2024 International Conference on Brain Computer Interface & Healthcare Technologies (iCon-BCIHT) | 979-8-3315-4006-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICON-BCIHT63907.2024.10882404

One-class SVM using Domain-based sEMG Features for Detection of Neuromuscular Disorders

Vaidehi Bhat
Dept. of Electrical Engineering, IIT Indore
Indore, India
ee230002077@iiti.ac.in

Varad Gaekwad
Dept. of Electrical Engineering, IIT Indore
Indore, India
ee230002081@iiti.ac.in

Advay Kunte
Dept. of Electrical Engineering, IIT Indore
Indore, India
ee230002007@iiti.ac.in

Nambiar Anand Sreenivasan
Dept. of Mechanical Engineering, IIT Indore
Indore, India
me230003046@iiti.ac.in

Soham Pandit
Dept. of Mechanical Engineering, IIT Indore
Indore, India
me220003080@iiti.ac.in

Hitendra Kumar
Dept. of Biosciences and Biomedical Engineering, IIT Indore, Indore, India
hitendra@iiti.ac.in

Vibhor Pandhare
Dept. of Mechanical Engineering, IIT Indore
Indore, India
vibhorpandhare@iiti.ac.in

***Abstract*— Neuromuscular Disorders (NMDs) are a group of conditions that affect nerves and muscles leading to muscle weakness, atrophy, and sometimes paralysis. The effects of these conditions cannot be reversed, causing permanent impairment if not diagnosed timely. Recent medical advancements have led to timely detection and successful prevention of many diseases with portable healthcare solutions backed by strong machine learning algorithms. A similar approach for NMDs can be designed using non-invasive Surface Electromyography (sEMG) signals. While machine learning approaches have been developed for detecting NMDs, generalized features and a lack of comprehensive datasets prevent the evaluation of the robustness of these methods. Thus, a systematic methodology that incorporates features representing NMD biomarkers and one-class support vector machine for unsupervised detection of neuromuscular disorders is proposed. The method is implemented on electromyography analysis of Human Activity – Database 1 (EMAHA DB-1) which is a comprehensive dataset for hand gesture recognition based on Indian population. For robust evaluation of the algorithm, functions are designed to simulate anomalies based on NMD bio-signals' knowledge by generating an augmented dataset that mimics the nature of NMDs. Multiple machine learning models are implemented and evaluated on the real healthy dataset as well as simulated NMDs dataset. One-class SVM emerged successful amongst other models with an accuracy of 81.45%, averaged over separately trained models for each limb action.**

Keywords — Machine Learning, Neuromuscular Disorders, Surface Electromyography, Detection, Healthcare

I. Introduction

Portable Medical Devices (PMDs) are small, portable (mobile, untethered) wireless computing devices, that enable people to monitor their health and wellbeing virtually at any time in any given context [1, 2]. Devices are advancing from portable to wearable and embedded forms, allowing real-time monitoring of health conditions. These innovations are anticipated to reduce healthcare costs by enhancing decision-making, prevention, treatment adherence, and decreasing hospital visits and administrative expenses. Ever since the pacemaker, first used in 1959, there have been a multitude of dedicated portable medical devices [3]. Today, health-related apps for mobile devices are rapidly expanding, covering nearly every aspect of human healthcare. Additionally, peripheral products that attach to mobile devices are enhancing functionality beyond what the original devices offer. Artificial Intelligence is one of the key enabling technologies that allow complex analysis of the vast amount of data being collected through these devices to provide healthcare inferences [4, 5]. These advanced systems support complex healthcare applications and enable continuous provision of medical needs. These devices can monitor a vast range of biomedical signals, such as electromyogram (EMG), electrocardiogram (ECG), respiratory rate, heartbeats, blood glucose levels, oxygen saturation, body or skin temperature and galvanic skin response. As an example, the CardioNet wearable system monitors patient heart rate, ECG and other data to help physicians diagnose and support patients suffering from arrhythmias [6, 7]. While advancement in healthcare technologies is paving the way for diagnosis, treatment and cure, there remain few diseases that lack a solution comprising a solid technological foundation. Neuromuscular Disorders (NMDs) include such diseases.

Neuromuscular disorders are a category of conditions that impair the function of muscles and the nerves that control them. These disorders disrupt the communication between the nervous system and the muscular system, leading to a variety of symptoms, most notably muscle weakness, atrophy, and sometimes paralysis [8]. The nervous system's ability to send signals to muscles is essential for voluntary movements such as walking, speaking, and even breathing. When this communication is compromised, it can result in significant challenges to mobility, daily functioning, and overall quality of life [9]. The term "neuromuscular disorders" encompasses a wide range of conditions, each with its own underlying cause,

979-8-3503-8168-9/24/$31.00 ©2024 IEEE.

progression, and impact [10]. Some of these disorders, like muscular dystrophies, are genetic and cause gradual muscle degeneration and weakness over time [11]. The onset, severity, and progression of neuromuscular disorders can vary widely, with some conditions appearing in childhood and others developing later in life [12]. Most NMDs are chronic and progressive, meaning that they worsen over time and, in many cases, are incurable [13]. Often, by the time a subject begins to experience noticeable changes or symptoms, it becomes impossible for healthcare professionals to reverse them. Currently, early detection is the most important step, as it can stop the disease from progressing further and causing permanent impairment.

