A hybrid WD-EEMD sEMG feature extraction technique for lower limb activity recognition

Ankit Vijayvargiya, *Student Member, IEEE,* Vishu Gupta, *Member, IEEE,* Rajesh Kumar, *Senior Member, IEEE*
Nilanjan Dey, *Senior Member, IEEE* and João Manuel R. S. Tavares, *Member, IEEE*

***Abstract*—Classification and analysis of surface EMG (sEMG) signals have been of particular interest due to their numerous applications in the biomedical field. They can be used for the diagnosis of neuromuscular diseases, kinesiological studies, and human-machine interaction. However, these signals are difficult to process due to their noisy nature. To overcome this problem, a hybrid of wavelet with ensemble empirical mode decomposition pre-processing technique called WD-EEMD is proposed for classifying lower limb activities based on sEMG signals in healthy and knee abnormal subjects. First, Wavelet De-noising is used for filtering out white Gaussian Noise (WGN) and unwanted signals (contribution of other muscle signals). Next, an Ensemble Empirical Mode Decomposition is used for filtering out power line interference (PLI) and baseline wandering (BW) noises, followed by extraction of a total of nine time-domain features. Finally, the performance parameters of the Linear Discriminant Analysis (LDA) classifier are calculated with a 3-fold cross-validation technique. This study involves 11 healthy and 11 individuals with a knee abnormality for three different activities: walking, flexion of the leg up (standing), and leg extension from sitting position (sitting). Different pre-processing techniques similar to that of WD-EEMD were compared. It was observed that the proposed method achieves an average classification accuracy of 90.69% and 97.45% for healthy subjects and knee abnormal subjects, respectively.**

***Index Terms*—Biomedical signal analysis, EMG classification, WD-EEMD, Ensemble Empirical Mode Decomposition, Wavelet denoising, Linear Discriminant Analysis, Gait activities.**

I. Introduction

KNEE problems are defined as a sensation of discomfort in the knee that are caused by lack of proper warm-up, poor form during physical activities, or osteoarthritis. According to [1], one out of every four individuals have joint symptoms or arthritis because of an underlying condition such as degenerative arthritis of the knee. The knee joint is a synovial joint that is formed with several surrounding structures, including ligaments, bones, cartilage and tendons, to perform its functions [2]. Any external harm to any of these can result in knee abnormality [3]. Knee osteoarthritis, cerebral palsy are some knee abnormalities that cause knee pain and reduce the quality of daily life of a person [4], [5].

Ankit Vijayvargiya, Vishu Gupta and Rajesh Kumar are with the Department of Electrical Engineering, Malaviya National Institute of Technology Jaipur, 302017, India (e-mail: ankitvijayvargiya29@gmail.com; vishu.gupta0607@gmail.com; rkumar.ee@mnit.ac.in).

Nilanjan Dey is with the Department of Computer Science, JIS University, Kolkata, India (e-mail: neelanjan.dey@gmail.com).

João Manuel R. S. Tavares is with the Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal (e-mail: tavares@fe.up.pt).

Assistive devices can be used to enhance the quality of the daily life of an unhealthy person. These devices are categorized into: orthosis and prosthesis. The prosthesis is an artificial limb for a missing body part while orthosis is used to improve the functionality of moving body parts for weak person. These devices are also classified based on power: active and passive. An active assistive device uses a power source to activate the actuators while a passive device has no power source. So, automatic control is possible with active devices while passive devices cannot be similarly controlled. As usual, the active devices could be body-powered, or electric-powered. Electric-powered lower limb assistive devices have been widely used, and may be operated by a pressure resistor, strain gauge, micro-switch, electroencephalogram signals (EEG), electromyogram signals (EMG), etc. In recent years, EMG signals have widely been used for controlling assistive devices because it allows the recognition of movement in advance [6] and provides faster detection of the signal variation [7]. Invasive or non-invasive techniques are used to acquire EMG signals from muscles in which non-invasive techniques are better than invasive techniques as no medical supervision is required and infection is also negligible during placement of non-invasive (sEMG) electrodes [8]. In the non-invasive technique, electrical activity produced by skeletal muscle is collected through a sEMG sensor.

The design of a neuro-fuzzy controller has been proposed by Kiguchi et al. for the upper limb robotic exoskeleton, which is an upper limb assistive device [9]. The purpose of this exoskeleton is to enhance the quality of life of the injured, disabled, elderly, and physically challenged people. A sEMG based low-cost elbow joint-powered exoskeleton was developed for bicep brachii strength augmentation by Krasin et al. [10]. The user who worn this exoskeleton can freely move in normal condition; however, when the biceps muscle is in an underloaded condition, then this muscle produces a different EMG signal which causes the exoskeleton to automatically switch to the assistive lifting movement and then returns to its normal condition once the muscle is relaxed. A low-cost sEMG controlled upper limb prosthetic arm has been developed by Sharmila et al. [11]. For automatic control of the prosthetic arm, the sEMG signal is acquired for different hand movements from the users and features are extracted after preprocessing the data. Based on these features, the signal is classified for various hand movements. After recognizing the hand movements by the classifier, a control signal is generated and given to the motor of the prosthetic arm to perform the intended movements.

