



EMG BASED DIAGNOSIS OF MYOPATHY AND NEUROPATHY USING MACHINE LEARNING TECHNIQUES

Anisha.C. D¹, Dr. Arulanand.N²

¹ME Software Engineering student, PSG college of Technology

²Professor and Associate HOD, PSG college of Technology

¹ani.c.dass@gmail.com

²naa.cse@psgtech.ac.in

Abstract: Myopathy and Neuropathy are non-progressive and progressive neuromuscular disorders which weakens the muscles and nerves respectively. Electromyography (EMG) signals are bio signals obtained from the individual muscle cells. EMG based diagnosis for neuromuscular disorders is a safe and reliable method. Integrating the EMG signals with machine learning techniques improves the diagnostic accuracy. The proposed system performs analysis on the clinical raw EMG dataset which is obtained from the publicly available PhysioNet database. The two-channel raw EMG dataset of healthy, myopathy and neuropathy subjects are divided into samples. The Time Domain (TD) features are extracted from divided samples of each subject. The extracted features are annotated with the class label representing the state of the individual. The annotated features split into training and testing set in the standard ratio 70: 30. The comparative classification analysis on the complete annotated features set and prominent features set procured using Pearson correlation technique is performed. The features are scaled using standard scaler technique. The analysis on scaled annotated features set and scaled prominent features set is also implemented. The hyperparameter space of the classifiers are given by trial and error method. The hyperparameters of the classifiers are tuned using Bayesian optimization technique and the optimal parameters are obtained. and are fed to the tuned classifier. The classification algorithms considered in the analysis are Random Forest and Multi-Layer Perceptron Neural Network (MLPNN). The performance evaluation of the classifiers on the test data is computed using the Accuracy, Confusion Matrix, F1 Score, Precision and Recall metrics. The evaluation results of the classifiers states that Random Forest performs better than MLPNN wherein it provides an accuracy of 96 % with non-scaled Time Domain (TD) features and MLPNN outperforms better than Random Forest with an accuracy of 97% on scaled Time Domain (TD) features which is higher than the existing systems. The inferences from the evaluation results is that

Bayesian optimization tuned classifiers improves the accuracy which provides a robust diagnostic model for neuromuscular disorder diagnosis.

Keywords—EMG, Bayesian optimization, MLPNN, Random Forest

I INTRODUCTION

Electromyography (EMG) signals are generated from the muscle cells and are measured in electric potentials [1][13]. EMG based diagnosis for neuromuscular disorders is reliable since it denotes the difference of electric potentials of the electrodes which is generated by the subject muscle cells while performing movements [1][2]. Myopathy is a non- progressive neuromuscular disorder which affects the muscle cells. Neuropathy is a progressive neuromuscular disorder which affects the nerve cells [3]. Integration of Machine Learning algorithms with EMG signals improves the speed and accuracy of diagnosis.

The drawback identified in the existing systems is that lack of analysis in tuning the hyperparameters of the classifiers since it is the most essential step in improving the accuracy of the model. Another drawback identified is that lack of analysis on impact of features based on feature selection techniques.

The paper is organized as follows: section II discusses about the related works on neuromuscular disorder diagnosis methods. Section III presents the methodologies of the proposed framework with the block diagram covering all processes. Section IV provides the results and its inferences. Section V covers the conclusion and future work.

II RELATED WORKS

Kehri et al [1], presents an EMG signal analysis for the diagnosis of myopathy and neuropathy. The analysis focuses on decomposition of signals using wavelet transform (WT). The statistical features are extracted from the decomposed signals. The classification algorithms namely Artificial Neural Network (ANN) and Non-Linear Support Vector Machine (SVM) are used for discriminating myopathy and neuropathy subjects from healthy subjects. V. Kehri et al [2], provides a review on different feature extraction techniques and classification techniques. The analysis focus on classification of neuromuscular disease based on EMG signals on various combinations of feature extracted techniques. Amit kumar singh et al [3], discriminates the neuromuscular disorder EMG signals by decomposing the signals using Empirical Mode

Decomposition (EMD) wherein the features namely mean, standard deviation, variance and entropy were obtained. Swaroop. R et al in this paper [5] has presented a back-propagation algorithm which classify the healthy, myopathy and neuropathy EMG signals. I. Elamvazhuthi et al [15] presents a classification of neuromuscular disorders using Artificial Neural Network (ANN) based on the features known as Auto Regressive (AR), Root Mean Square (RMS), Zero Crossing (ZC), Mean Absolute Value (MAV) and Waveform Length (WL).