Multiple works have used machine learning algorithms for evaluating neuromuscular conditions. A previous study summarized and illustrates various algorithms for efficient ways of understanding the electromyography (EMG) signals [14]. Artificial Neural Networks have been used to identify neuromuscular conditions of myopathy and neuropathy using EMG signals [15]. While EMG signals are able to provide insights towards detecting NMDs, it is an invasive process where the electrodes are inserted into the muscles and thus may not be suitable for application in portable/wearable devices. On the other hand, Surface Electromyography (sEMG) uses electrodes that can be used on the skin for collecting signals. A two-stage binary classifier model was reported in a previous study to identify neuropathy and myopathy using feature extraction and feature selection methods [16]. Chandrasekhar et al. proposed an ML algorithm to classify upper limb actions for assessing improvement in post-stroke patients [17]. Various techniques for processing and classifying sEMG data for various applications involving lower limb activity have also been reported recently [18] along with a detailed analysis of sEMG signal for eight hand motions [19]. These studies state a common conclusion of a need to evaluate the methods in more robust settings given the lack of availability of large sEMG datasets.

To overcome this, recently a comprehensive dataset of Indian healthy subjects performing 22 activities primarily aimed at hand gesture classification was presented. This electromyography analysis of Human Activity – Database 1 (EMAHA-DB1) [20] is a recent dataset and unexplored for detection of neuromuscular disorders. The dataset aims to provide valuable insights into understanding limb disabilities, ageing effects, and neuromotor deficits. This dataset can be particularly useful for classification studies and statistical analysis of sEMG signals. So far, it has been used for enhancing keystroke dynamics [21], automated categorization of microsurgical tools in the fields of surgery [22] and designing a human-machine interface for lower limb exoskeleton assistance system [23].

Thus, in this work, EMAHA-DB1 is used to develop an unsupervised machine learning approach to detect neuromuscular disorders as anomalies. Further, domain knowledge is used to design relevant features representing NMD biosignal markers to increase confidence in decision making. Additionally, this domain knowledge is also utilized for synthesizing a dataset that mimics the presence of neuromuscular disorders for robust evaluation of the machine learning algorithms. The paper is organized as follows: Section II describes the dataset. Section III presents the proposed methodology, which includes data preprocessing, feature design and extraction, and algorithm selection. Section IV presents the experiments performed and Section V presents the results and discussion.

II. Dataset Description

The Electromyography Analysis of Hand Activities – Data Base 1 (EMAHA-DB1) [20] is a comprehensive dataset designed to facilitate the study of surface electromyography (sEMG) signals collected during various activities of daily life (ADL). The study participants comprised 25 healthy subjects (22 males, 3 females) with no history of upper limb pathology with an average age of 28 years. The dataset used the following protocol for signal acquisition:

1) Actions: 10 different hand activities, each performed 10 times by each subject, were used in this work as: 1) Coin Tossing; 2) Finger Snapping; 3) Pulling Draw (Empty); 4) Pulling Draw (Heavy); 5) Pulling Draw (Empty) (Palmar View); 6) Pulling Draw (Heavy) (Palmar View); 7) Pushing Drawer (Empty); 8) Pushing Drawer (Heavy); 9) Hand Clasping; 10) Hand Clapping;

2) Recording Device: 5-channel Noraxon Ultium wireless sEMG sensor setup.

3) Electrode Placement: Self-adhesive Ag/AgCl dual electrodes placed at five key muscle sites on the right arm, given as: (i) Electrode 1: Brachioradialis (BR); (ii) Electrode 2: Flexor carpi radialis (FCR); (iii) Electrode 3: Flexor carpi ulnaris (FCU); (iv) Electrode 4: Biceps Brachii (BB); and (v) Electrode 5: Abductor Pollicis Brevis (APB).

4) Session Details: Each session lasted up to one hour. Each activity consisted of an action phase, a release phase (if applicable), and a rest phase. Activities were performed for 8 seconds with a 5-second rest between repetitions and a 30-second rest between different activities. The procedure within each activity is shown in Fig. 1. A sample signal from Electrode 5 for Action 1 is shown in Fig. 2.

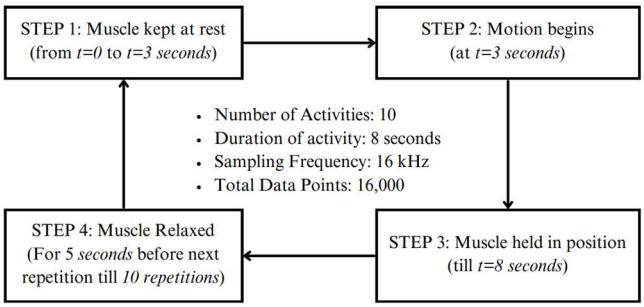

Fig. 1. Data acquisition within each activity in EMAHA-DB1.

