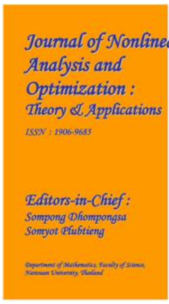
**J**ournal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1 : 2024
ISSN : **1906-9685**




# ELECTROMYOGRAPHY SIGNALS ANALYSIS USING MACHINE LEARNING

**Dr. Mallika Jain1, Dr. V. Elanangai2, *Dr. Tejashree Tejpal Moharekar3 and Dr. Tejpal Jaysing Moharekar4**
1Senior lecturer, Department of Electronics, Kalaniketan Polytechnic College, Jabalpur, MP, India.
2Assistant Professor, Department of EEE, St. Peter's Institute of Higher Education and Research, Chennai, TN, India.
3Assistant Professor, Yashwantrao Chavan School of Rural Development, Shivaji University, Kolhapur, Maharashtra, India.
4Assistant Professor, Department of Commerce and Management, Shivaji University, Kolhapur, Maharastra, India. Email: ttm.50649@unishivaji.ac.in

Abstract—
In this study, we will analyse EMG data using ml techniques. Electromyography (EMG) is a method employed in biomedical & biomechanics studies to quantify electrical impulses in myocytes. The mission's overarching goal is to create and assess ml algorithms for EMG signal analysis & processing, with the expectation that this will lead to better diagnostic and therapeutic results. Several unsupervised & supervised ml techniques will be investigated in this study to better identify trends in EMG information and make accurate findings. The findings of this study will enhance clinical & biomechanics studies by increasing our comprehension of the possibilities of ml in the processing of EMG data.

Keywords—: ML, EMG, RF, NN, Supervised learning, Regression

## I. INTRODUCTION

EMG is a technique that has been studied extensively for its potential medicinal and biomechanical uses. Muscular functioning & pathology may be better understood with the use of data provided by EMG, which detects brain impulses in muscles. As our ability to collect and store EMG data has improved, so has the need for more precise and time-saving techniques of processing and interpreting that data. Applying machine learning approaches to the processing of EMG data has the potential to completely reshape the way we think about studying muscles.

The purpose of this study is to examine the potential of using ml techniques to the analysis and interpretation of EMG data. The goal of this study is to enhance the reliability of predictions by discovering patterns and traits in EMG data via the use of ml techniques. Additionally, the initiative will provide light on the potential pitfalls of using ml to analyse EMG data and suggest avenues for future study in the area.

The study will also explore the various EMG signal types and their medicinal & biomechanics implications. Intramammary EMG data, for instance, give information on more profoundly activated muscles than do surface EMG signals, which are extensively employed in the study of movement patterns including muscle activity rhythms. Both kinds of EMG readings will be run through ml algorithms to see whether we can better diagnose patients and tailor our treatments.

Analysis of electromyographic (EMG) data is complicated by the fact that the signals often include unwanted noise and artefacts that might compromise the reliability of any predictions made from the data. In order to reduce the influence of noise and artefacts on the analysis outcomes, this research will also investigate several methods for pre-processing & purification of EMG signals.

In the first phase, we will choose the right ml techniques and build the EMG signal pre-processing pipeline. The next step is to test the algorithms on a number of electromyography (EMG) datasets in order to determine which approaches are most promising for analysing EMG data. Last but not least, the findings will be examined and compared to more conventional EMG analysis techniques in order to highlight the strengths and weaknesses of the ml strategy.

Benefiting both the medical & biomechanical communities, this project's findings will aid in the creation of cutting-edge techniques for analyzing EMG data. The initiative will also set the path for future studies in this area by providing meaningful information into the possibility of ml approaches for tackling complicated challenges in other domains.

## II. LITERATURE REVIEW

These last several years have seen a lot of focus on using ml for diagnosing. The Gulshan's group at the University of Texas published in JAMA in 2017 that AI could identify retinopathy in more than 100,005 retinal fundus images. There are 54 optometrists in the United States with active medical licenses, and the ai technique outperforms human doctors in terms of precision and sensitivity [2]. In 2017, Golden suggested using dl to rapidly scan pathology images for signs of lymph node metastases in breast cancer patients.

