Classification of Neuromuscular Diseases using Dominant MUAP Based on Wavelet Domain Features and Improving Its Accuracy using SVM

J.Defino[1], S. Mary Vasanthi[2]

[1]*P.G. Scholar, M.E. Applied Electronics,*
St. Xavier's Catholic College of Engineering, Chunkankadai,
Nagercoil-629003, Tamil Nadu, India

[2]*Assistant Professor, Department of ECE,*
St. Xavier's Catholic College of Engineering, Chunkankadai,
Nagercoil-629003, Tamil Nadu, India

Abstract--**In this paper, the neuromuscular diseases are classified into Myopathy, Amyotrophic Lateral Sclerosis (ALS) and Normal persons by extracting wavelet features from dominant Motor Unit Action Potential (MUAP). Initially, the electromyography (EMG) signal is preprocessed by three techniques namely, sampling, normalization and band pass filtering. The selection of dominant MUAP is done by dividing the preprocessed EMG signal into several frames using windowing technique and evaluating the energy associated with each frame. Among all frames, the frame which has the highest energy content is selected as the dominant MUAP. Wavelet features of dominant MUAP are obtained using Discrete Wavelet Transform (DWT). Statistics over the set of wavelet features are extracted to reduce the dimensionality. For the classification purpose, first k-nearest neighbor classifier is used and its performance is evaluated. Further, to improve the performance, Support Vector Machine (SVM) classifier is used. The performance of the proposed method is measured in terms of specificity, sensitivity and total classification accuracy.**

Keywords--- Electromyography (EMG), dominant MUAP, Discrete Wavelet Transform, k-Nearest Neighbor Classifier, Support Vector Machine Classifier.

I. INTRODUCTION

Electromyography (EMG) evaluates and measures the electrical activities produced by the activation of the muscles. The contraction of a muscle may be either voluntary or involuntary. The functional unit of muscle contraction is a motor unit. The motor unit contains a single alpha motor neuron and all the muscle fibres associated with the neurons. When the action potential of the motor unit reaches a depolarization threshold, the muscle fibre contracts. An electromagnetic field will be generated by depolarization and the potential will be measured as a voltage.The depolarization spreads along the muscle membranes and generates muscle action potential. The summation of the individual muscle action potentials from the muscle fibres that surrounds the single motor unit is called the Motor Unit Action Potential (MUAP). EMG signal consists of the algebraic summation the motor unit action potential within the pickup area of the electrode that has been used.The pick-up area of an electrode contains more than one motorunit because muscle fibers of several motor units are mingled together throughout the entire muscle.

Neuromuscular diseases affect the neuromuscular system which includes the muscles, the peripheral motor nerves, the neuromuscular junction and the motor nerves in the spinal cord. There are over more than hundreds of neuromuscular diseases. Some of them include Amyotrophic Lateral Sclerosis (ALS), Neuropathy and Myopathy. A disorder in which a muscle dysfunction is caused due to inherent muscle defect is called myopathy. Amyotrophic Lateral Sclerosis affects the neurons that are responsible for controlling the voluntary muscle.To understand the interrelationship among various neuromuscular disorders, EMG signal classification is very important.

In [1] Empirical Mode Decomposition (EMD) technique is introduced for EMG signal filtering and compared the results with wavelet transform. In [2] EMG signals are segmented and classified using Artificial Neural Network (ANN) technique. The combined wavelet and autoregressive features are classified using Adaptive Neuro Fuzzy Inference System (ANFIS), SVM and Fuzzy SVM (FSVM) in [3].In [4] the