III. Proposed Methodology

The flow diagram presented in Fig. 3 gives an overview of the proposed solution. The selected data from Section II is split into training and validation sets. On the training set, a preprocessing pipeline is developed which includes biomedical signal processing. Features were extracted from the preprocessed data for model training. Additionally, feature-based functions were

designed to mimic disorder behavior and generate testing data

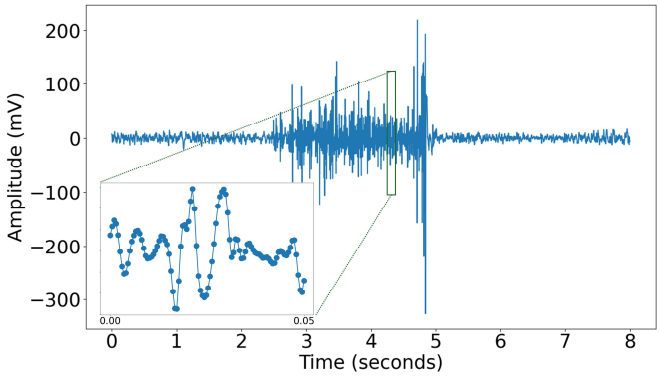

Fig. 2. Sample sEMG signal for Electrode 5, Action 1.

that includes a combination of healthy patient data from EMAHA-DB1 and generated anomalous patient data. The performance of the trained model was first evaluated on the validation set, followed by the testing set.

A. Signal Preprocessing

The raw sEMG signals underwent signal filtering processes [24] before being used for feature extraction. First, data was centered around the mean, followed by applying an IIR Notch Filter at frequency 50 Hz, as the powerline interference present in the recording conditions was 50 Hz [20] as shown in Fig. 4. Then, a Butterworth lowpass filter was applied and all frequencies above 400 Hz were removed, as shown in Fig. 5, as these frequencies are not present in human sEMG signals [25]. A further smoothening of the signal was carried out by applying PyWavelet's Denoising [20] as shown in Fig. 6. The processing features were applied on the dataset on a piece of signal.

B. Feature Extraction

Certain features in the sEMG signal are indicators of various neuromuscular disorders. The following features have been extracted as previously reported [25] and shown in Fig.7 and Fig. 8.

1) Zero Crossing Rate: This refers to the number of times the signal changes sign. Rapid firing of neurons is often an indication of Myopathy, while slow firing of neurons is an indication of Neuropathy. This is linked to the concept of recruitment frequency, where abnormal patterns are indicators of NMDs.

2) Peak Amplitude: This refers to the maximum value of the signal. Amplitudes higher than usual are an indication of misfiring in neurons, and lower amplitudes can be an indication of atrophy.

3) Peak Frequency: This refers to the highest frequency value in the spectrum. Higher frequencies are present in subjects with myopathy and a lower frequency in subjects with anterior horn cell lesions.

4) Mean Frequency and *Median Frequency*: These refer to the average and midpoints of frequency spectrum. While peak frequency is only a measure of the maximum in a range, for the analysis of the entire range and to notice if abnormally higher or abnormally lower frequencies are predominantly making up a signal, frequency is computed. This predominance of a frequency that is outside the normal range can potentially