III METHODOLOGIES

Figure 1 depicts the block diagram of the Proposed Framework wherein it presents the techniques used in each process.

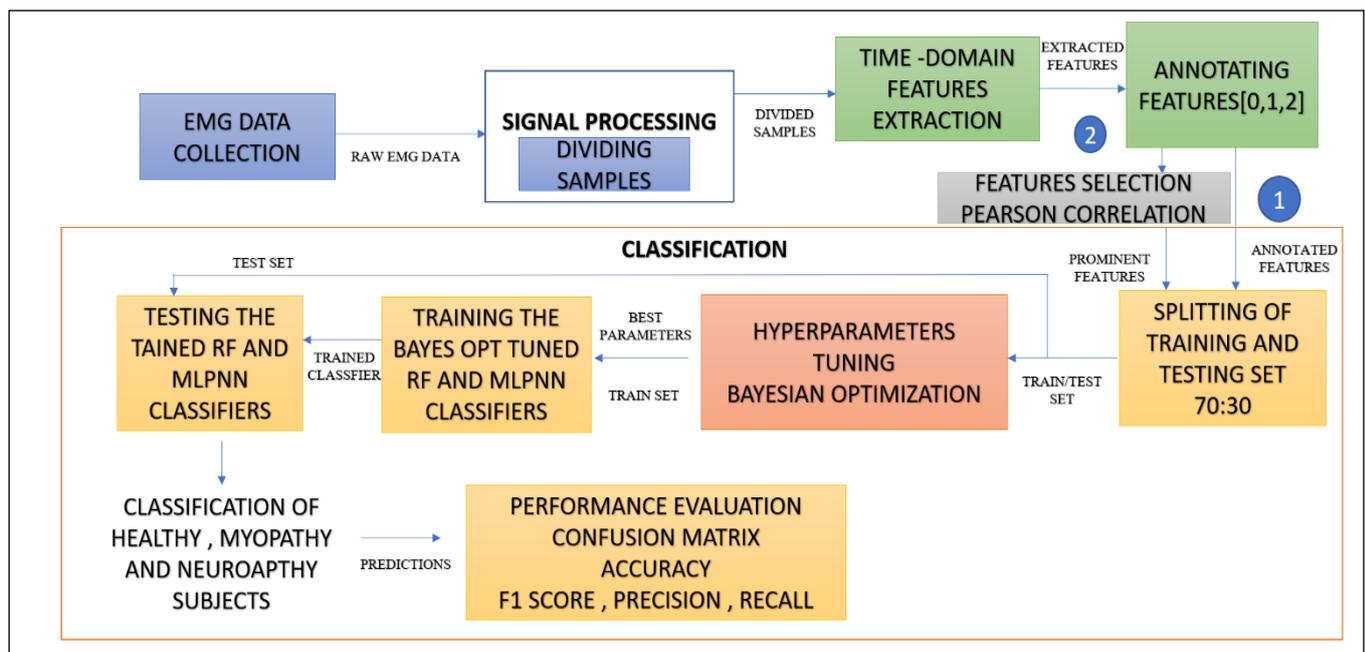


Figure 1 Block Diagram of the Proposed Framework

A. Dataset Description

The EMG data has been retrieved from publicly available PhysioNet database. EMG signals has been recorded by inserting a needle electrode into Tibialis anterior muscle and the subjects has been asked to dorsiflexed the foot. The repositioning of electrodes has been performed until the Motor Unit Potentials

(MUP) were obtained. EMG data from subjects is acquired using using Medelec synergy N2 EMG monitoring system.[12]

B. Signal Processing-Dividing the EMG signals into samples

The EMG signals of healthy, myopathy and neuropathy subjects are divided into samples with 100 instances per sample for healthy subject and 200

instances per sample for myopathy and neuropathy subjects. The division of samples provides about 509 samples for healthy subjects, 552 samples for myopathy subjects, 740 samples for neuropathy subjects. The total of 1801 instances of all three subjects is obtained as a result.