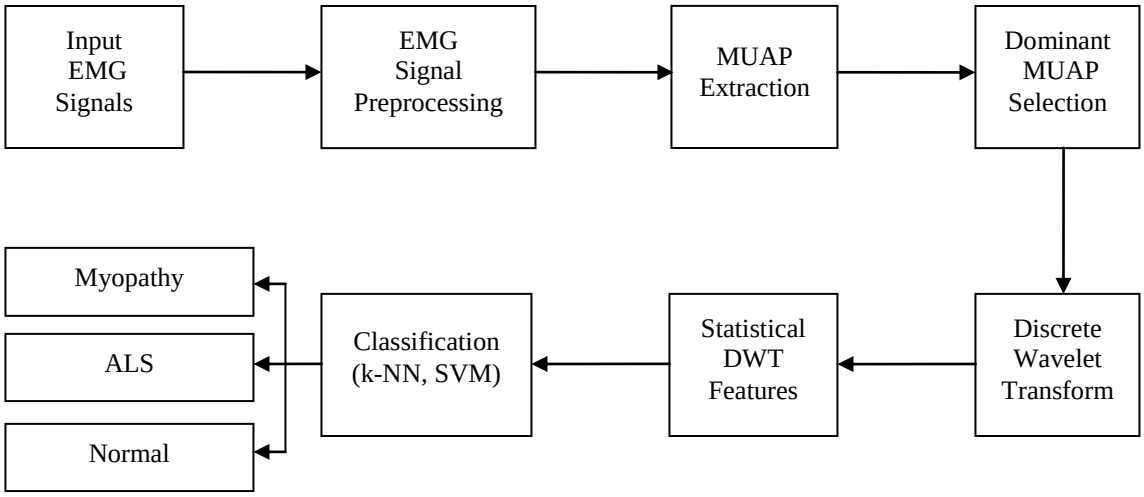

Fig.1 Block Diagram of Proposed Method for Classification of Neuromuscular Diseases

neuromuscular diseases are classified using SVM by Symlet 4 wavelet technique. It is reported in [6] that the time frequency based features such as Short Time Fourier Transform, wavelet transform and wavelet packet transform are efficient for EMG signal classification. Mean and maxima of DWT coefficients are used to classify neuromuscular diseases using k-NN classifier in [7]. Using Fast Fourier Transform features Multilayer perceptron and SVM classifiers are used to classify neuromuscular diseases in [8]. By using MUAP detection and clustering technique the EMG signals are classified by a two stage classification method in [9] and SVM in [10]. In [11] statistical pattern recognition technique is used for the classification of EMG signals. A system that detects motor unit potentials, cluster them and classify the detected motor unit potentials into motor unit potential trains using supervised classification algorithm is proposed in [12].

Various EMG signal decomposition techniques like blind source separation of convolved mixtures, ANN and high precision EMG signal decomposition techniques are compared in [14]. In [15] four different wavelets are analyzed that describes MUAP morphology in the time frequency plane. A hybrid classifier fusion scheme is proposed in [16] for analyzing simulated and real EMG signals. Wavelet packet transform features are extracted from MUAPs and classified using fuzzy C-means clustering technique in [17]. Using DWT features the neuromuscular diseases are classified by neuro fuzzy classification technique in [18], Fuzzy SVM in [19] and particle swarm optimization SVM in [20]. In [21] autoregressive features are extracted and classified using wavelet neural network classifier. By extracting wavelet features from dominant MUAP the neuromuscular diseases are classified using k-NN classifier in [22].

In the proposed method, dominant MUAP based Discrete Wavelet Transform (DWT) feature extraction scheme is used for the classification of neuromuscular diseases into normal, myopathy and ALS.

II. MATERIALS AND METHODS

The steps involved in the proposed method are EMG signal preprocessing, MUAP extraction, dominant MUAP selection, DWT, statistical DWT features and classification as shown in Fig.1.

A. EMG Signal Description

In the proposed method, a database of clinical signals including a group of normal persons, a group of ALS patients and a group of myopathy patients are used. The EMG signals are recorded at low voluntary and constant level of contraction. To monitor the signal quality visual and audio feedback is used. Using a standard concentric needle electrode the EMG signals are taken from five different places in the muscle. The normal group consists of6 male and 4 female subjects in the age of 21 – 37 years. The ALS group consists of 4 male and 4 female subjects in the age of 35 – 67 years. The myopathy group consists of 5 male and 2 female subjects in the age of 19 – 63 years [13].