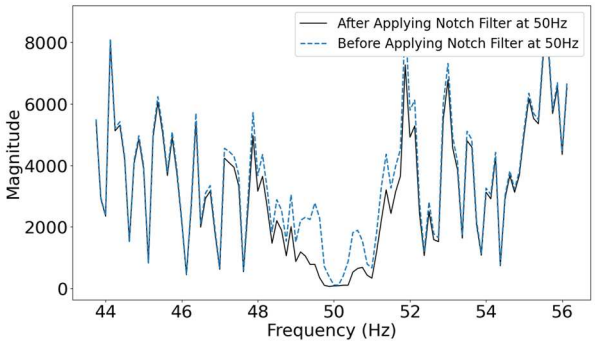

Fig. 4. Effect of applying notch filter on signal

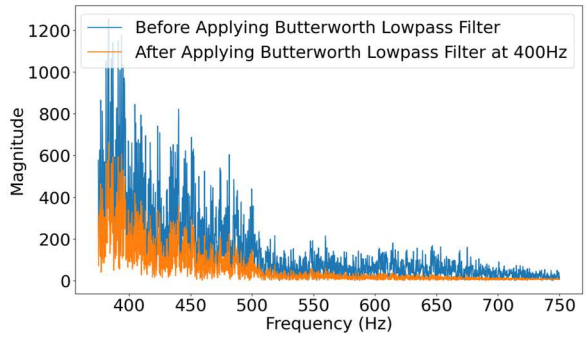

Fig. 5. Effect of applying Butterworth Lowpass filter

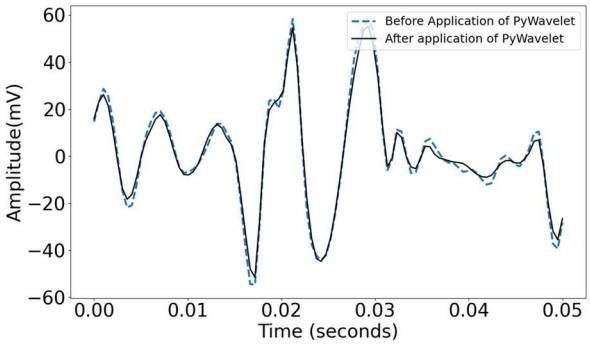

Fig. 6. Effect of applying PyWavelet Denoising

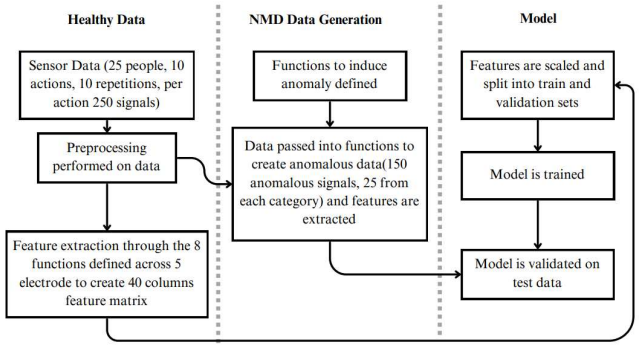

Fig. 3. An overview of the proposed solution

indicate a neuromuscular disorder. Median can support mean frequency evaluation if there are any outliers in data that do not indicate an anomaly but rather are a measurement artifact or systematic error.

5) Mean Absolute Value (MAV): This is the mean of the absolute values of the signal. Measuring the MAV can provide insights into the approximate variations in the amplitude across the signals because as discussed above, high peaks in each Motor Unit Action Potential (MUAP) indicate misfiring of neurons.

6) Integrated EMG (iEMG): This is the area under the absolute signal curve. A decreased area under the EMG curve denotes reduced MUAPs which are an indicator of neurogenic involvement, and an increased area may indicate hyperactivity of the neurons involved.

7) Root Mean Square (RMS): The RMS is often a reflection into the signal power and hence is utilized for measuring muscular activation levels and neuromuscular efficiency.

These 8 features were extracted for each sample. Data across all 5 electrodes was considered separately, and hence (8x5) features were obtained. Thus, the feature matrix was created with 40 columns and 250 rows denoting 250 samples per action. While the standard normal ranges for these features were suggested earlier [25], they may not stand valid due to various influencing factors. To address this, machine learning was employed to establish boundaries based on all features, rather than relying on simple threshold comparisons, to distinguish normal and anomalous signals.

For each action, a separate machine learning model was trained. For each model, based on the action, the suitable electrode was chosen out of the five available locations. This was performed due to multiple reasons. Firstly, the human joints are connected using different muscle groups. Each muscle group is placed at a specific location and length to facilitate a set of or a single motion. When any action is performed, different muscle groups are activated, and they contract to move the bone groups and execute the motion. This is true for even holding a particular pose. The muscles can be imagined as spring-dampers that pull or push the bones to move them or keep them stationary against a load [26]. For muscles to contract (activate), they require signal from motor neurons. This signal is in the form of electrical impulses. For activating specific muscle groups, corresponding motor neurons will provide the impulses [27]. These signals are eventually recorded using the sEMG electrodes, placed at designated regions. So, for specific types of motion, corresponding muscle groups will be activated by providing electrical signals from motor neurons and those signals are recorded by sEMG sensors. Different muscle groups were focused for different actions as described previously and their selection is specified in Table 1[26,27,28].

Table 1: Muscle group selected for each action

Action	Action Name	Muscle Group
1	Coin Tossing	APB
2	Finger Snapping	APB
3	Pulling Draw (Empty)	BB
4	Pulling Draw (Heavy)	BB
5	Pulling Draw (Empty) (Palmar View)	APB
6	Pulling Draw (Heavy) (Palmar View)	BR
7	Pushing Drawer (Empty)	BB
8	Pushing Drawer (Heavy)	APB
9	Hand Clasping	APB
10	Hand Clapping	APB

C. *Creation of synthetic data mimicking nature of NMDs*

Based on the domain-based features described in the previous section, this section describes the functions designed to inject anomalies into the healthy sEMG signals.

1) Boosting the Zero Crossing Rate (ZCR): Diseases like neuropathy and myopathy are characterized by their sharp deviation from normal sEMG signal due to their ZCR characteristic, i.e., an anomalous signal rapidly changes its phases. Hence, this effect was induced through an anomaly by changing the sign of the initial point to make it different from its neighbor point. This way the signal crosses the axis an additional time in this skip length.

2) Reducing the Zero Crossing Rate: A threshold ZCR was set, which was a random integer in the range (a,b). Then, a loop commenced, where random batches of 200 data points were taken into account. The number 200 signifies the approximate length where clear Motor Unit Action Potential (MUAP) was observed. This was then flattened in the similar way ZCR was reduced, until the ZCR dropped below the threshold ZCR.