C. Features Extraction

The following 15 Time Domain (TD) based features are extracted from the divided samples.[11]

1. Enhanced Mean Absolute Value (EMAV): It is an enhancement of Mean Absolute Value (MAV) with computation including a threshold value p .
2. Enhanced Wavelength (EW): It is an enhancement of Wavelength(W) with computation including a threshold value p .
3. Mean Absolute Value: This parameter represents the amplitude of the signal. It provides the addition of the absolute value of EMG signal.
4. Wavelength: This parameter provides the cumulative length of the signal covering all neighboring peaks.
5. Root Mean Square: This parameter directly implies the amplitude of the signal. It is computed by obtaining the square root of the average of the EMG signal.
6. Average Amplitude change: It represents the average of the amplitude of the EMG signal.
7. Difference Absolute Standard Deviation value: It denotes the standard deviation of the adjacent samples.
8. Log detector: It is associated with the exerted force of the movement.
9. Modified Mean Absolute Value: It is an extension of MAV with assigned weights.
10. Modified Mean Absolute Value 2: It is an extension of MAV with continuous weights assigned.
11. Myopulse Percentage Rate: It is a mean of the output wherein the output exceeds the absolute value exceed the predefined threshold value.

12. Simple Square Integral: It is the sum of the squared values of the EMG signal.

13. Variance of EMG: It is the average of the power of the signal.

14. Willison Amplitude: Changes in EMG signal when it exceeds the threshold values is specified using this parameter.

15. Maximum Fractal Length: The strength of the contraction is specified by this parameter.

D. Annotating Features

The extracted features are annotated with class labels wherein class label 0 represents healthy subjects, 1 represents myopathy subjects and 2 represents neuropathy subjects.

E. Feature Selection

i. Working Principle of Pearson Correlation Coefficient (PCC)

PCC is used to find linear dependency between variables using the covariance between the features and class label and dividing the covariance value with square root of the product of features and class label [8].

ii. Output of Pearson Correlation coefficient (PCC)

The correlation between features and with the class label is provided in the range from -1 to 1 where value

- nearer to zero refers to weaker correlation
- closer to 1 denotes to stronger positive correlation
- closer to -1 represents stronger negative correlation

The nine prominent features namely EW, MAV1, RMS1, W, RMS2, LD, MMAV, MMAV2, SSI1 are obtained from this Feature selection process.

iii. Uncorrelated Features

From this nine prominent Features, uncorrelated features are procured. The uncorrelated features such as Enhanced Wavelength and Root Mean Square are procured.

F. Splitting of Training and Testing Set

The annotated features are split into training and testing set with 70 % data contributing to training set and remaining 30 % data contributing to testing set.

G. Hyperparameters Tuning

The hyperparameters of the classifiers are tuned using Bayesian optimization technique.

i. Working Principle of Bayesian optimization

The Bayesian optimization setup consists of an objective function, parameter space on which the search algorithm is applied and trials database. The search algorithm used is tree of Parzen algorithm.[9]

ii. Input Hyperparameter Space of MLPNN

The hidden_layer_sizes and max_iter parameters values are specified using trial and error method.

- activation: ['relu', 'logistic', 'tanh', 'identity']
- solver: ['lbfgs', 'adam']
- hidden_layer_sizes: [(1,),(2,),(3,),(4,),(5,),(6,),(7,),(8,),(9,),(10,),(11,),(12,),(13,),(14,),(15,),(16,),(17,),(18,),(19,),(20,),(21,)]
- learning_rate: ['constant', 'invscaling', 'adaptive']
- 'max_iter': [1000,2000,3000,4000,5000,6000...,20000]

iii. Input Hyperparameter Space of Random Forest

Here max_depth and n_estimators parameter values are specified using trial and error method.