B. EMG Signal Preprocessing

Preprocessing the EMG signal is important to reduce the amount of data to be processed and to reduce the effect of noise. The preprocessing steps includes sampling, normalization and band pass filtering. Sampling process which is the first step in EMG preprocessing reduces the number of samples in the EMG signal. The absolute amplitude values of EMG signal are influenced by factors like skin filtering influence, electrode repositioning, etc. So in repeated measurements of the same subject, it is not possible to realize reliable comparisons. Furthermore, comparisons of values of certain muscle in different subjects are also not possible on the absolute scale. Therefore it is important to normalize the EMG signal in some way. The amplitude of the signal measured during maximal voluntary isometric contraction of the corresponding muscle is chosen as the value to which normalization is made. Normalization of EMG signal $x(n)$ with M number of samples is expressed as

$$y(n) = \frac{x(n)}{\substack{Maximum value \\ of EMG signal}} \;;\; n = 0,1,\dots,M-1 \qquad (1)$$

Generally, two types of artifacts exists in an EMG signal namely, technical and biological artifacts. Technical artifacts are caused by amplification noise, bad electrical contact points and interference from external power sources. It will be observed as a strong 50 Hz or 60 Hz frequency component in the signal. The biological artifacts are caused due to the movement of the patients, sweating and heart rate activity. It will appear as a 250 Hz frequency component in the EMG signal. In order to remove these artifacts the EMG signal is band pass filtered [9] in the range of 100 to 200 Hz.

C. MUAP Extraction

MUAP extraction involves segmentation of the composite EMG signal into its component motor unit action potential trains [12]. In the proposed method, a rectangular window function is used to divide the EMG signals into MUAP trains called frames. A rectangular window function is constant inside chosen interval and zero elsewhere. A rectangular window function $w(n)$ with N number of samples in the window is expressed as

$$w(n) = \begin{cases} 1\;;\; 0 \le n \le N-1 \\ 0\;;\; Otherwise \end{cases} \qquad (2)$$

To obtain the MUAP frames, the EMG signal is convolved with the rectangular window function. The l^{th} MUAP frame is expressed as

$$f_l(n) = y(n) * w(n) \qquad (3)$$

$$f_l(n) = \sum_{m=0}^{M-1} y(m)w(n-m) \qquad (4)$$

where, $y(n)$ is the band pass filtered EMG signal, l is the frame number and M is the total number of samples in the signal.

D. Dominant MUAP Selection

The energy of MUAP will be different in different groups i.e. normal, ALS and myopathy. In myopathy patients, the energy of MUAP is low and in ALS patients, the energy of MUAP is high when compared with the energy of MUAP in normal persons. Dominant MUAP is the region of EMG signal where the energy content is high. To determine the dominant MUAP the energy associated with each MUAP frame has to be evaluated. The energy associated with the l^{th} frame is expressed as

$$E_l = \sum_{n=0}^{N-1} [f_l(n)]^2 \qquad (5)$$

where, $f_l(n)$ is the MUAP in the l^{th} frame and$N$ is the total number of samples in the l^{th} frame. After the calculation of the energy associated with each frame, the frame with highest energy is chosen as the dominant MUAP. The dominant MUAP $d(n)$ is expressed as

$$d(n) = \max[E_1, E_2, \dots, E_l] \qquad (6)$$

E. DWT Feature Extraction

The Discrete Wavelet Transform (DWT) is performed on the dominant MUAP. The DWT decomposes the signal into coarse coefficients and detailed coefficients. The dominant MUAP signal is decomposed by passing it through a high pass filter $\varphi(n)$ and a low pass filter $\psi(n)$ simultaneously. The output of both low pass and high pass filters are then sub sampled by 2. This constitutes one level of DWT decomposition as shown in Fig. 2.

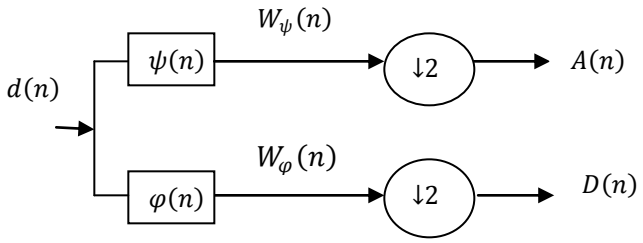

Fig.2 Level 1 DWT Implementation.