These studies have had some medicinal successes as a result of the massive data collection. Even if this technology can't replace pathologists entirely, it would greatly improve their diagnosing effectiveness & relieve some of their workload [3]. While there has been some success in using AI and ML to electromyography (EMG) data over the last several decades, this work is seldom reported in the open literature. EMG data were mostly employed for qualitative data [4, 5] from the time of the technology's inception in the eighteenth century until the turn of the past century. This fast-growing topic was reviewed in 1985 by JVBasmajian&CJDeLuca, who provided reliable criteria for practical application to aid physicians in identification [5]. Similar ideas were presented in a book by MJAminoff in 1987 [6]. Academics have put a significant amount of time and effort into doing quantitative analysis on EMG data since the turn of the century, which has greatly aided in the advancement of EMG.

TJDoherty&DWStashuk first introduced the approach and basic framework of EMG statistical studies in 2003 [7]. Since then, statistical investigations of electromyography have exploded in interest among academics.The frequency of motor units in the both proximal and distal muscles of the upper limbs was quantified and examined quantitatively by SBoe and DWStashuk in 2005 [8]. This is useful information for EMG-based gesture identification. It wasn't until 2006 that DWStashuk and Lino offered a quantitative interpretation of electromyography [9].

Numerous efforts has being undertaken to apply ml algorithms to the interpretation of EMG data during the last decade, thanks in large part to the strong development and use of ml & dl. In 2013, AbdulhamitSubasi of the Neurological Department at the of Gaziantep used the SVM approach to categorise the EMG data from the biceps muscles of 27 patients. This led to promising findings for the desktop diagnostic. To enhance classification accuracy, he suggested integrating particle swarm optimization (PSO) with SVM[10].

In 2014, Yousefii& Jamileh et al. used conventional ml techniques such as SOFM, DT, BT, ANN, as well as NFS to categorise 57 individuals into those with and without nonspecific arm pain. NFS was

shown to have a more accurate classification impact on EMG findings when compared to the other methods [11]. Review study on the role of electromyography (EMG) in aiding medical assessment and future directions in the big data age was published in 2018 by Phinyomarkk. The report also suggested a detailed framework and set of assumptions for everything from data collecting through data analysis [12].
The medical sector has benefited much from the investigations stated above, but there is still much more to learn.

To begin, these recent successes have relied heavily on dl and the massive data collection with many examples. Numerous real-world studies, however, have demonstrated that even in a prestigious hospital in a major city, each kind of test may only yield hundreds of data for an entire year after categorising the EMG exams. Consequently, it is very difficult to collect enough data to meet the volume requirements of deep learning. Because of this, efforts to study and use deep learning in the medical field have been stymied. When manually picking characteristics, classical ml methods may achieve excellent accuracy even when working with little data [13].

Therefore, conventional ml techniques have found widespread use in studies of TCM therapy and diagnosis. These research efforts have been crucial to our understanding of how to properly categorise medical conditions. Second, there is a great untapped potential for AI-related studies including face and cranial EMG data, particularly involving F-MNCS and ABR data. For the purpose of a "motor neuron conduct study," MCV is the standard academic term. This experimental apparatus (MEB-9200K) produces data relevant to studies of motor nerve conduction. As a result, this research uses a motor nerve conduction study and abides by the aforementioned electronics company naming guidelines throughout data processing and publishing (MNCS).

In this study, we use a conventional ml technique on EMG data from a very small data set to investigate the clinical relevance of this topic. To begin, over the course of ten months, data from 2,352 EMG examinations at the Sichuan Medical Center of Conventional Chinese Medical are gathered. 575 F-MNCS files and 233 ABR reports are chosen once the inclusion criteria are designed and applied. After data is cleaned, two sets are created. LR, SVM, & RFA, three of the most common techniques, are then used on each dataset in turn. As an added bonus, the processed outcomes of four algorithms are compared and discussed in depth, with special attention paid to contrasting the effects in the scenarios with and without data standardisation. In addition, it is discovered that the RFA may provide a rating of the characteristics' relevance. One last conclusion: random forest is the best algorithm for CAD

## III. EXISTING SYSTEM
Visual assessment of impulses, signals filter, & extraction of features are only few examples of the manual or semi-automated approaches used by current EMG systems to process data. Although there has been much success with these techniques, they do have certain restrictions.

The subjective nature of manual examination is one of the major drawbacks, since it might introduce inconsistencies and inaccuracies into the analysis. The manual tasks are also more likely to result in delays and mistakes.

The current systems also have the flaw of not making optimal use of the abundant EMG data that is now accessible. Traditional approaches have inherent limitations because to their dependence on pre-defined characteristics and criteria, which may not reflect the intricate connections and patterns in the information. The analysis's capacity for discovery is stunted and subpar outcomes are possible as a consequence.