3) Boosting Amplitude: It is often the case that affected muscles can generate significantly different power from normal signals, and that affects the amplitudes of the MUAP. So, the section of the signal where the movement of the gesture begins in the sEMG signal was taken into account, and identification of the MUAPs began by finding global maximas and minimas

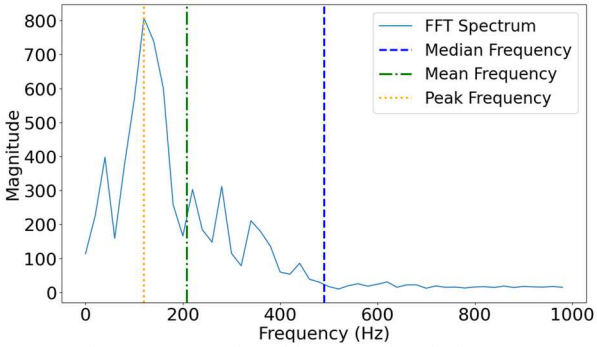

Fig. 7. Representation of frequency domain features

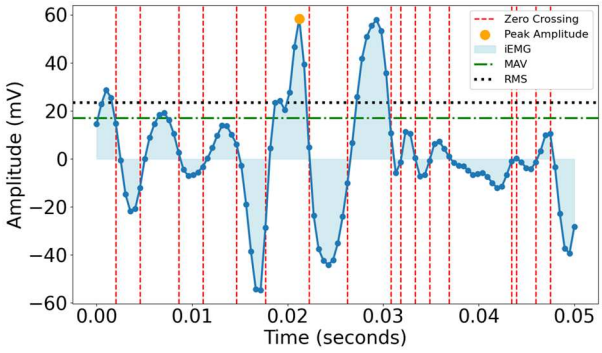

Fig. 8. Representation of time domain features

of that batch and increasing their magnitude by a random integer in the range (a,b).

4) Reducing Amplitude: Similar to the way the boosting amplitude function was created, the portion of the signal that comprises movement was identified, and further the MUAPs in the signal were identified. Next, their amplitudes were reduced by a factor, which was a random integer in the range (a, b). Fig. 9 shows a sample implementation of this anomaly.

5) Creating dips in signal in time domain: If a group of neurons is inactive or malfunctioning in a particular muscle region, MUAPs might not occur in a consecutive fashion as they might do in a healthy signal, and at some points MUAPs may entirely disappear. For this, the portion of the signal that corresponds to the motion part of the signal was taken up; further, in batches of a few MUAPs, a dip in that region was created by flattening variable length depending on a random integer selection between (a,b).

6) Creating dips in signal in frequency domain: A well-functioning muscle with good motor unit recruitment will show a broad spectrum of frequencies. If lapses are present, there might be dips or gaps in this frequency spectrum during those periods. For this, the fourier transform of the signal was taken, then random frequencies in random amounts around the mean and median frequencies were removed to create dips in the frequency spectrum, then an inverse fourier transform was applied to reconstruct the anomalous signal.

The pseudocodes for these functions are presented in Table 2. In order to determine the range (a, b) in each use case, the distribution of metrics of healthy data was analyzed. The lower boundary was set to barely bring anomalous data out of the ordinary zone, and the upper limit makes sure that changes do not go beyond 10% over the mean. Random numbers were selected within this range, and it allowed for creating a variety of anomalous signals, from minor to prominent anomalies, based on the random values chosen within the range and the signal's existing feature value. Thus, a dataset was synthesized, which included features of 150 unhealthy signals and 75 healthy signals to be used for testing of the developed machine learning algorithms.

D. Algorithm

One-Class Support Vector Machines (OCSVMs) are a type of machine learning algorithm designed for anomaly detection