Over the past several years, applications of the upper limb using sEMG signal have been focused on by researchers compared to the applications of the lower limb, because acquiring the sEMG signals of lower limb muscles is more complex due to the contribution of multiple motor units at a time and their dependency on neuromuscular activity, physiological and anatomical properties of the involved muscles. Recognition of various gait activities based on the sEMG signal of the lower limb has an important role in controlling the exoskeleton for the knee abnormal person or in the prosthesis control for the lower limb amputee. Neural network based myopathy and neuropathy classification using sEMG signal was proposed in [13]. Kugler et al. have recognized Parkinson's disease using sEMG signal [14]. The classification of six different move-ments of lower limb using machine learning classifiers was studied by Khimraj et al. [12]. Vijayavargiya et.al. [15] worked on identifying knee abnormality in the subject, by collecting imbalanced surface EMG signals data as a result of different sized signal lengths, of healthy and unhealthy individuals. To realize multi-step classification, various schemes were adopted and employed a computational classifier for conclusive recog-nition. In other research [16], they used various machine learning algorithms to provide a comparative analysis between them. In this, walking, standing, and sitting three lower limb activities are observed and the movements are recorded in terms of sEMG signals for the purpose of classification of subjects with knee abnormality. Various different steps are performed on sEMG data to achieve classification. And five machine learning algorithms including Decision tree, Extra tree, KNN (k-nearest neighbor), SVM (support vector ma-chine), and Random forest to provide performance comparison in terms of accuracy, sensitivity, specificity, and F1-score to identify knee abnormality in unhealthy subjects. A. Gautam et al [45] introduced a novel classification approach to incor-porate lower limb activities accompanying prognostication of the knee joint angle. Convolutional Neural Network (CNN) and LSTM combinedly make architecture to classify lower limb activities where CNN is used for extracting features from sEMG signal data and LSTM is used for joint angle prediction and to interpret the features follow up dense layer is connected for classification. Combining these three blocks they have proposed MyoNet model to predict lower limb activities (out of walking, standing, and sitting) simultaneously with joint angle prediction.

Despite several successful applications, sEMG based knee activity recognition remains a challenging problem due to their noisy nature. Due to the mixing of different noise signals or artifacts such as inherent noise, ambient noise and motion artifacts, the identity of an actual sEMG signal originating in the muscle is lost. For this purpose, different methods have been proposed for sEMG noise elimination by the researchers. The frequency range of the sEMG signal is 10 to 500 Hz and an amplitude range of 0 to 10 mV [17]. The conventional filtering methods such as low-pass, high-pass, and band-pass filters can be used to remove the noises that are not in the range of the sEMG signals. However, they are unable to remove random noises such as white Gaussian noise that is in the range of active sEMG signal spectrum band. The frequency ranges of the motor unit in an sEMG signal can be represented by Wigner-Ville distribution (WVD) [18]. It exhibits excellent localization properties, but has a cross-term effect and thus, cannot deal with multi-component signals. Wavelets overcome the limitations caused by WVD. It does not have a cross-term and thus, has the capability of handling multi-resolution problems. Various adaptive filtering [19] techniques, like Wiener filtering [20], based on the Fourier approach have also been proposed for the removal of noise in surface myoelectric signals. As mentioned in many of earlier works [21], [22], the Wavelet Transform (WT) has been used in processing sEMG signals, as it is an extremely flexible approach to signal decomposition with a lot of choices in wavelet functions. The properties of the wavelet function and the characteristic of the signal to be analyzed need to be more carefully matched, before the classification process [21]. WT is also used for de-noising these signals by selecting an optimal wavelet function for them [23]. The use of wavelets has also gained widespread acceptance when extracting features or analyzing signals, especially for sEMG signals [24], [25]. They have advantages over classical techniques like Fourier transform or autoregressive models in analyzing physical situations where the signal contains discontinuities and sharp spikes [26], [27].

Studies on the decomposition of sEMG signals have been done since the 1960s [28], [29], where efforts have been made to segregate individual contributing motor potentials. A powerful technique called Empirical Mode Decomposition (EMD) was introduced [30] in 1998, which decomposed sEMG signals into Intrinsic Mode Functions (IMFs). EMD proved to be very effective and useful yet there were a few limitations associated with it. Improvements have been made in this algorithm, one of which is the Ensemble EMD (EEMD) [31].

Motivated by the need for neural control of the lower limb exoskeleton or prosthesis, investigated different types of mobility tasks that could be correctly identified using sEMG signals obtained from the leg muscles. The results of this study will help in the future development of neural-controlled artificial exoskeleton or lower limbs prosthesis with versatile activities for injured or disabled persons. Due to mixing of various noises in the sEMG signal, sEMG based knee activity recognition is a challenging problem. Therefore, in this article, a hybrid pre-processing technique WD-EEMD is proposed for the analysis of sEMG signal for the recognition of lower limb activities.

The major contributions of this research are:

1) Identification of the lower limb activity using the sEMG signals obtained from the leg muscles in individuals with and without knee abnormality.
2) A hybrid pre-processing technique WD-EEMD (Wavelet Denoising - Ensemble Empirical Mode Decomposition) which is proposed for the analysis of sEMG signals for lower limb activities recognition.
3) A total of nine time-domain features are extracted from the sEMG signals of the four lower limb muscles using the overlapping windowing technique.

4) Performance parameters of LDA classifier are calculated for the recognition of gait activities in healthy and knee abnormal subjects.

The structure of the article is as follows: a survey of data collection is given in section II. Section III presents the proposed methodology. Section IV comprises the results and discussion. Conclusion and future scope are given in Section V.

II. Data Acquisition

In this study, we have considered publicly available datasets from UCI for the classification of lower limb activity [32]. It consists of 22 volunteer participants above the age of 18 years, among which 11 participants appear to be fit and the rest exhibit knee abnormalities. No preceding case history was found regarding pain or injury in the knee of healthy participants. Among participants with knee abnormalities, four were affected with a meniscus injury, six suffered from anterior cruciate ligament (ACL), and one was encountered with sciatic nerve injury. The left leg for healthy subjects and affected limb for the knee abnormal subjects were chosen for acquiring the EMG signal. All the subjects undergo three different exercises: walking, flexion of the leg up (standing), and leg extension from sitting position (sitting) to analyze the behavior of the knee muscle. The data was recorded by four surface electrodes around the muscles: biceps femoris, rectus femoris, semitendinosus and vastus medialis, and the goniometer was attached to the external side of the knee joint. The used data collection equipment for obtaining the sEMG data was a MWX8 by Biometrics Ltd. that has 4 analog channels and 8 digital channels, out of which 4 for sEMG and 1 for goniometry were used. All the data was stored directly from MWX8 storage to the computer and transmitted to the Datalog software by the bluetooth adapter in real-time. The sampling frequency was of 1000 Hz and a 14-bit resolution was used. This study only focuses on the effect of EMG signals of the lower limb muscles during the lower limb activities, so only the sEMG signals are considered.