In addition, the current methods are not optimal for handling the storage & interpretation of big and complicated information, like those produced by multi-channel EMG observations. Such information might be extremely demanding, necessitating substantial resources for analysis and processing.

Considering the current state of affairs, it is clear that new and better approaches to EMG data processing are required. Ml approaches may be able to help researchers analyse EMG data more precisely and quickly, therefore overcoming these restrictions. Insights gained from this project's findings will help researchers in the future realise the full capability of ml for EMG data interpretation.

## IV. PROPOSED SYSTEM

The suggested system will use ml to analyse EMG data. In order to increase the reliability of predictions, we will use machine learning techniques to analyse EMG data for recurring patterns and characteristics. As opposed to the current approaches, the suggested system will offer several benefits, such as:

1) Reducing variability & volatility in analysis, ml algorithms are designed to be impartial and immune to human biases.
2) The proposed system would automated a number of the procedures in analysing EMG data, such as signal analysis, extraction of features, including classifications, cutting down on the amount of time and effort spent on the process.
3) Since the suggested system is scalable, it can process and analyse multi-channel EMG signals without slowing down.
4) Accuracy Enhancement ML algorithms may "learn" from data and quickly uncover complicated patterns and correlations that leading to more precise and insightful analytical findings.

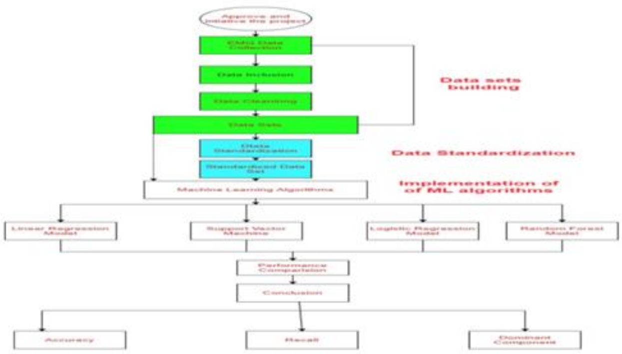



Fig.1. Our architecture

## V. METHODOLOGY

1) Data Collection -

Here, we'll learn how to collect EMG datasets for future study. Both publicly accessible datasets and original studies will be mined for their data.

To begin, the 2352 EMG medical records collected from the Regional Hospital. It has taken around ten months to collect these over two thousand three hundred reports. All information supplied is raw EMG data. After that, send along your Data Set.

2) Pre-processing the data -

Cleansing & pre-processing the EMG signals to eliminate noise and artefacts in order to get them ready for analysis is what this section is all about. Filtration, normalising the signal, and artefact reduction will be used if necessary.

When the data set has been collected, it will be put through a data cleaning procedure to look for missing or duplicate information and erroneous values. Processing of raw data from the F-MNCS. Total Size of Data Set is 794kb.

3) Extraction of features -

Here, we'll features extracted from the EMG signals that are meaningful for representing the data in a way that can be fed into ml algorithms. Features will be chosen according on the task at hand and the ml methods used.

4) Model Selection -

During this phase of the process, we will determine which ml algorithms will be used to conduct the study. Algorithms will be chosen depending on the nature of the EMG signals to be analysed, the desired outcomes of the analysis, and the findings of past research.

In the first step SVM algorithm is implemented to create the decision boundary that can aggregate data into classes. In the second step RFA is implemented to improve the predictive accuracy of the data set. In third step LR is implemented to solve the classification problems. In fourth step Linear regression shows a linear relationship between a dependent (y) and one or more independent (y) variables.

A. SVM -

There is the option of using a ml method called a SVM to analyse the EMG data collected for this project. For both regression & classification, SVM, as a SL algorithm, is a useful tool.

SVM may be used to sort EMG signals into several groups according to their properties, which is useful for the project at hand. EMG signals, for instance, may be categorized using SVM into many categories, such as extending, gripping, and moving. The primary benefit of SVM is that it can determine the optimal class-dividing line using the maximum margin criteria. This helps to enhance the model's generalisation performance by decreasing the likelihood of overfitting. To describe non-linear connections in the data, SVM provides a set of kernel parameters that map the input data to a higher-D pace. This makes SVM a useful tool for analyzing and processing non-linear, complicated inputs like EMG signals.