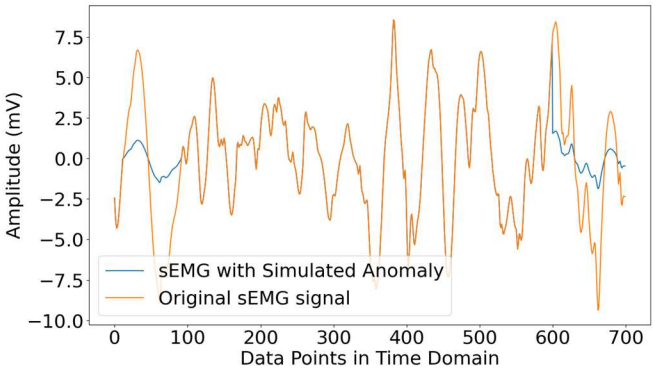

Fig. 9. Comparison of original signal with simulated anomalous signal

Table 2: Pseudocodes for anomaly inducing functions

Function	Pseudo Code
Boosting ZCR	function boost_ZCR(signal, a, b): step_size = random_integer(a, b) for i from 0 to length(signal) taking steps=step_size: if sign(signal[i]) == sign(signal[i + 1]): signal[i] = -signal[i] return signal
Reducing ZCR	function reduce_ZCR(signal, a, b): threshold_ZCR = random_integer(a, b) while calculate_ZCR(signal) > threshold_ZCR: batch_start = random_integer(0, length(signal) - 200) batch_end = batch_start + 200 for i from batch_start to batch_end - 1: signal[i] =0 return signal
Boosting Amplitudes	function boost_amplitudes(signal, a, b): MUAP=identify_muap(signal) multiplier=random_integer(a,b) MUAP=MUAP*multiplier result= reconstructer(signal+MUAP) return result
Reducing Amplitudes	function reduce_amplitudes(signal, a, b): MUAP=identify_muap(signal) factor=random_integer(a,b) MUAP=MUAP/factor result= reconstructer(signal+MUAP) return result
Creating dips in amplitude spectrum	function create_dips_amp_spectrum(signal): random range=(a,b) where signal[i]==random range: signal[i]=0 return signal
Creating dips in frequency spectrum	function create_dips_freq_spectrum(signal): frequencies=fourier transform(signal) frequencies_removed=random integer(a,b) frequencies=frequencies-frequencies_removed reconstructed_signal= inverse fourier transform(modified frequencies) return reconstructed signal

that learns a decision function using a high-dimensional feature space. The goal is to distinguish between normal data and outliers, which are rare or unusual observations. The algorithm finds the smallest hypersphere or hyperplane that encompasses most of the training data, treating points outside this boundary as anomalies. Mathematically, let this hyperplane have center a and radius $R > 0$. Let the data points be x_i, where i = 1,2,3…. n. The following constraint stands valid:

$$|x_i - a^2| \leq R^2 \; \forall \; i$$

R^2 as a cost function is sensitive to outliers and therefore may lead to inaccurate capturing of training data. Thus, the cost function becomes,

$$R^2 + C \sum_i \xi_i$$

and the constraint can be written as,

$|x_i - a^2| \leq R^2 + \xi_i \; \forall \; i$ where ξ_i are the positive weights associated with each data point. Combining this with the method of Lagrange Multipliers, the optimization problem to be solved is,

$$\mathcal{L}(R, a, \alpha_i, \gamma_i, \xi_i) = R^2 + C \sum_i \xi_i - \sum_i \alpha_i (R^2 + \xi_i - (|x_i|^2 - 2a \cdot x_i + |a|^2)) - \sum_i \gamma_i \xi_i$$

where α_i and γ_i are non-negative Lagrange multipliers. $\mathcal{L}$ should be maximised with respect to α_i and γ_i, but minimised with respect to R, a and ξ_i.

IV. EXPERIMENTS

As described previously, a series of data processing steps are followed before using it for model training and performance evaluation. Experiments were carried out at the level of selecting a suitable model as well as the hyperparameters of the model. For each choice of model, the hyperparameters were varied in the feasible range for performance comparison. The proposed OCSVM model was compared with two other models as described below:

1) Elliptic Envelope: This method is an outlier detection technique that fits an elliptic envelope to the data, assuming a Gaussian distribution. This method estimates the robust covariance of the data and constructs an ellipse or ellipsoid that covers the central data points while excluding outliers. It works well when the data is approximately normally distributed, as it leverages the Mahalanobis distance to detect points that fall outside the expected distribution.

2) Isolation Forests: Isolation Forests represent a robust method for anomaly detection in high-dimensional datasets. This unsupervised learning technique constructs multiple decision trees to isolate observations through random feature selection and value splitting. The fundamental principle is that anomalies, being both rare and different, require fewer splits to isolate, resulting in shorter average path lengths within the trees. Each observation's anomaly score is derived from these path lengths, effectively distinguishing anomalous data points from normal ones. This approach is advantageous due to its efficiency and scalability.

V. RESULTS AND DISCUSSION

All three models were passed through the same analysis pipeline using the same set for features. For each action out of the ten selected actions, a separate instance was trained for each model. A 4-fold cross validation was performed on the training and validation combined dataset, and the achieved mean and standard deviation of the accuracy across the three models and the actions are reported in Table 3 and Table 4. Similarly, a 4-fold cross validation was performed on the training and testing combined dataset, reporting the same metrics.

Based on the results, a very noticeable difference was visible in the accuracies of the same models across the two evaluation mechanisms– validation and testing. One class SVM performed poorly on the validation data compared to both Elliptic Envelope and Isolation Forests. Both the other algorithms exhibited significantly better performance. However, when the performance was evaluated on the testing dataset, OCSVM performed better than the other two algorithms. This substantiates the well-known fact that several anomaly detection models tend to underfit or overfit and do not form a good decision boundary. One Class SVM, on the other hand, maintained approximately constant performance, which shows that even when the test data consisted of anomalous cases, it was able to distinguish between normal and anomalous, confirming that the decision boundary it forms is robust and captures the necessary features. Table 5 shows the confusion matrix of the proposed OC-SVM model for action 7. Corresponding to the value of achieved accuracy, it can be noted there are cases of false negatives that would need further minimization and performance improvement.