III. Methodology

This section provides a description of the methods used in building the proposed methodology.

A. Wavelet De-noising (WD)

To remove noise from the sEMG signal, a technique called the thresholding method is used based on wavelets. Wavelets can be visualized as small waves or ripples that have a very short and finite period. In the wavelet analysis method, first is selected a wavelet function that is called mother wavelet. Then, the low frequency version of wavelet is used to perform frequency analysis while the high frequency version of wavelet is used for temporal analysis [26]. Different mother wavelets are generated from a single basic wavelet $\psi(t)$, by scaling and translation [33]:

$$\psi_{s,\tau} = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right). \tag{1}$$

There are various kinds of mother wavelets and each has its own characteristics. Some of the most popular wavelet families are: Haar, Daubechies, Coiflet and Symlet.

When performing Discrete Wavelet Transform (DWT), wavelet coefficients can be generated by passing the signal through high-pass (detail coefficients) and low-pass (approximate coefficients) filters. The number of detail coefficients generated depends on the adopted level of decomposition.

After the wavelet decomposition of a signal, approximate and detailed coefficients are obtained. To eliminate the noise, small signal details can be excluded without any loss of important information. This, thresholding sets all coefficients to zero that are less than a particular threshold [14]. There are various modes for it such as soft and hard thresholding; however, both present their challenges. Garotte threshold function was proposed to overcome the shortcomings in the soft and hard thresholding methods. The universal threshold, that is used in this study, is defined as:

$$\lambda = \sigma\sqrt{2ln(N)}, \tag{2}$$

where σ = (MAD)/0.6745, with MAD referring to the Median Absolute Deviation of the wavelet coefficient and N is the length of signal.

As studied in [25], db7 from the Daubechies family is used till the fourth decomposition level in this work, and the decomposition is selected upto four level where one level of approximate coefficients and four levels of details coefficients are obtained. Garotte thresholding is applied on the second detail coefficient level (D2).

B. Ensemble Empirical Mode Decomposition

Due to the non-linearity and non-stationarity of sEMG signals, decomposition techniques that assume a process to be linear and stationary may yield deceptive results [34]. Empirical Mode Decomposition (EEMD) is a powerful tool for decomposing non-stationary and non-linear signals with complicated spatial and temporal structures into complete or almost orthogonal components, called Intrinsic Mode Functions [35]. IMF is a mono-component function or an oscillatory mode with one instantaneous frequency [36].

Using the EMD algorithm, a given signal $x(t)$ can be decomposed into a number of IMFs iteratively through a shifting algorithm. The procedure is as follows [30]:

1) An upper $u(t)$ and a lower $l(t)$ envelopes are created through interpolation (here, cubic) of all local maxima and minima of $x(t)$.
2) A running mean envelope $m(t)$ is then calculated using $m(t) = \frac{u(t)+l(t)}{2}$.
3) The mean envelope is then subtracted from the signal, which gives $k(t) = x(t) - m(t)$.
4) It is verified that whether $k(t)$ satisfies the following conditions of being an IMF:
   - The number of local extrema and zero crossings in the entire length of $k(t)$ must either be equal or at most differ by one;
   - At any point in the series, the mean value of $k(t)$ should be zero.

TABLE I: Extracted sEMG features and their mathematical formulation.

	Extracted Feature	Mathematical formulation
1	Mean Absolute Value (MAV)	$\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert$ where x_i is a sample of the sEMG signal
2	Root Mean Square (RMS)	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert^2}$
3	Zero Crossing (ZC)	$\sum_{i=1}^{N-1} f(x_i)$ where $f(x_i)=\begin{cases}1 & if,(x_i>0 \; and \; x_{i+1}<0) \\ & or\,(x_i<0 \; and \; x_{i+1}>0) \\ 0 & otherwise\end{cases}$
4	Slope Sign Change (SSC)	$\sum_{i=2}^{N-1} f(x_i)$ where $f(x_i)=\begin{cases}1 & if,if,(x_i>x_{i-1} \; and \; x_i>x_{i+1}) \\ & or\;(x_i<x_{i-1} \; and \; x_i<x_{i+1}) \\ 0 & otherwise\end{cases}$
5	Variance (VAR)	$\frac{1}{N-1}\sum_{i=1}^{N}x_i^2$
6	Difference Absolute Standard Deviation Value (DASDV)	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(x_{i+1}-x_i)^2}$
7	Average Amplitude Change (AAC)	$\frac{1}{N}\sum_{i=1}^{N-1}\lvert x_{i+1}-x_i\rvert$
8	Skewness (Skew)	$\frac{E[(x-\mu)^3]}{\sigma^3}$
9	Kurtosis (Kurt)	$\frac{E[(x-\mu)^4]}{\sigma^4}$ where σ is the Standard deviation of the signal dataset, μ = Mean of the dataset and E is the Expected value estimator of the dataset.

5) If $k(t)$ does not satisfy the conditions of being an IMF, $x(t)$ is replaced by $k(t)$ and sifting is continued, i.e., steps 1-4 are reiterated until the signal obtained satisfy the conditions. The sifting process can also be stopped if $k(t)$ is a monotonic function.