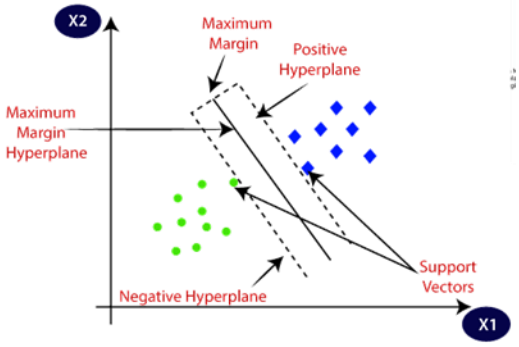



Fig.2. SVM network

B. RFA -

The project might employ the ml method known as RF to analyse the EMG data. For both regression and classification problems you may turn to RF, an ensemble technique.

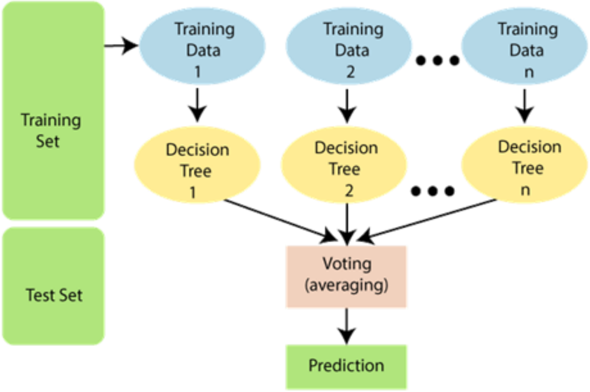



Fig.3. RFA network

Random Forest may be used to sort EMG signals into several groups according to their properties, which is useful for our research. Stretching, gripping, & striding are just some of the motions that may be identified from EMG data using Random Forest.

Random Forest's key benefit is that it can deal with huge and complicated datasets while minimising the possibility of overfitting. The predictions from each individual decision tree are combined in a RF to get a final forecast. Overfitting is mitigated and the model's accuracy & stability are improved by this method. The amount of trees in a forest, the height of the trees, and the amount of variables employed at each split are just some of the parameters in RF that may be tweaked to improve the model's efficiency.

C. Linear Regression -

The research might benefit from the application of the ml method known as linear regression to analyse the EMG data. For regression problems in which a continuous result is to be predicted from a collection of input characteristics, a supervised learning approach called linear regression is often utilised.

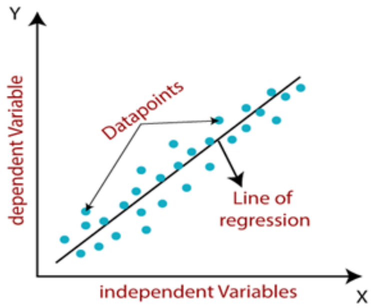



Fig.4. Regression’s Network

Linear Regression may be used to describe the association between EMG data and the corresponding movement or activity as it pertains to this project. As an illustration,

Function of Linear Regression -

Linear Regression may be used to foretell the path of an arm's reaching motion in response to electromyographic (EMG) data. Linear regression's key benefit is that it is easy to understand and use. A smooth line represents the link between the input data and the output in a regression model, which simplifies the analysis and interpretation of the findings. It is also relatively effective and works well with sets that have a limited number of input attributes. Since EMG signals often consist of numerous channels, the flexibility of the algorithm makes it a helpful tool for analyzing and processing these signals.

D. Logistic Regression -

Logistic Regression's key benefit is the ease with which it may be understood and used. With Logistic Regression, the findings may be easily understood and interpreted since a logistic curve is used to describe the connection among the input data and the outcome.

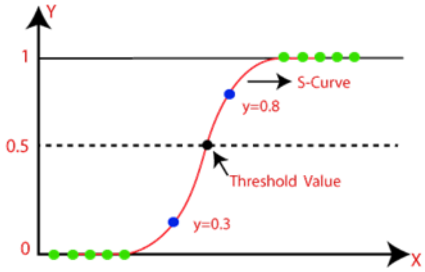



Fig.5. LR network

E. Decision Tree–

A decision tree is a machine learning model that can be used for classification and regression problems. It involves partitioning the feature space into smaller subsets based on the values of the features. The decision on which feature to split on and the split value is made using an impurity measure, such as Gini impurity or entropy. There are different algorithms for building a decision tree, such as ID3, C4.5, and CART. Each algorithm differs in the way they select the splitting feature and the split value, as well as the stopping criteria for tree construction. The tree structure can be easily visualized and understood, making them useful for explaining the decision-making process to stakeholders. However, decision trees can be prone to overfitting if the tree is too deep or if the stopping criteria are not well-defined. To avoid overfitting, techniques such as pruning and setting a maximum depth can be used. Decision trees can be used as a standalone model or as a building block in more complex models, such as random forests and gradient boosting machines. Their versatility and interpretability make them a popular choice for data analysis. However, it is important to properly interpret the results and to avoid overfitting.