- Criterion: ['entropy', 'gini']
- max_depth: [10, 100, 10]
- max_features: ['auto', 'sqrt', 'log2', None]
- n_estimators: [10, 60,600,1500]

i. Classification

Random Forest and MLPNN are considered for classification since Random Forest is an ensemble technique which combines the prediction of several decision trees which would be better than the prediction provided by single classifier and MLPNN integrates predictions of several hidden layers.

i. Working Principle of Multi-Layer Perceptron Neural Network

The Multi-Layer Perceptron is the commonly used form of ANN[7]. It consists of input layers, hidden layers and output layers. Each layer consists of neurons. The neurons in input layer distributes the input and the neurons in hidden layer performs weighted computations and produce output to the neurons in the output layer.[14]

ii. Working Principle of Random Forest

Training Phase: It trains several decision trees with the training data in parallel.[10]

Testing Phase: It combines the predictions of several decision tree on testing data using majority voting technique. [10]

j. Performance Evaluation

- i. Confusion Matrix: provides the matrix of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).
- ii. Precision: It is the ratio of True Positives (TP) and the total number of True Positive (TP) and False Positives (FP).
- iii. Recall: It is the ratio of True Positives (TP) and the total number of True Positive (TP) and False Negatives (FN).
- iv. F1 score denotes the weighted mean of Precision and Recall.

IV RESULTS ANALYSIS

The proposed work analysis is performed in three experiments. The following are the conducted experiments:

1. Experiment 1: The experiment 1 performs classification analysis of Bayesian opt tuned Random forest and MLPNN with the Time Domain Feature set which consists of 19 features.
2. Experiment 2: The experiment 2 performs features selection analysis on the extracted Time Domain Features using Pearson Correlation technique and the prominent features are fed to the Bayesian opt tuned Random forest and MLPNN.

3. Experiment 3: The experiment 3 focuses on the classification with uncorrelated features which are obtained by eliminating the most correlated features from the prominent features procured using experiment 2.

Figure 2 presents the highest accurate confusion matrix of Scaled Multi-Layer Perceptron

and Non-Scaled Random Forest procured in experiment 1; where x axis represents the predicted label count and y axis represents the true label. The number of correctly predicted data points will be present in the diagonal axis. The class label 0 represents healthy subject ,1 represents myopathy subject and 2 represents neuropathy subjects. Figure 3 depicts the accuracy graph.