The output of low pass filter $W_\psi(n)$ is expressed as

$$W_\psi(n) = \sum_{k=0}^{N-1} d(k)\psi(n-k) \qquad (7)$$

where, $\psi(n)$ is the Daubechies Mother wavelet and is defined as

$$\psi(n) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{n-k2^j}{2^j}\right) \qquad (8)$$

where, j is the scale parameter, k is the shift parameter and j, k are integers. The output of high pass filter $W_\varphi(n)$ is expressed as

$$W_\varphi(n) = \sum_{k=0}^{N-1} d(k)\varphi(n-k) \qquad (9)$$

where, $\varphi(n)$ is the Daubechies Father wavelet and is defined as

$$\varphi(n) = (-1)^n \psi(N-1-n) \qquad (10)$$

After sub sampling by 2, the approximate coefficients $A(n)$ and detailed coefficients $D(n)$ are given by

$$A(n) = W_\psi(n) \qquad (11)$$

$$D(n) = W_\varphi(n) \qquad (12)$$

The dimension of the extracted wavelet coefficients is high. To reduce the dimensionality, statistics over the set of the wavelet coefficients are used. The statistical features include the maximum value of the wavelet coefficients in each sub band and the standard deviation of the wavelet coefficients in each sub band. The statistical features are expressed as follows:

$$A_{max} = max[A(n)] \qquad (13)$$

$$D_{max} = max[D(n)] \qquad (14)$$

$$A_{std} = \sqrt{\frac{1}{N-1}\left[\sum_{n=0}^{N-1}[A(n)]^2 - \frac{1}{N}\left(\sum_{n=0}^{N-1} A(n)\right)^2\right]} \qquad (15)$$

$$D_{std} = \sqrt{\frac{1}{N-1}\left[\sum_{n=0}^{N-1}[D(n)]^2 - \frac{1}{N}\left(\sum_{n=0}^{N-1} D(n)\right)^2\right]} \quad (16)$$

where, A_{max} , A_{std} are the maximum value and standard deviation of approximate coefficients respectively and D_{max} , D_{std} are the maximum value and standard deviation of detailed coefficients. The extracted statistical feature vectors are combined as a single feature vector. The combined feature vector or the test vector$x^{'}$ is expressed as

$$x^{'} = [A_{max}, A_{std}, D_{max}, D_{std}] \quad (17)$$

F. Classification

Classification is a process where the unknown labels of the test vectors are predicted. It requires two types of dataset namely, training and testing dataset. In training process, a classification model is formed using the training dataset. In testing process, the obtained classification model will be used to predict the unknown class label of the test dataset.

1) k-Nearest Neighbor Classifier: k-Nearest Neighbor (k-NN) classifier assigns the class label to the test vector that occurs most frequently among the k nearest training vectors. The steps involved in k-NN classifier are as follows:

Step 1: Choose the value of k. The square root of total number of training signals is taken as the k value.
Step 2: Determine the combined feature vector of all training signals (i.e., training vectors).
Step 3: Determine the distance between the test vector and all the training vectors using Euclidean distance measure. The Euclidean distance between two vectors $d(x^{'},x)$ is given by

$$d(x^{'},x) = \sqrt{\sum_{i=1}^{R}(x_i^{'} - x_i)^2} \quad (18)$$

where, $x^{'}$ is the test vector, x is the training vector and R is the number of features.
Step 4: Combine the class label along with the Euclidean distance of the training vector. i.e., $[d(x^{'},x)y]$, where, y is the class label of training vectors (y is 0 for ALS, 1 for myopathy and 2 for ALS).
Step 5: Sort the Euclidean distance along with the class labels.
Step 6: Assign the class label to the training vector that occurs most frequently among the k nearest training vectors.