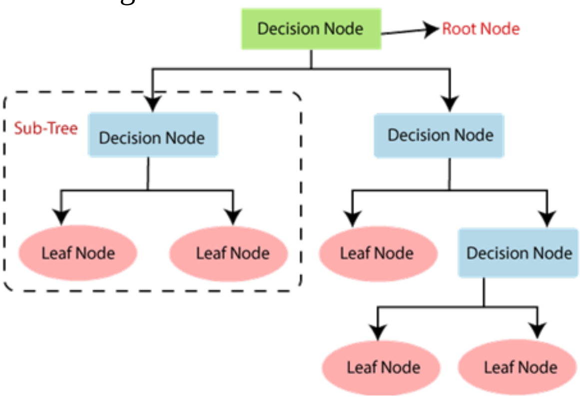



Fig.6. Decision Tree network

5) Model Training -

In this stage, the EMG samples will be pre-processed, and the ml algorithms will be trained on those datasets. The data will be used to teach the algorithms how to recognise patterns and correlations, allowing them to then make predictions.

6) Validation -

In this section, we will assess the efficacy of the ml models we have developed. The algorithms' precision and applicability will be evaluated by applying them to data sets that have not been used in development thus far.

When all of these algorithms are combined, RFA produces the highest accuracy.

# VI. RESULTS

Accuracy of the comparative models are as follows –

| Model | Accuracy |
| --- | --- |
| SVM | 94.26523297491039 |
| Random Forest | 98.56630824372759 |
| Logistic Regression | 98.2078853046595 |
| Linear Regression | 27.2 |

Using a ML approach, the RFA achieved the maximum accuracy of 98.56 when analysing electromyographic data. Classifying motions and spotting patterns in muscle activity may be accomplished with this strategy, yielding important insights into the data and guiding future investigations.

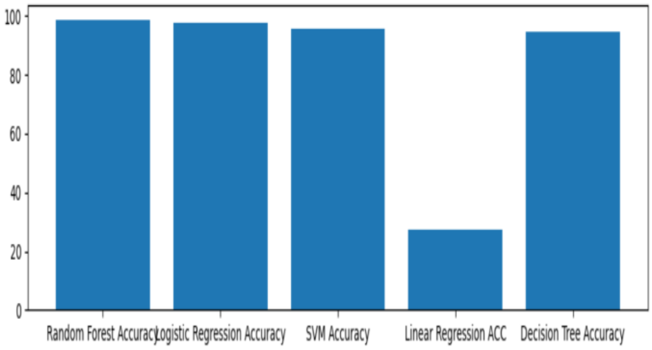
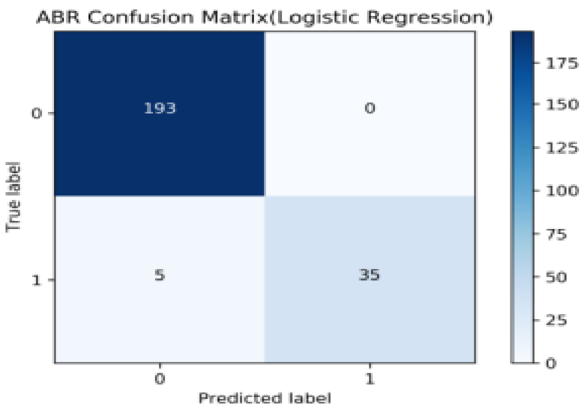
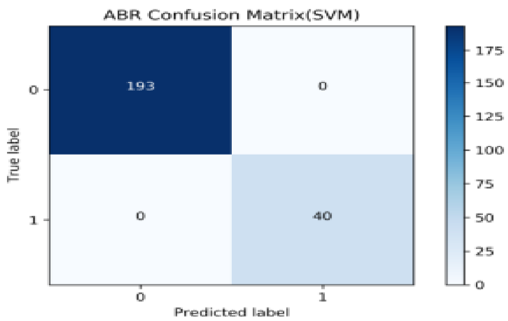