The original signal $x(t)$ may be obtained by summation of IMFs and the residual term:

$$x(t)=\sum_{m=1}^{M-1} IMF_m(t)+r_M(t) \tag{3}$$

where r_M is the residual term after extracting $M-1$ IMFs.

However, the EMD algorithm has a problem of frequent appearance mode mixing due to its sensitivity to noise [37]. To alleviate this problem, the noise-assisted data analysis method EEMD was proposed, which describes the IMFs as the average of an ensemble of trials [35].

The procedure of Ensemble EMD is as follows [35]:

1) An equally distributed and independent white noise series having the same standard deviation is added to the targeted data [37].

$$x_m(t)=x(t)+n_m(t) \tag{4}$$

where $n_m(t)$ is the m_{th} white noise series added to $x(t)$, and $x_m(t)$ is the noise-added signal.

2) The signal $x_m(t)$ is then decomposed using the EMD algorithm into P IMFs $h_{p,m}$, where P denotes the number of IMFs and $h_{p,m}$ is the p_{th} IMF of the m_{th} trial.
3) The above two steps are repeated M times, but with different white noises, where M indicates the number of ensembles.
4) The final IMFs are calculated by averaging each of the P IMFs over M trials.

$$IMF_P=\frac{1}{M}\sum_{1}^{M}h_{p,m}, \quad p=1,2,...,P \;\; m=1,2,...,M \tag{5}$$

The added noise cancels each other during averaging of corresponding IMFs in the EEMD process. The final IMFs remain inside the natural dyadic filter windows, hence decreasing the chances of mode-mixing [31].

Wavelet De-noising is used for filtering out white Gaussian Noise (WGN) and unwanted signals, e.g. contributions of other muscle signals. It also helps in preserving critical features. Ensemble Empirical Mode Decomposition is used for filtering out power line interference (PLI) and baseline wandering (BW) noises. It is also used here for decomposing the signal to extract relevant features from it. In this article, the preprocessing technique using Wavelet De-noising (WD) and Ensemble Empirical Mode Decomposition (EEMD) is referred to as WD-EEMD, and together, they form an excellent hybrid approach.

C. Feature Extraction

After the WD-EEMD analysis, a sliding windowing technique is used to extract the features rather than considering the entire signal at once, due to its stochastic nature [38]. A sliding window procedure, of adjacent or overlapped nature, is used for segmentation [39]. In [40], the results exemplify that the overlapped windowing approach outperforms the disjoint or adjacent windowing scheme on the basis of classification accuracy. Segmenting the data into short windows clinches constant local mean, which assures stationarity of the data during the feature extraction process [41]. The overlapped windowing approach was implemented in this study to divide each of the temporal series into optimal segments or subframes of 256ms time windows and a leap of 64ms [42].

After preprocessing and signal segmentation, a feature extraction stage is used to emphasize the relevant structures of sEMG signals. The features must be selected in such a way that condenses the suitable information and maximally separate the output classes. Time-domain features are normally employed for muscle activity detection, muscle contraction, and onset detection whereas, frequency domain features are used to detect neural abnormalities and muscle fatigue.

In this research, nine time-domain features: Mean Absolute Value, Root Mean Square, Zero Crossing, Slope Sign

Changes, Variance, Difference Absolute Standard Deviation Value, Skewness, and Kurtosis, are extracted from the three IMFs generated using the EEMD algorithm. As these features do not require any transformation and also due to their computational simplicity, they are generally quick and easily implemented for sEMG pattern recognition. A brief description of the features used is given in Table I.

D. Feature classification

As per the literature survey, many different machine learning models, like Support Vector Machine classifier and Decision Tree, have been used for sEMG classification; however, Linear Discriminant Classifier has been shown to be effective on applications to lower limb sEMG signal [42]. In this classifier, the discriminant property of LDA is enhanced during classification by maximizing the ratio of the between-class variance to the within-class variance. The within the class matrix (S_w) and between-class matrix (S_b) can formally be defined as:

$$S_w = \sum_{k=1}^{C}\sum_{i=1}^{N_k}(f_i^k - \mu_k)(f_i^k - \mu_k)^T, \tag{6}$$

$$S_b = \sum_{k=1}^{C}(\mu_k - \mu)(\mu_k - \mu)^T, \tag{7}$$

where f_i^k is the i^{th} sample of class k, μ_k is the mean of class k, C is the number of classes, N_k is the number of samples in class k, and μ is the mean of all classes.

E. Performance Evaluation Metrics

To analyze the performance of machine learning models true positive rate, false positive rate, true negative rate, false negative rate are estimated, and their outcome helps in the building of confusion metrics. A confusion matrix facilitates the visualization of the performance of the model on the test dataset. In this study, we have three different classes: walking (W), sitting (S) and standing (T). The confusion matrix is formed as:

$$C = \begin{bmatrix} C_{WW} & C_{WS} & C_{WT} \\ C_{SW} & C_{SS} & C_{ST} \\ C_{TW} & C_{TS} & C_{TT}, \end{bmatrix} \tag{8}$$

where C_{WW} is the number of cases in walking class predicted as walking, C_{WS} is the number of cases in walking class predicted as Sitting, C_{WT} is the number of cases in walking class predicted as standing and others can be defined similarly.

Total number of data points = $C_{WW}+C_{WS}+C_{WT}+C_{SW}+C_{SS}+C_{ST}+C_{TW}+C_{TS}+C_{TT}$

Total number of cases as predicted walking (P_W) = $C_{WW}+C_{SW}+C_{TW}$

Total number of cases as actual walking (A_W) = $C_{WW}+C_{WS}+C_{WT}$

Total number of cases as predicted sitting (P_S) = $C_{WS}+C_{SS}+C_{TS}$

Total number of cases as actual sitting (A_S) = $C_{SW}+C_{SS}+C_{ST}$

Total number of cases as predicted standing (P_T) = $C_{WT}+C_{ST}+C_{TT}$

Total number of cases as actual standing (A_T) = $C_{TW}+C_{TS}+C_{TT}$

The performance parameters for three class dataset are:

Accuracy - defined as the ratio of all correct predictions to the total number of instances in the dataset:

$$\textbf{Acc=}\ \frac{C_{WW}+C_{SS}+C_{TT}}{Total\ number\ of\ dataset}. \tag{9}$$

Specificity - the ratio of correct negative prediction to the total number of actual negative instances in the dataset.