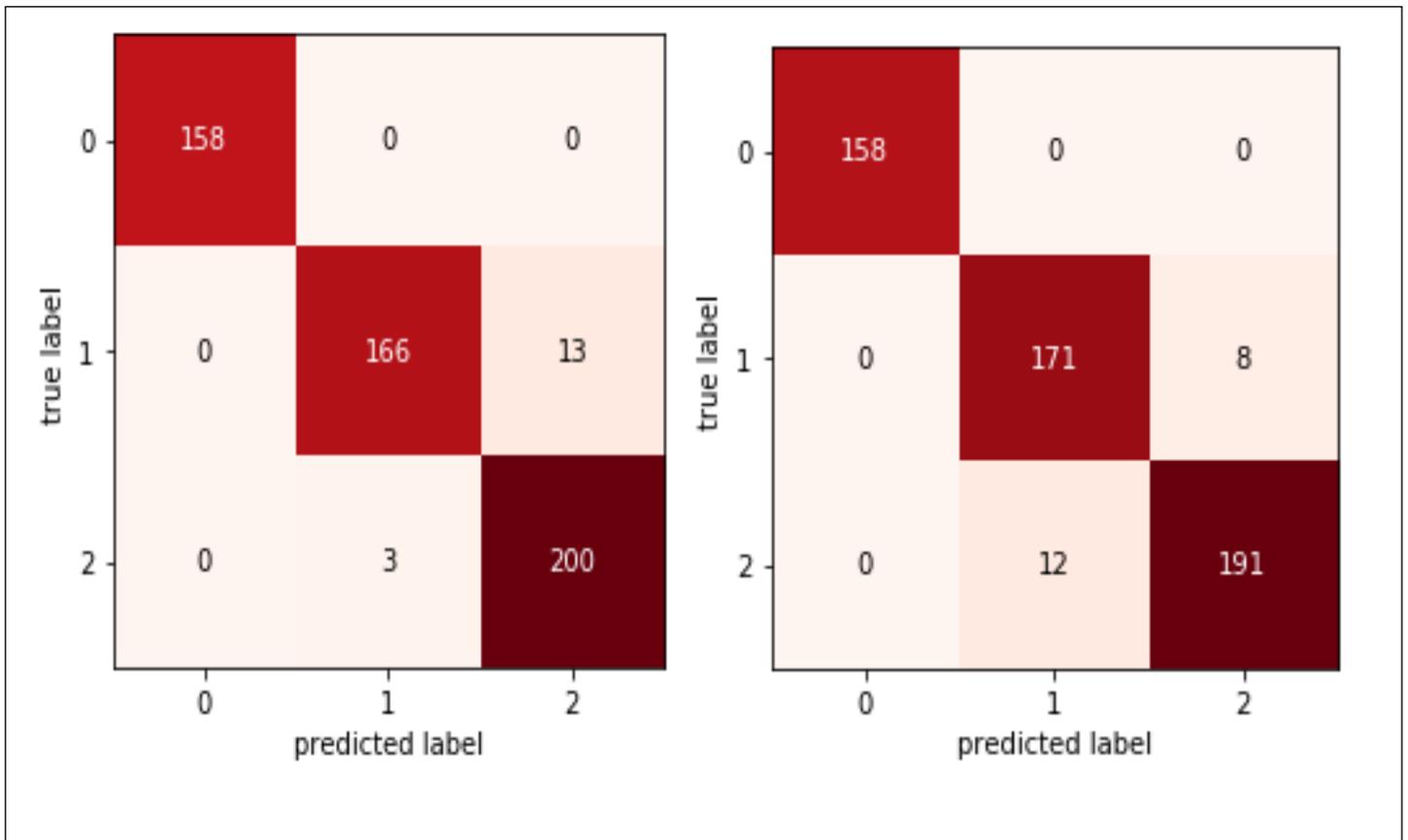


Figure 2 Best Accurate Confusion Matrix of Scaled MLPNN(Left) And Non-Scaled Random Forest (Right)