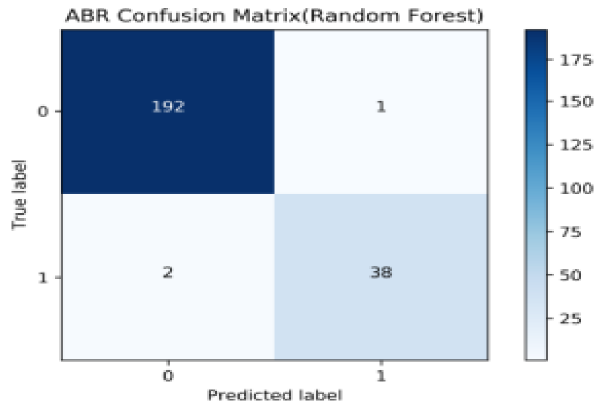

Fig.6. Accuracy comparison graph




**(a) Confusion matrix of logistic regression**




**(b) Confusion matrix of SVM**




(c) **Confusion matrix of random forest**

## VII. CONCLUSION & FUTURE WORK

Using a ml approach, the study examined EMG data, finding that the RFA best distinguished between motion types and identified patterns in muscle movement. The results show the promise of ml for EMG data processing.

In light of the encouraging findings of the current investigation, more research and development are warranted. To improve the model's precision, one may, for instance, investigate other ml techniques or add more features. To further verify the model's efficacy, it might be applied to additional demographics or kinds of EMG data. Signal processing is only one example of a technology that might be used with the models to provide a more thorough and in-depth examination of the data.

## REFERENCES

[1] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, ‘‘Deep learning for healthcare: Review, opportunities and challenges,’’ Briefings Bioinformatics, vol. 19, no. 6, pp. 1236–1246, Nov. 2018.
[2] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, ‘‘Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,’’ JAMA, vol. 316, no. 22, p. 2402, Dec. 2016.
[3] J. A. Golden, ‘‘Deep learning algorithms for detection of lymph node metastases from breast cancer: Helping artificial intelligence be seen,’’ JAMA, vol. 318, no. 22, p. 2184, Dec. 2017.
[4] M. J. Aminoff, Electromyography in Clinical Practice: Clinical and Electrodiagnostic Aspects of Neuromuscular Disease, 2nd ed. New York, NY, USA: Churchill Livingstone, 1987.
[5] J. V. Basmajian, C. De Luca, and C. J. Deluca, Muscles Alive: Their Functions Revealed by Electromyography, 5th ed. Baltimore, MD, USA: William & Wilkins, 1985, pp. 1–20.
[6] M. J. Aminoff, Electromyography in Clinical Practice: Clinical and Electrodiagnostic Aspects of Neuromuscular Disease, 2nd ed. New York, NY, USA: Churchill Livingstone, 1987, pp. 3–15.
[7] T. J. Doherty and D. W. Stashuk, ‘‘Decomposition-based quantitative electromyography: Methods and initial normative data in five muscles,’’ Muscle Nerve, vol. 28, no. 2, pp. 204–211, Aug. 2003.
[8] S. G. Boe, D. W. Stashuk, and T. J. Doherty, ‘‘Within-subject reliability of motor unit number estimates and quantitative motor unit analysis in a distal and proximal upper limb muscle,’’ Clin. Neurophysiol., vol. 117, no. 3, pp. 596–603, Mar. 2006, doi: 10.1016/j.clinph.2005.10.021.
[9] D. W. Stashuk, L. Pino, A. Hamilton-Wright, T. Doherty, and S. Boe, ‘‘Interpretation of QEMG data,’’ in Proc. Gen. Meeting Amer. Assoc. Neuromuscular Electrodiagnostic Med. (AANEM), Phoenix, AZ, USA, 2007.
[10] A. Subasi, ‘‘Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders,’’ Comput. Biol. Med., vol. 43, no. 5, pp. 576–586, Jun. 2013.

[11] J. Yousefi and A. Hamilton-Wright, ‘‘Characterizing EMG data using machine-learning tools,’’ Comput. Biol. Med., vol. 51, pp. 1–13, Aug. 2014.

[12] A. Phinyomark and E. Scheme, ‘‘EMG pattern recognition in the era of big data and deep learning,’’ Big Data Cogn. Comput., vol. 2, no. 3, p. 21, Aug. 2018.

[13] T. Ziemniak, ‘‘Use of machine learning classification techniques to detect atypical behavior in medical applications,’’ in Proc. 6th Int. Conf. IT Secur. Incident Manage. IT Forensics, May 2011.