Table 3: Results on validation dataset.

Action Num	Accuracy of Elliptic Envelope		Accuracy of Isolation Forest		Accuracy of One Class SVM	
	Mean	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***
1	99.34	0.0115	99.67	0.0057	87.00	0.0238
2	99.34	0.0115	97.66	0.0057	84.67	0.0346
3	96.68	0.0115	96.67	0.0221	84.34	0.0197
4	98.67	0.0000	98.34	0.0057	82.33	0.0455
5	99.00	0.0057	98.67	0.0163	93.33	0.0163
6	97.34	0.0000	97.34	0.0000	80.68	0.0200
7	98.67	0.0000	99.00	0.0057	85.00	0.0455
8	99.34	0.0115	100.00	0.0000	77.33	0.0432
9	99.34	0.0115	96.00	0.0094	83.00	0.0173
10	98.34	0.0057	98.67	0.0000	80.66	0.0199
Avg.	**98.60**	0.0068	**98.19**	0.0070	**83.73**	0.0285

Table 4: Results on test dataset.

Action Num	Accuracy of Elliptic Envelope		Accuracy of Isolation Forest		Accuracy of One Class SVM	
	Mean	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***
1	68.56	0.0057	72.77	0.0145	75.11	0.0099
2	58.67	0.0224	73.11	0.0022	77.11	0.0049
3	66.11	0.0319	74.00	0.0091	79.45	0.0119
4	64.44	0.0174	77.56	0.0179	83.00	0.0036
5	45.00	0.0036	83.77	0.0049	86.12	0.0048
6	76.33	0.0127	75.89	0.0090	80.89	0.0121
7	77.11	0.0111	79.34	0.0091	83.89	0.0036
8	60.22	0.0503	83.67	0.0048	82.77	0.0065
9	44.33	0.0630	77.56	0.0209	81.55	0.0155
10	55.78	0.1072	84.34	0.0072	85.34	0.0133
Avg.	**60.67**	0.0325	**78.08**	0.0099	**81.45**	0.0086

Table 5: Confusion Matrix for OCSVM for Action 7.

		True Labels	
		Faulty	***Healthy***
Predicted Label	***Faulty***	138	23
	Healthy	12	52

VI. Conclusion and Future Work

This study presents an unsupervised machine learning framework for the detection of neuromuscular disorders (NMDs) using a novel human dataset - Electromyography Analysis of Hand Activities – Database 1 (EMAHA-DB1). The study effectively integrated domain-specific knowledge into the feature engineering process, enabling the synthesis of an augmented NMD dataset that enables a robust evaluation of the proposed One-Class Support Vector Machine (OCSVM) model. Based on the NMDs biosignal markers, anomalous signal fragments were imposed on the dataset to obtain the modified anomalous sEMG data while applying suitable constraints to maintain characteristic of the real human sEMG signals. A comparative performance analysis of OCSVM with Elliptic Envelope and Isolation Forest models revealed that, while the OCSVM exhibited lower accuracy on the validation dataset, it outperformed the other models on the test dataset containing both normal data and synthetic anomalous data. This demonstrated the ability of OCSVM to establish a more resilient decision boundary for anomaly detection in sEMG signals. Although this study aims to establish OCSVM as a suitable model for NMD detection, it will be expanded further to include enhancing the model generalizability through the incorporation of larger, more heterogeneous datasets. Datasets tagged with specific NMDs will also be explored because, as seen with the current dataset, number of false positives mark a striking percentage. Advanced feature extraction techniques will be included to closely mimic the differences in the action potential and nerve conduction velocity in the genesis of sEMG signals. Additionally, the development of portable, non-invasive sEMG acquisition systems could extend the practical applicability of this approach in clinical environments.

Acknowledgement

This work has been supported by Charak Center for Digital Healthcare, established at IIT Indore together with IITI DRISHTI CPS Foundation under the National Mission on Inter-disciplinary Cyber Physical System, Govt. of India. The authors would like to thank Dr. Sourav Chandra from the Department of Biosciences and Biomedical Engineering at IIT Indore for his guidance during the work.

References

[1] Greenspun, H., Coughlin, S.: mHealth in an mWorld: How mobile technology is transforming health care. Deloitte Center for Health Solutions, Washington (2012)

[2] Geneau, R., Stuckler, D., Stachenko, S., McKee, M., Ebrahim, S., Basu, S., Chockalingham, A., Mwatsama, M., Jamal, R., Alwan, A., Beaglehole, R.: Raising the priority of preventing chronic diseases: A political process. Lancet 376, 1689–1698 (2010)

[3] Giselbrecht, S., Rapp, B., Niemeyer, C.: The Chemistry of Cyborgs—Interfacing Technical Devices with Organisms. Angew. Chem. Int. Ed. 52, 13942–13957 (2013)

[4] Qayyum, A., Qadir, J., Bilal, M. and Al-Fuqaha, A., 2020. Secure and robust machine learning for healthcare: A survey. IEEE Reviews in Biomedical Engineering, 14, pp.156-180.