2) Support Vector Machine Classifier: Support Vector Machine (SVM) classifier is a binary classifier that builds a hyper plane in multidimensional space using support vectors. Support vectors are training points closer to the decision hyper plane. Consider L training vectors each with R dimension. The class labels in SVM are represented by either +1 or -1. The decision hyper plane is defined as

$$w^t x + b = 0 \quad (19)$$

where, w is the weight vector and b is bias. The support hyper planes are defined as

$$w^t x + b = 1 \quad \text{for class 1} \quad (20)$$
$$w^t x + b = -1 \quad \text{for class 2} \quad (21)$$

The distance between two support hyper planes is called margin. There exist many hyper planes that separate the two classes. SVM maximizes the margin of each hyperplane until it hit the training points. The optimal hyper plane is chosen as the one with maximum margin. This is equivalent to the quadratic optimization problem expressed as follows:

$$\text{Minimize } \frac{\|w\|^2}{2} \text{ such that } y_i(x_i w + b) - 1 \geq 0 \ \forall i \quad (22)$$

This problem can be solved by forming Lagrangian multipliers $\alpha_i, i = 1,2, \ldots, L$for each constraint. The primary problem is of the form

$$L_p = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{L}\alpha_i y_i(x_i w + b) + \sum_{i=1}^{L}\alpha_i \quad (23)$$

Differentiate L_p with respect to w and b and equating it to zero results in

$$\frac{\partial L_p}{\partial w} = 0 \quad \Rightarrow w = \sum_{i=1}^{L}\alpha_i y_i x_i \quad (24)$$

$$\frac{\partial L_p}{\partial b} = 0 \quad \Rightarrow \sum_{i=1}^{L}\alpha_i y_i = 0 \quad (25)$$

Substituting the above results in L_p leads to dual problem of the form
Maximize

$$L_D = \sum_{i=1}^{L}\alpha_i - \frac{1}{2}\sum_{ij}\alpha_i\alpha_j y_i y_j x_i x_j \quad (26)$$

such that $\alpha_i \geq 0 \ \forall i, \sum_{i=1}^{L}\alpha_i y_i = 0$.
$\alpha_i$values are determined by solving equations (24) and (25). Training vector with $\alpha_i \neq 0$are considered as support vectors. The bias is given by

$$b = \frac{1}{N_s}\sum_{s\in S}\left(y_s - \sum_{m\in S}\alpha_m y_m x_m . x_s\right) \quad (27)$$

where, N_s is the total number of support vectors. A test vector is classified by evaluating

$$y^{'} = sgn(w^t x^{'} + b) \quad (28)$$

If $y^{'}$ is positive then $x^{'}$ belongs to class 1. If $y^{'}$ is negative then $x^{'}$ belongs to class 2.
SVM can only solve a two class problem to solve a three class problem multiclass classification schemes has to be used. The multiclass classification schemes include One – Against – All (OAA) and One – Against – One (OAO) scheme. In OAA scheme, N-1 classifiers are used where N is the number of classes and each classifier is trained with all the classes. In OAO scheme, N(N-1)/2 classifiers are used and each classifier is trained with only two classes.

III. RESULTS AND DISCUSSIONS

In the proposed method, a dataset with 750 recordings (250 signals from each type i.e., normal, myopathy and ALS) has been used. Fig. 3 shows an input signal corresponding to a myopathy patient. The total number of samples in the signal is 2,62,134 samples.Fig. 4 shows the steps involved in EMG signal preprocessing. First, the EMG signal is sampled by a

factor of two thus reducing the number of samples to 1,31,067.

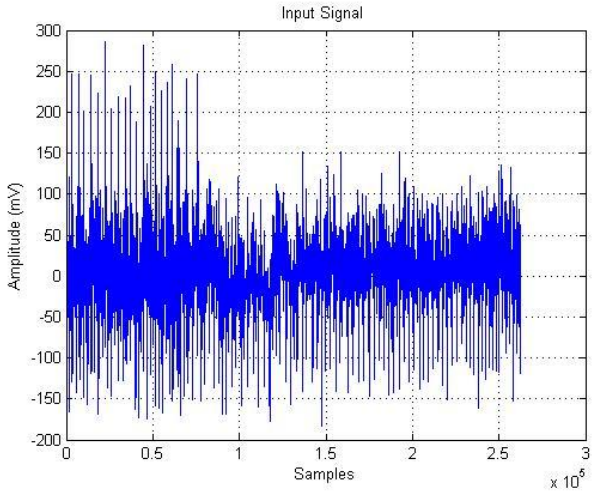

Fig.3 Input EMG Signal.