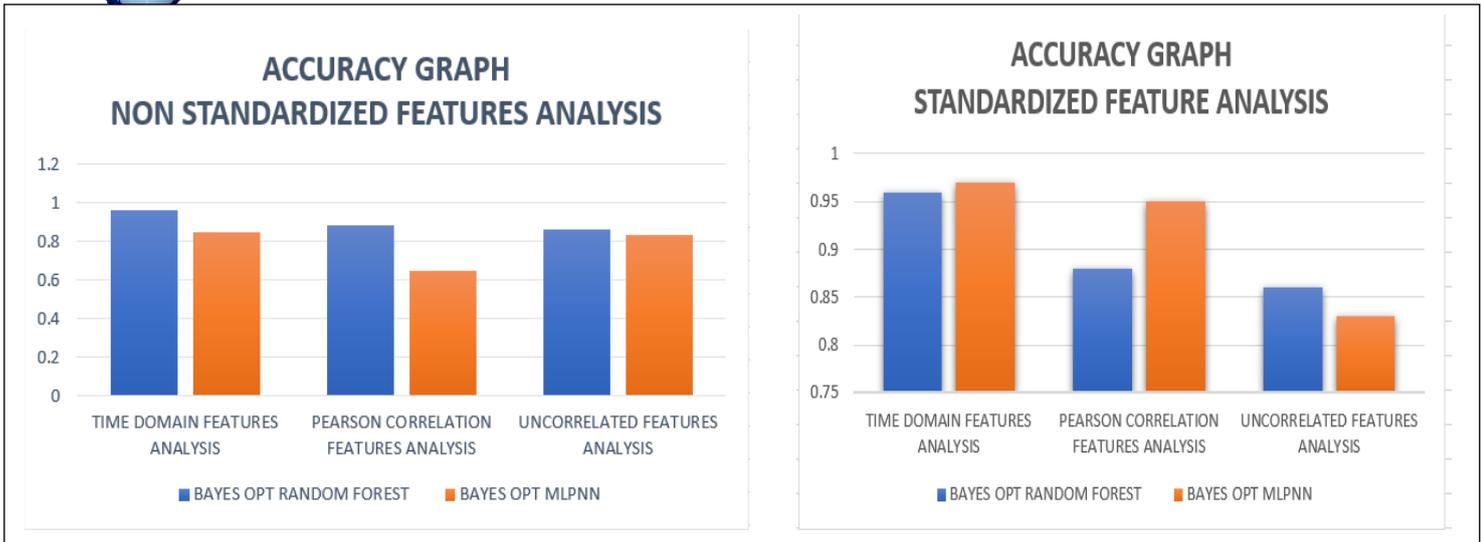


Figure 3 Accuracy Graph of Three Experiments.

TABLE I CLASSIFIER PERFORMANCE EVALUATION

EXP.NO	EXPERIMENT	CLASSIFIER	PRECISION	RECALL	F1 SCORE	SUPPORT
NO SCALING OF FEATURES						
1	TIME DOMAIN FEATURES ANALYSIS	BAYESIAN OPT TUNED RANDOM FOREST	0.96	0.96	0.96	540
		BAYESIAN OPT TUNED MLPNN	0.86	0.86	0.86	540
2	PEARSON CORRELATION(PC) FEATURES ANALYSIS	BAYESIAN OPT TUNED RANDOM FOREST	0.88	0.89	0.88	540
		BAYESIAN OPT TUNED MLPNN	0.67	0.67	0.64	540
3	UNCORRELATED FEATURES ANALYSIS OF PEARSON CORRELATION FEATURES	BAYESIAN OPT TUNED RANDOM FOREST	0.87	0.87	0.86	540
		BAYESIAN OPT TUNED MLPNN	0.85	0.84	0.83	540
SCALING OF FEATURES						
1	TIME DOMAIN FEATURES ANALYSIS	BAYESIAN OPT TUNED RANDOM FOREST	0.96	0.96	0.96	540
		BAYESIAN OPT TUNED MLPNN	0.97	0.97	0.97	540
2	PEARSON CORRELATION(PC) FEATURES ANALYSIS	BAYESIAN OPT TUNED RANDOM FOREST	0.88	0.89	0.88	540
		BAYESIAN OPT TUNED MLPNN	0.95	0.95	0.95	540
3	UNCORRELATED FEATURES ANALYSIS OF PEARSON CORRELATION FEATURES	BAYESIAN OPT TUNED RANDOM FOREST	0.87	0.87	0.86	540
		BAYESIAN OPT TUNED MLPNN	0.84	0.84	0.83	540



a. Inferences from the Experimental results:

1. Random Forest performs better without scaling the features than Multi-Layer Perceptron since the Random forest trains several decision trees and combines the predictions of several decision trees, it is more accurate and robust after tuning the parameters.
2. Random Forest provides same accuracy with and without scaling. MLPNN outperforms random forest after scaling with an accuracy of 97% for time domain-based features and with an accuracy of 95% with Pearson Correlation Features.
3. Time Domain Features set provided a highest accuracy result of 97% for MLPNN and 96% for Bayesian optimization tuned random forest than Pearson correlation features and uncorrelated features.
4. MLPNN classifier provided an accuracy of 97% and it outperforms than random forest after scaling the Time Domain Features using Standard Scaler.
5. Random Forest provided good accuracy of 96% without any feature selection techniques since the feature importance attribute which is present in tree-based classifier is more accurate than the explicit feature selection techniques.