Specificity for walking class:

$$SP_W = \frac{C_{SS}+C_{ST}+C_{TS}+C_{TT}}{A_S+A_T}. \tag{10}$$

Specificity for sitting class:

$$SP_S = \frac{C_{WW}+C_{WT}+C_{TW}+C_{TT}}{A_W+A_T}. \tag{11}$$

Specificity for standing class:

$$SP_T = \frac{C_{WW}+C_{WS}+C_{SW}+C_{SS}}{A_W+A_S}. \tag{12}$$

Precision - gives the ratio of correct positive prediction to the total number of predicted positive instances in the dataset:

$$PR_i = \frac{C_{ii}}{P_i}. \tag{13}$$

Sensitivity (Recall) - the ratio of correct positive prediction to the total number of actual positive instances in the dataset:

$$RC_i = \frac{C_{ii}}{A_i}. \tag{14}$$

F-Score - measures the balance between precision and recall and is equal to harmonic mean of precision and recall:

$$F_i = \frac{2*RC_i*PR_i}{RC_i+PR_i}, \tag{15}$$

where i $\in$ {W, S, T }.

IV. Result and Discussion

The current study is based on the application of WD-EEMD based classification of lower limb movement using sEMG signals. A total of 11 healthy and 11 subjects suffering from knee abnormality were considered and the performance parameters for the classification of three different movements: walking, sitting and standing, were computed. The activity signal of the subjects was pre-processed using WD-EEMD and then divided into training and testing sets according to the 3-fold cross-validation technique. An overlapping window with a window size of 256 msec and 25% overlapping was chosen for the segmentation of the signal. Then, the LDA classifier was trained with the training dataset and the performance parameters of the model with the testing dataset calculated.

The K-fold cross-validation methodology is a resampling strategy that utilizes a compelled information test to assess the performance of AI models. In this technique, the samples are divided into groups of k equal size. After that, the training of the model is performed with $k-1$ groups of samples and the

TABLE II: Subject-wise Mean accuracy of the LDA classifier in percentage with the 3-fold cross validation technique for healthy individuals during walking, sitting and standing, with different pre-processing techniques (best values found are in bold).

Subject	Wavelet			EMD			WD-EEMD		
	Walking	Sitting	Standing	Walking	Sitting	Standing	Walking	Sitting	Standing
S1	96.43	87.84	91.03	82.14	52.70	61.54	85.71	97.30	97.44
S2	91.89	94.59	93.59	64.86	62.16	70.51	89.19	91.89	100.00
S3	71.43	68.75	91.86	45.71	43.75	46.51	68.57	37.50	72.09
S4	76.32	78.57	75.00	52.63	60.00	62.50	68.42	92.86	93.75
S5	93.94	94.87	92.75	66.67	71.79	81.16	93.94	82.05	92.75
S6	67.57	70.89	92.38	81.08	64.56	85.71	100.00	96.20	100.00
S7	97.56	92.68	94.78	95.12	90.24	97.39	95.12	100.00	95.65
S8	68.09	88.30	92.74	82.98	90.43	90.32	97.87	100.00	95.97
S9	65.38	100.00	99.08	69.23	100.00	78.41	61.54	98.31	95.41
S10	86.21	97.38	96.69	89.66	97.91	96.03	100.00	100.00	94.70
S11	93.10	88.89	96.51	62.07	53.70	79.07	75.86	79.63	90.70
Mean	82.54	87.52	92.40	72.01	71.57	77.20	**85.11**	**88.70**	**93.50**

TABLE III: Subject-wise Mean accuracy of the LDA classifier in percentage with the 3-fold cross validation technique for knee abnormal individuals during walking, sitting and standing, with different pre-processing techniques (best values found are in bold).

Subject	Wavelet			EMD			WD-EEMD		
	Walking	Sitting	Standing	Walking	Sitting	Standing	Walking	Sitting	Standing
S1	84.84	96.55	96.51	60.60	75.86	83.72	98.48	87.93	98.83
S2	94.64	89.23	91.42	82.14	78.46	60.00	98.21	100	82.85
S3	96.90	84.21	98.01	96.90	90.35	99.00	100	91.22	100
S4	96.58	95.37	97.42	95.90	97.10	99.35	98.63	100	96.67
S5	95.79	93.06	100	96.63	98.02	99.23	99.16	99.00	100
S6	87.22	75.67	88.61	84.59	85.84	91.39	99.62	100	98.61
S7	97.5	84.81	91.67	86.21	88.61	90.28	100	91.14	95.83
S8	73.08	84.70	85.52	48.08	71.76	64.47	96.15	94.12	100
S9	86.14	67.86	78.13	85.54	76.78	87.5	100	99.10	97.91
S10	90.68	94.19	84.16	92.37	96.51	90.10	97.22	97.67	96.03
S11	94.37	86.00	98.46	95.07	75.00	89.23	100	100	97.69
Mean	90.70	86.51	91.81	84.00	84.91	86.75	**98.86**	**96.38**	**96.77**

testing is performed with k^{th} group of samples. This process is repeated for all the groups obtained from the input data.

Table II presents the subject-wise mean of the correct classification percentage for the three movements under study obtained from the sEMG data acquired from healthy subjects, whereas the performance concerning the knee abnormal subjects is given in Table III. The similar pre-processing techniques can be compared as to with and without knee abnormal subjects from the data in these tables, which confirms that WD-EEMD performed significantly better than the Wavelet Transform or EMD when apllied individually.