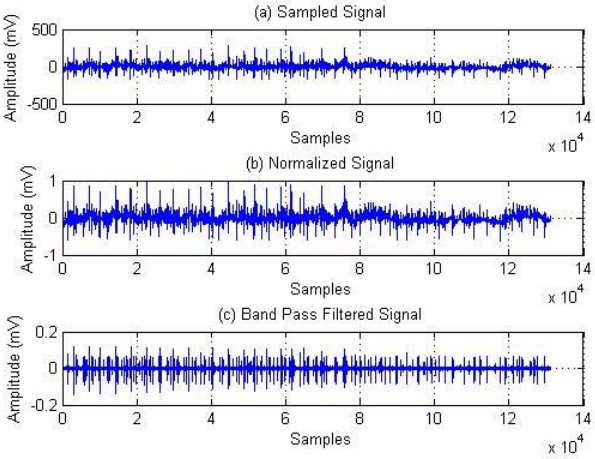

Fig.4 EMG Signal Preprocessing.

Second, normalization is performed by dividing the amplitude of every sample by the maximum amplitude of the entire EMG signal. Third, band pass filtering is done to remove the effect of noise.

The preprocessed EMG signal is decomposed into 131 MUAP frames with 1000 samples eachusing windowing technique as shown in Fig. 5. The energy associated with the 131 MUAP frames are computed and is shown in Fig. 6. Of these 131 frames, the frame with the highest energy is chosen as the dominant MUAP.

Next, the DWT is applied to the dominant MUAP. To perform DWT, the dominant MUAP is simultaneously low pass and high pass filtered. This reduces the total number of samples by half. The output of low pass and high pass filters are then down sampled by 2 to obtain the approximate and detailed coefficients respectively as shown in Fig. 7. This again reduces the total number of samples into half.

Statistical features over the wavelet coefficients are considered to reduce its dimensionality. The statistical features include A_{max}, A_{std}, D_{max} and D_{std}. The extracted features are

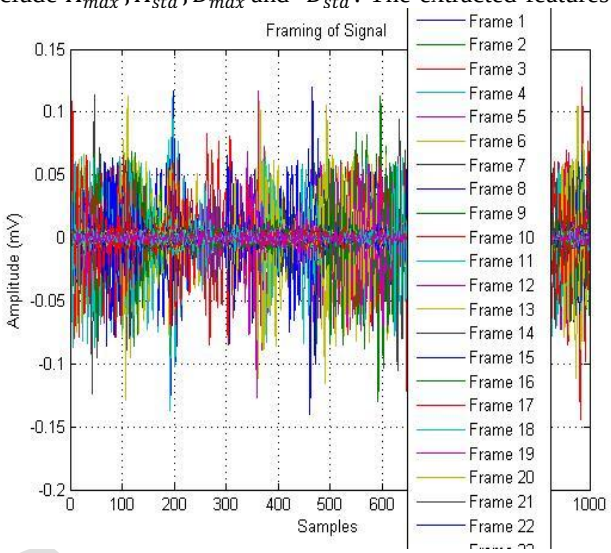

Fig.5 Windowed EMG Signal.

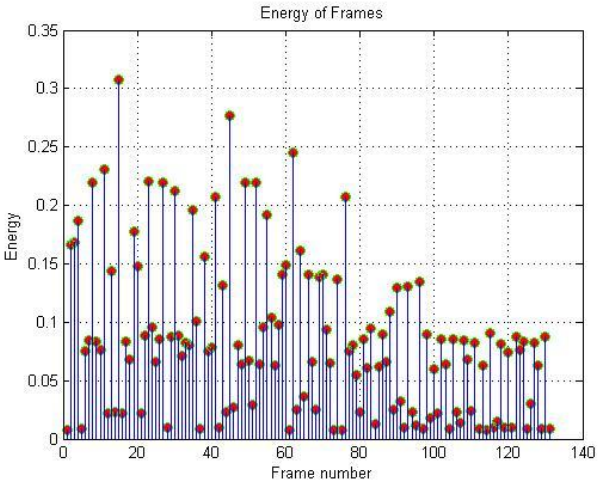

Fig.6 Energy associated with the frames.

