.687	3.945	2.897	0.053	0.058	0.059	0.081
7d)	2.317	2.277	2.458	2.380	8.910	7.185	0.021	0.028	7.911	8.278
8.1d)	4.397	3.220	4.723	1.335	4.399	1.323	0.108	0.153	0.627	0.663
3d)	2.120	2.112	2.333	2.155	9.904	3.861	0.024	0.080	7.429	8.770
7d)	4.656	4.388	5.960	5.576	5.826	5.398	0.241	0.310	0.097	0.112
9.1d)	2.233	1.749	2.638	2.057	1.494	1.123	0.038	0.053	7.057	3.416
3d)	2.142	2.101	2.175	2.444	3.767	1.122	0.010	0.027	2.471	3.037
7d)	5.244	1.972	2.891	2.177	2.883	9.266	0.086	0.030	0.050	2.927
10.1)	1.990	1.445	2.654	1.572	2.246	6.852	0.086	0.040	0.054	8.876
3d)	1.690	1.411	2.104	1.532	1.564	6.045	0.083	0.047	0.031	8.516
7d)	1.667	1.507	2.073	1.746	1.514	1.030	0.007	0.066	0.028	3.675
11.1d)	2.208	2.087	2.302	2.118	6.513	3.573	0.014	0.008	2.605	2.467
3d)	2.195	2.092	2.240	2.102	4.592	2.119	0.009	0.003	2.502	2.421
7d)	1.278	1.382	1.462	1.426	7.520	3563	0.045	0.015	3.258	2.556
12.1d)	6.709	6.414	8.445	8.114	8.314	7.933	0.136	0.136	0.087	0.091
3d)	6.819	1.780	8.640	4.883	8.489	4.608	0.134	0.114	0.089	0.066
7d)	2.196	1.598	7.692	1.492	7.847	2.074	0.670	0.120	0.014	0.069

Table 6.2 Classification performance of support vector machine and complex tree

	Complex Tree		Support Vector Machine	
	Accuracy	Kappa Value	Accuracy	Kappa Value
1.	89.6	0.84	92.7	0.88
2.	93.8	0.89	94.8	0.91
3.	96.9	0.94	99.0	0.98
4.	96.9	0.94	97.9	0.96
5.	97.9	0.96	97.9	0.96
6.	90.6	0.85	91.7	0.86
7.	95.3	0.92	100	0.92
8.	100	0.81	97.9	0.95
9.	96.9	0.95	90.6	0.85
10.	83	0.74	87.9	0.80
11.	86	0.79	91.2	0.86
12.	82	0.73	89.0	0.83

Table 6.2 shows that accuracy and kappa value are estimated from the recorded EMG data. The accuracy and kappa value were observed for complex tree and support vector machine. The accuracy of the support vector machine is observed high as compared to the complex tree classifier. Similarly, the Kappa value is high as compared to a complex tree. Therefore these time and frequency features could also be used to study the muscle activities in order to improve the classification accuracy.

6.4 Conclusion

This study represents the classification of EMG data to assess the fatigue level of different neck pain patients while giving traction therapy. First, the time and frequency domain features were extracted from the raw EMG data. Subsequently, SVM and decision tree classifiers are used to evaluate the classification performance. The obtained results show that the accuracy and kappa value of the SVM is higher than that of the complex tree classifier.