Tables IV and V allow the comparison in terms of performance indices between the different pre-processing techniques under studied when applied to subjects with and without knee abnormalities. This comparison allows to conclude that the WD-EEMD pre-processing technique obtained the highest performance indices relatively to other techniques.

Many different methods have been proposed for lower limb activity recognition. Herrera-Gonzalez et al. have developed a classifier for the classification of three different exercises using MP-ANN with an accuracy of 88%, 94% and 92% for walking, sitting and standing tasks, respectively Table VI [43]. On the other hand, Zhang et al. have classified different lower limb movements of healthy subjects by using the Empirical Mode Decomposition based approach obtaining the results given in Table VI [44].

TABLE VI: Comparison of the performance obtained by the proposed methodology against the ones obtained by literature studies with same dataset.

Approach	Subject	Walking	Sitting	Standing
EMD [44]	Healthy	0.64	0.67	0.69
MEMD [44]	Healthy	0.73	0.79	0.82
NA-EMD [44]	Healthy	0.79	0.83	0.83
MP-ANN [43]	Knee Abnormal	88	94	92
Transfer Learning based LRCN [45]	Healthy	98.2	97.7	98.4
	Knee Abnormal	92.8	92.3	92.2
ICA-EBM [42]	Healthy	96.0	96.2	96.2
	Knee Abnormal	86.6	86.4	85.5
Proposed Method	Healthy	85.11	88.70	93.50
	Knee Abnormal	**98.86**	**96.38**	**96.77**

Naik et al. developed a classifier to classify the walking, sitting and standing activities with an accuracy of 96.14 and 86.17% for healthy subjects and subjects suffering from knee abnormalities, respectively [42]. Gautam and collaborators introduced the transfer learning-based LRCN model to classify the walking, sitting and standing activities obtaining an accuracy of 98.2, 97.7 and 98.4% for healthy subjects

TABLE IV: Subject-wise performance indices of the LDA classifier in percentage with the 3-fold cross validation technique for healthy individuals during walking, sitting and standing, with different pre-processing techniques (best values found are in bold).

Subject	Wavelet				EMD				WD-EEMD			
	Accuracy	Specificity	Sensitivity	F-Score	Accuracy	Specificity	Sensitivity	F-Score	Accuracy	Specificity	Sensitivity	F-Score
S1	92.86	96.19	92.64	92.93	74.29	86.64	73.40	73.71	95.71	97.75	95.08	95.41
S2	91.62	95.66	91.77	92.02	72.77	86.17	73.53	73.38	93.19	96.56	93.69	93.54
S3	94.64	97.36	93.04	93.34	96.30	98.14	95.42	95.66	98.15	99.02	97.08	97.66
S4	96.46	98.21	96.46	96.30	97.10	98.53	97.46	97.20	98.55	99.34	98.47	98.33
S5	96.38	98.21	96.29	95.92	97.65	98.92	97.96	97.43	99.36	99.71	99.39	99.31
S6	84.74	92.26	83.83	83.95	87.79	93.70	87.27	87.47	99.30	99.68	99.41	99.27
S7	91.34	95.74	91.33	91.24	88.52	93.98	88.36	88.97	95.67	97.87	95.66	95.58
S8	82.16	90.52	81.10	82.14	63.38	80.93	61.44	62.04	96.71	98.28	96.76	96.88
S9	78.61	88.36	77.38	78.34	83.42	91.32	83.28	83.63	99.20	99.52	99.01	99.20
S10	89.51	94.55	89.67	89.82	92.79	96.37	92.99	92.84	96.95	98.49	96.98	96.92
S11	93.55	96.74	92.94	93.13	87.63	93.77	86.43	86.67	99.19	99.57	99.23	99.26
Mean	90.17	94.89	89.68	89.92	85.60	92.59	85.23	85.36	**97.45**	**98.71**	**97.34**	**97.40**

TABLE V: Subject-wise performance indices of the LDA classifier in percentage with the 3-fold cross validation technique for knee abnormal individuals during walking, sitting and standing, with different pre-processing techniques(best values found are in bold).

Subject	Wavelet				EMD				WD-EEMD			
	Accuracy	Specificity	Sensitivity	F-Score	Accuracy	Specificity	Sensitivity	F-Score	Accuracy	Specificity	Sensitivity	F-Score
S1	90.56	94.97	91.76	90.72	61.11	79.30	65.46	62.18	95.56	97.83	93.48	93.50
S2	93.65	96.48	93.36	94.02	66.14	82.20	65.85	65.70	94.71	97.19	93.69	94.12
S3	81.07	89.45	77.35	78.42	45.56	72.22	45.33	44.56	61.54	79.64	59.39	57.72
S4	76.60	87.63	76.63	76.87	59.57	78.60	58.38	58.70	88.30	94.24	85.01	85.40
S5	93.89	96.93	93.85	93.21	74.44	86.94	73.21	72.46	88.33	94.74	89.58	86.94
S6	80.54	90.29	76.94	76.15	77.38	89.19	77.12	74.28	98.64	99.35	98.73	98.32
S7	94.54	96.82	95.01	95.27	94.54	96.76	94.25	94.91	97.06	98.64	96.92	96.46
S8	86.79	92.68	83.04	84.47	89.06	94.48	87.91	87.39	97.74	98.96	97.95	97.40
S9	96.05	97.77	88.16	90.95	88.36	94.36	82.55	81.25	93.28	96.51	85.09	86.23
S10	96.23	97.64	93.43	95.17	96.50	97.85	94.53	95.45	97.84	99.04	98.23	95.79
S11	93.49	95.97	92.83	93.61	68.05	81.74	64.95	65.82	84.62	91.61	82.06	82.62
Mean	89.40	94.24	87.49	88.08	74.61	86.69	73.59	72.97	**90.69**	**95.25**	**89.10**	**88.59**

and 92.8, 92.3 and 92.2% for individual suffering from knee abnormalities, respectively [45]. The sEMG data of lower limb muscles that we have considered here is the same as the other contributors. Table VI allows the comparative performance analysis between the proposed model against literature studies, which allows to conclude that the proposed WD-EMD based pre-processing technique gave high performance for lower limb activity recognition in knee abnormal subjects while in healthy subjects other techniques gave better results than the WD-EMD. The controlling of lower limb assistive devices, for example, is required for individuals with knee abnormality, so the proposed technique seems to be better than other literature methods.