combined into a single feature vector called the test vector. The extracted test vector is found to be

$$x' = [0.1551 \;\; 0.0242 \;\; 0.0335 \;\; 0.0055]$$

In the proposed method, first k-NN classifier (k = 24) is used to classify the neuromuscular diseases into normal, myopathy and ALS. It assigns a class label to the test vector that occurs most frequently among the k nearest training vectors. Second SVM classifier using OAA and OAO schemes is used to classify the neuromuscular diseases. The performance of the proposed method is measured using three parameters namely, specificity, sensitivity and total classification accuracy. The performance parameters are defined as follows:

$$\text{Specificity} = \frac{\text{Numberofcorrectly classifiednormalsignals}}{\text{Totalnumberofnormalsignals}} \quad (29)$$

$$\text{Sensitivity} = \frac{\text{Numberofcorrectlyclassified signalswithparticaldisease}}{\text{Totalnumberofsignals withparticulardisease}} \quad (30)$$

$$\text{ClassificationAccuracy} = \frac{\text{Totalnumberofcorrectly classifiedtestsignals}}{\text{Totalnumber oftestsignals}} \quad (31)$$

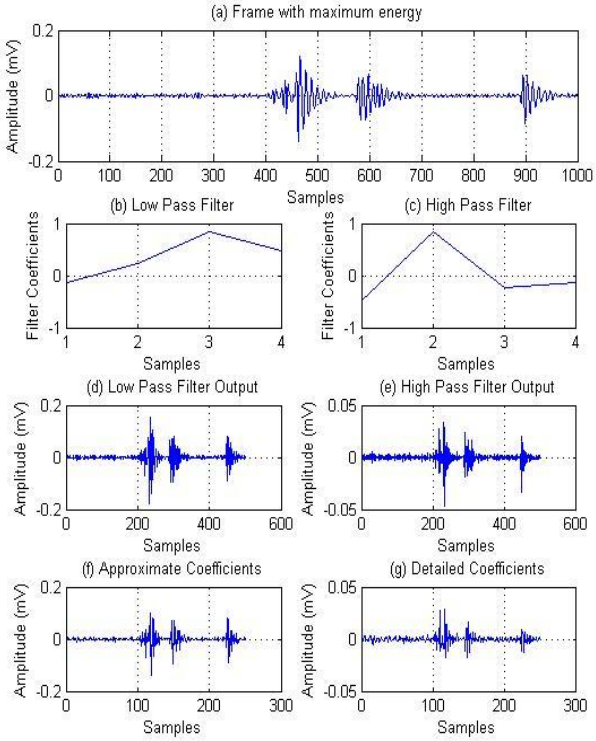

Fig.7 DWT on Dominant MUAP

Table I shows the performance parameters of k-NN and SVM classifiersfor 150 test recordings (50 signals from each category). Of the three classifiers the accuracy of SVM One – Against – One OAO scheme is higher compared with other classifiers.

TABLE I
PERFORMANCE PARAMETER OF k-NN AND SVM CLASSIFIERS.

Performance Parameter	k-NN	SVM OAA	SVM OAO
Sensitivity(ALS)	59 %	89 %	93 %
Sensitivity(Myopathy)	93 %	90 %	91 %
Specificity	92 %	81 %	88 %
Total Classification Accuracy	81.3 %	86.6 %	88.6 %

IV. CONCLUSION

In this work a classification scheme based on dominant motor unit action potential is proposed to classify the neuromuscular diseases into ALS, Myopathy and Normal. To accomplish this dominant motor unit action potential is found by using energy criteria. From dominant motor unit action potential discrete wavelet transform coefficients are obtained and certain statistical features like maximum value and standard deviation values are extracted. Using the extracted features the EMG signals are classified into ALS, Myopathy and Normal. For classification, initially k-Nearest Neighbor classifier is used. To improve the classification accuracy further Support Vector Machine classifier is used and has resulted in 88.6 % accuracy.

REFERENCES

[1]. Andrade A.O., Nasuto S., Kyberd P., Reed C.M.S. and Kanijn F.R.V. (2006), 'EMG signal filtering based on empirical mode decomposition', Biomedical Signal Processing and Control, Vol.1, pp.44-55.