V CONCLUSION AND FUTURE WORK

Myopathy and neuropathy are neuromuscular disorders which affect the cells of the muscle and the nerve. EMG diagnosis is essential since it measures the muscle activity and processing of EMG requires an appropriate signal processing technique. Machine Learning Techniques improves the diagnostic accuracy by training the models with the processed EMG signals. The existing systems lacks focuses on tuning of the classifiers hyperparameters which is the most important procedure for gaining good accurate and robust model. The proposed framework mainly focuses on tuning the Random Forest and Multi-Layer Perceptron Neural Network (MLPNN) classifier using Bayesian Optimization technique which makes informed decisions at each iteration in forming the combination of parameters and

provides the optimal parameters for the classifiers. Another problem identified in existing system is that lack of impact analysis of features using feature selection techniques. The above specified problem is covered in the proposed framework by conducting three experiments which is based on Time Domain Feature Analysis, Pearson Correlation Technique Feature Analysis and Uncorrelated Feature Analysis. The Bayesian optimization tuned MLPNN provides highest accuracy of 97% and Random Forest of 96% with Time Domain Features. Future Enhancement to this work is to build a robust diagnostic Machine Learning (ML) Model by tuning the parameters of different classifiers and combine the classifier's predictions using Ensemble techniques and also to performs analysis with different feature selection techniques.

REFERENCES

1. Vikram Kehri et al, "EMG Signal Analysis for Diagnosis of Muscular Dystrophy Using Wavelet Transform SVM and ANN", Biomedical and Pharmacology Journal, September 2018.
2. V. Kehri et al, "Techniques of EMG signal analysis and classification of Neuromuscular diseases", Advances in Intelligent Systems Research. Vol. 137, Pp. 485-491, 2017.
3. Amit Kumar Singh et al, "Discrimination of Myopathy, Neuropathy and Healthy EMG Signals", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, Issue 5, May 2017.
4. Amit Kumar Singh et al, "Approach for Classification of Neuromuscular Disorder using EMG Signals", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 5, May 2017.
5. Swaroop R et al, "Classification of Myopathy and Neuropathy EMG signals using Neural Network", 2017 International Conference on circuits Power and Computing Technologies [ICCPCT], IEEE, 2017
6. Anjana Goen, "Classification of EMG Signals for Assessment of Neuromuscular Disorder", International Journal of Electronics and Electrical Engineering Vol. 2, No. 3, September, 2014.
7. Ö. Galip Saracoglu, "An Artificial Neural Network Approach for the Prediction of Absorption Measurements of an Evanescent Field Fiber Sensor", Sensors 2008.
8. Obilor et al, "A Test for Significance of Pearson's Correlation Coefficient(r)", International Journal of Innovative Mathematics, Statistics & Energy Policies 6(1):11-23, Jan-Mar, 2018
9. James Bergstra et al, "Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms.", Proc. of the 12th Python in Science Conf. (Scipy 2013)
10. Pedregosa et al, "Scikit-learn: Machine Learning in Python.", Journal of Machine Learning Research 12 (2011) 2825-2830.



11. Michala (2020), "Electromyographic signal Analysis Tool – EMGAT2014", MATLAB Central File Exchange. Retrieved March 29, 2020.
12. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals", *Circulation*. 101(23): e215-e220, 2003
13. S. Boisset, F Goubel, "Integrated electromyography activity and muscle work," *J Applied Physiol*, vol 35, pp. 695-702,1972.
14. P. Mishra, S. Mishra, J. Nanda and K. V. Sajith, "Multilayer perceptron neural network (MLPNN) controller for automatic generation control of multiarea thermal system," *North American Power Symposium*, Boston,MA, 2011, pp. 1-7. 2011
15. I. Elamvazuthia et al, "Electromyography (EMG) based Classification of Neuromuscular Disorders using Multi-Layer Perceptron", *IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2015)*.