V. Conclusion

In this article, a hybrid pre-processing technique called Wavelet Denoising - Ensemble Empirical Mode Decomposition (WD-EEMD) was proposed for the analysis of sEMG signals for recognition of lower limb activity in subjects with and without knee abnormality. Both WD and EEMD filter different types of noises commonly associated with the EGG signal, and hence provide an integrated approach to de-noising and decomposition the inout signal. Another advantage of using EEMD is to decompose the signal into several IMFs to assist with the signal segmentation and feature extraction phases using the overlapping windowing technique. The results were compared with the ones obtained by similar pre-processing techniques with the hybrid approach proving to be superior than them. For performance evaluation, 3-fold cross-validation was implemented on the used dataset, and the proposed method achieved an average classification accuracy of 90.69 and 97.45% for healthy subjects and knee abnormal subjects, respectively.

There are still some extensions in the future for the proposed work. First, the used dataset includes data acquired with a relative low number of subjects. Hence, the proposed approach should be validated with a large number of the subject, which will reduce the biasing issue due to the use of a small dataset. Second, the proposed methodology was validated using an offline dataset, and so, further research can aim its validation using a real-time dataset for its clinical validation. Additionally, the other advanced machine learning algorithm can be implemented and one may also try to reduce the extracted features space by using feature reduction techniques.

Acknowledgment

This publication is supported by Visvesvaraya PhD Scheme, Meity, Govt. of India, MEITY-PHD-2942.

Compliance with ethical standards

Ethical approval: This article does not contain any study with human participants performed by any of the authors. Declaration of Competing Interest: The authors report no declarations of interest.

References

[1] A. Foundation, "Arthritis By The Numbers," https://www.arthritis.org/getmedia/e1256607-fa87-4593-aa8a-8db4f291072a/2019-abtn-final-march-2019.pdf, 2019.

[2] Y. Sun, E. Teo, and Q. Zhang, “Discussions of knee joint segmentation,” in *International Conference on Biomedical and Pharmaceutical Engineering*. IEEE, 2006.

[3] P. Richebé, X. Capdevila, and C. Rivat, “Persistent postsurgical painpathophysiology and preventative pharmacologic considerations,” *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 129, no. 3, 2018.

[4] A. Vijayvargiya, P. L. Singh, S. M. Verma, R. Kumar, and S. Bansal, “Performance comparison analysis of different classifier for early detection of knee osteoarthritis,” in *Sensors for Health Monitoring*. Elsevier, 2019.

[5] M. Oskoui, F. Coutinho, J. Dykeman, N. Jette, and T. Pringsheim, “An update on the prevalence of cerebral palsy: a systematic review and meta-analysis,” *Developmental Medicine & Child Neurology*, vol. 55, no. 6, 2013.

[6] B.-S. Yang and S.-T. Liao, “Fall detecting using inertial and electromyographic sensors,” in *Proceedings of the 36th Annual Meeting of the American Society of Biomechanics, Gainsville, FL, USA*, 2012.

[7] J. Cheng, X. Chen, and M. Shen, “A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals,” *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, 2012.

[8] D. Farina and F. Negro, “Accessing the neural drive to muscle and translation to neurorehabilitation technologies,” *IEEE Reviews in Biomedical Engineering*, vol. 5, 2012.

[9] K. Kiguchi, T. Tanaka, and T. Fukuda, “Neuro-fuzzy control of a robotic exoskeleton with emg signals,” *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 4, 2004.

[10] V. Krasin, V. Gandhi, Z. Yang, and M. Karamanoglu, “Emg based elbow joint powered exoskeleton for biceps brachii strength augmentation,” in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015.

[11] K. Sharmila, T. Sarath, and K. Ramachandran, “Emg controlled low cost prosthetic arm,” in *IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics*. IEEE, 2016.

[12] P. K. Shukla, A. Vijayvargiya, R. Kumar *et al.*, “Human activity recognition using accelerometer and gyroscope data from smartphones,” in *International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)*. IEEE, 2020.

[13] R. Swaroop, M. Kaur, P. Suresh, and P. K. Sadhu, “Classification of myopathy and neuropathy emg signals using neural network,” in *International Conference on Circuit, Power and Computing Technologies (ICCPCT)*. IEEE, 2017.

[14] P. Kugler, C. Jaremenko, J. Schlachetzki, J. Winkler, J. Klucken, and B. Eskofier, “Automatic recognition of parkinson’s disease using surface electromyography during standardized gait tests,” in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013.

[15] A. Vijayvargiya, C. Prakash, R. Kumar, S. Bansal, and J. M. R. Tavares, “Human knee abnormality detection from imbalanced semg data,” *Biomedical Signal Processing and Control*, vol. 66, 2021.

[16] A. Vijayvargiya, R. Kumar, N. Dey, and J. M. R. Tavares, “Comparative analysis of machine learning techniques for the classification of knee abnormality,” in *IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*. IEEE, 2020.

[17] N. Nazmi, A. Rahman, M. Azizi, S.-I. Yamamoto, S. A. Ahmad, H. Zamzuri, and S. A. Mazlan, “A review of classification techniques of emg signals during isotonic and isometric contractions,” *Sensors*, vol. 16, no. 8, 2016.