[2]. Christodoulou C.I. and Pattichis C.S (1999), 'Unsupervised pattern recognition for the classification of EMG signals', IEEE Transactions on Biomedical Engineering, Vol.46, No.22, pp.169-178.

[3]. Devi B.V.A. and Priyadharsini S. (2013), 'Diagnosis of neuromuscular disorders using soft computing techniques', Vol. 3, Issue 5, pp.105-111.

[4]. Doborowolski A.P.WierzbowskiM. andTomczykiewicz K. (2010), 'Wavelet analysis for support vector machine classification of motor unit action potentials', IEEE, pp.4632-4635.

[5]. Englehart K., Hudgins D., Parker P.A. and Stevenson M. (1999), 'Classification of the myoelectric signal using time-frequency based representations', Medical Engineering and Physics, Vol.21, pp.431-438.

[6]. Fattah S.A. and Zhu W. (2013), 'Identification of Motor Neuron disease using wavelet domain features extracted from EMG signal', IEEE, pp.1309-1311.

[7]. Guler N.F. and Kocer S. (2005), 'Classification of EMG signals using PCA and FFT', Journal of Medical Systems, Vol. 29, No. 3, pp.241-250.

[8]. Katsis C.D., Exarchos T.P., Papaloukas C., Goletsis Y., Fotiadis D.I. and Sarmas I. (2007), 'A two-stage method for MUAP classification based on EMG decomposition', Computers in Biology and Medicine, Vol.37, pp.1232-1240.

[9]. Katsis C.D., Goletsis Y., Likas A., Fotiadis D.I. and Sarmas I. (2006), 'A novel method for automated EMG decomposition and MUAP classification ', Artificial Intelligence in Medicine, Vol.37, pp.55-64.

[10]. Kaur N. and Singh T. (2010), 'Statistical technique for classification of MUAPs for neuromuscular diseases', International Journal of Computer Applications, Vol. 1, No.2, pp.8-11.

[11]. Nikolic M. (2001), 'Detailed analysis of clinical electromyography signals EMG decomposition, findings and firing pattern analysis in controls and patients with myopathy and amyotrophic lateral sclerosis'.

[12]. Parsaei H. and Stashuk D.W. (2013), 'EMG signal decomposition using motor unit potential train validity', IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol.21, No.2, pp.265-274.

[13]. Parveen and Duhan M. (2012), 'Comparative study of different EMG signal decomposition techniques', International Journal of Engineering, Science and Metallurgy, Vol.2, No.3, pp.717-719.

[14]. Pattichis C.S. and Pattichis M.S. (1999), 'Time-Scale analysis of motor unit action potentials', IEEE Transactions on Biomedical Enginering, Vol. 46, No.11, pp.1320-1329.

[15]. Rasheed S., Stashuk D.W. and Kamel M.S. (2007), 'A hybrid classifier fusion approach for motor unit potential classification during EMG signal decomposition', IEEE Transactions on Biomedical Engineering, Vol.54, No.9, pp.1715-1721.
[16]. Ren X., Huang H. and Deng L. (2009), 'MUAP classification based on wavelet packet and fuzzy clustering technique', IEEE, pp.01-04.
[17]. Subasi A. (2012), 'Classification of EMG signals using combined features and soft computing techniques',Applied Soft Computing, Vol.12, pp.2188-2198.
[18]. Subasi A. (2012), 'Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines', Computers and Biology and Medicine, Vol.42, pp.806-815.
[19]. Subasi A. (2013), 'Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular diseases', Computers and Biology and Medicine, Vol.43, pp.576-586.
[20]. Subasi A., Yilmaz M. and Ozcalik H.R. (2006), 'Classification of EMG signal using wavelet neural network', Journal of Neuroscience Methods, Vol.156, pp.360-367.
[21]. Doulah A. B. S. M. U., Fattah S. A. and Zhu W. P. (2014), 'Wavelet domain feature extraction scheme based on dominant motor unit action potential of EMG signal for neuromuscular disease classification', Vol.8, No.2, pp.155-164.