[5] Zhang, A., Xing, L., Zou, J. and Wu, J.C., 2022. Shifting machine learning for healthcare from development to deployment and from models to data. Nature Biomedical Engineering, 6(12), pp.1330-1345.

[6] Anita Pal, Ranjeet Srivastva, Yogendra Narain Singh: CardioNet—An Efficient ECG Arrhythmia Classification System Using Transfer Learning

[7] Rafael Gomez and Anna Harrison: Beyond Wearables—Experiences and Trends in Design of Portable Medical Devices

[8] Mercuri, E., & Muntoni, F. (2013). Muscular dystrophies. The Lancet, 381(9869), 845-860.

[9] Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., ... & Zoing, M. C. (2011). Amyotrophic lateral sclerosis. The Lancet, 377(9769), 942-955.

[10] Dimachkie, M. M., & Barohn, R. J. (2013). Chronic inflammatory demyelinating polyneuropathy. Seminars in Neurology, 33(2), 163-173.

[11] Emery, A. E. H. (2002). The muscular dystrophies. The Lancet, 359(9307), 687-695.

[12] Carter, G. T., Abresch, R. T., Fowler, W. M., Johnson, E. R., Kilmer, D. D., & McDonald, C. M. (1995). Profiles of neuromuscular diseases: Duchenne muscular dystrophy. American Journal of Physical Medicine & Rehabilitation, 74(5), S70-S92.

[13] Darras, B. T., Urion, D. K., & Ghosh, P. S. (2018). Dystrophinopathies. In GeneReviews® (University of Washington, Seattle).

[14] Techniques of EMG signal analysis: detection, processing, classification and applications M. B. I. Reaz, M. S. Hussain and F. Mohd-Yasin

[15] Electromyography (EMG) based Classification of Neuromuscular Disorders using Multi-Layer Perceptron I. Elamvazuthi, N.H.X. Duy, Zulfiqar Ali , S.W. Su , M.K.A. Ahamed Khan and S. Parasuraman

[16] Jeong, Se-Jin Ahn, Two-stage binary classifier for neuromuscular disorders using surface electromyography feature extraction and selection, Medical Engineering & Physics

[17] V. Chandrasekhar, V. Vazhayil and M. Rao, "Design of a real time portable low-cost multi-channel surface electromyography system to aid neuromuscular disorder and post stroke rehabilitation patients," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 2020, pp. 4138-4142, doi: 10.1109/EMBC44109.2020.9176058.

[18] Vijayvargiya, A., Singh, B., Kumar, R. et al. Human lower limb activity recognition techniques, databases, challenges and its applications using sEMG signal: an overview. Biomed. Eng. Lett. 12, 343–358 (2022). https://doi.org/10.1007/s13534-022-00236-w

[19] Analysis of Extracted Forearm sEMG Signal Using LDA, QDA, K-NN Classification Algorithms Firas AlOmari and Gunhai Liu

[20] Naveen Kumar Karnam, Anish Turlapaty, Shiv Ram Dubey, and Balakrishna Gokaraju. (2023). EMAHA-DB1 [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/6814847

[21] Yaacob, M.N., Idrus, S.Z.S. and Mohammed, A.H., 2024, May. Enhancing keystroke dynamics accuracy with optimal SVM kernel usage. In AIP Conference Proceedings (Vol. 3135, No. 1). AIP Publishing.Li, W., Ma, Y., Shao, K., Yi, Z., Cao, W., Yin, M., Xu, T., & Wu, X. (2024). The Human–Machine Interface Design Based on sEMG and Motor Imagery EEG for Lower Limb Exoskeleton Assistance System. IEEE Transactions on Instrumentation and Measurement, 73, 1-14.

[22] Reddy, M., Jonna, P., Perala, S., Rao, M., & Vazhiyal, V. (2023). Automated Microsurgical Tool Categorization using a Surface-Based EMG System. 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 1-5.

[23] Li, W., Ma, Y., Shao, K., Yi, Z., Cao, W., Yin, M., Xu, T., & Wu, X. (2024). The Human–Machine Interface Design Based on sEMG and Motor Imagery EEG for Lower Limb Exoskeleton Assistance System. IEEE Transactions on Instrumentation and Measurement, 73, 1-14.

[24] Study of signal processing techniques for EMG analysis Manoj Duhan and Chanderpal Sharma

[25] Electrodiagnosis in Diseases of Nerve and Muscle, Jun Kimura

[26] On control of reaching movements for musculo-skeletal redundant arm model, Kenji Taharaa, Suguru Arimotob, Masahiro Sekimotob and Zhi-Wei Luo

[27] In Vivo Neuromechanics: Decoding Causal Motor Neuron Behavior with Resulting Musculoskeletal Function, Massimo Sartori, Utku Ş. Yavuz & Dario Farina

[28] Zhang, A., Xing, L., Zou, J. and Wu, J.C., 2022. Shifting machine learning for healthcare from development to deployment and from models to data. Nature Biomedical Engineering, 6(12), pp.1330-1345.