[18] A. L. Ricamato, R. G. Absher, M. T. Moffroid, and J. P. Tranowski, “A time-frequency approach to evaluate electromyographic recordings,” in *Proceedings Fifth Annual IEEE Symposium on Computer-Based Medical Systems*. IEEE, 1992.

[19] T. W. Beck, J. M. DeFreitas, J. T. Cramer, and J. R. Stout, “A comparison of adaptive and notch filtering for removing electromagnetic noise from monopolar surface electromyographic signals,” *Physiological Measurement*, vol. 30, no. 4, 2009.

[20] G. Aschero and P. Gizdulich, “Denoising of surface emg with a modified wiener filtering approach,” *Journal of Electromyography and Kinesiology*, vol. 20, no. 2, 2010.

[21] D. K. Kumar, N. D. Pah, and A. Bradley, “Wavelet analysis of surface electromyography,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 4, 2003.

[22] R. Constable and R. Thornhill, “Using the discrete wavelet transform for time-frequency analysis of the surface emg signal.” *Biomedical Sciences Instrumentation*, vol. 29, 1993.

[23] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, “An optimal wavelet function based on wavelet denoising for multifunction myoelectric control,” in *6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, vol. 2. IEEE, 2009.

[24] K. Englehart, B. Hudgins, P. A. Parker, and M. Stevenson, “Classification of the myoelectric signal using time-frequency based representations,” *Medical Engineering & Physics*, vol. 21, no. 6-7, 1999.

[25] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, “Application of wavelet analysis in emg feature extraction for pattern classification,” *Measurement Science Review*, vol. 11, no. 2, 2011.

[26] A. Graps, “An introduction to wavelets,” *IEEE Computational Science and Engineering*, vol. 2, no. 2, 1995.

[27] M.-F. Lucas, A. Gaufriau, S. Pascual, C. Doncarli, and D. Farina, “Multichannel surface emg classification using support vector machines and signal-based wavelet optimization,” *Biomedical Signal Processing and Control*, vol. 3, no. 2, 2008.

[28] G. Gerstein and W. Clark, “Simultaneous studies of firing patterns in several neurons,” *Science*, vol. 143, no. 3612, 1964.

[29] D. G. Keehn, “An iterative spike separation technique,” *IEEE Transactions on Biomedical Engineering*, no. 1, 1966.

[30] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, 1998.

[31] Z. Wu and N. E. Huang, “Ensemble empirical mode decomposition: a noise-assisted data analysis method,” *Advances in Adaptive Data Analysis*, vol. 1, no. 01, 2009.

[32] O. Sanchez, J. Sotelo, M. Gonzales, and G. Hernandez, “Emg dataset in lower limb data set,” *UCI Machine Learning Repository*, vol. 2, 2014.

[33] V. Clemens, “A really friendly guide to wavelets,” http://agl.cs.unm.edu/~williams/cs530/arfgtw.pdf, 1999.

[34] C. Sapsanis, G. Georgoulas, and A. Tzes, “Emg based classification of basic hand movements based on time-frequency features,” in *21st Mediterranean Conference on Control and Automation*. IEEE, 2013.

[35] Y. Lei, Z. He, and Y. Zi, “Eemd method and wnn for fault diagnosis of locomotive roller bearings,” *Expert Systems with Applications*, vol. 38, no. 6, 2011.

[36] S. Gaci, “A new ensemble empirical mode decomposition (eemd) denoising method for seismic signals,” *Energy Procedia*, vol. 97, 2016.

[37] G. R. Naik, S. E. Selvan, and H. T. Nguyen, “Single-channel emg classification with ensemble-empirical-mode-decomposition-based ica for diagnosing neuromuscular disorders,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 7, 2015.

[38] F. D. Farfán, J. C. Politti, and C. J. Felice, “Evaluation of emg processing techniques using information theory,” *Biomedical Engineering Online*, vol. 9, no. 1, 2010.

[39] R. N. Khushaba, S. Kodagoda, M. Takruri, and G. Dissanayake, “Toward improved control of prosthetic fingers using surface electromyogram (emg) signals,” *Expert Systems with Applications*, vol. 39, no. 12, 2012.

[40] K. Englehart and B. Hudgins, “A robust, real-time control scheme for multifunction myoelectric control,” *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 7, 2003.

[41] S. A. Fattah, O. Iqbal, S. Zahin, C. Shahnaz, and G. Rosul, “Basic hand action classification based on surface emg using autoregressive reflection coefficient,” in *IEEE Region 10 Conference (TENCON)*. IEEE, 2017.

[42] G. R. Naik, S. E. Selvan, S. P. Arjunan, A. Acharyya, D. K. Kumar, A. Ramanujam, and H. T. Nguyen, “An ica-ebm-based semg classifier for recognizing lower limb movements in individuals with and without knee pathology,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 3, 2018.

[43] M. Herrera-González, G. A. Martínez-Hernández, J. L. Rodríguez-Sotelo, and Ó. F. Avilés-Sánchez, “Knee functional state classification using surface electromyographic and goniometric signals by means of artificial neural networks,” *Ingeniería y Universidad*, vol. 19, no. 1, 2015.

[44] Y. Zhang, P. Xu, P. Li, K. Duan, Y. Wen, Q. Yang, T. Zhang, and D. Yao, “Noise-assisted multivariate empirical mode decomposition for multichannel emg signals,” *Biomedical Engineering Online*, vol. 16, no. 1, 2017.

[45] A. Gautam, M. Panwar, D. Biswas, and A. Acharyya, “Myonet: A transfer-learning-based lrcn for lower limb movement recognition and knee joint angle prediction for remote monitoring of rehabilitation progress from semg,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, 2020.