Analyzing Neuromuscular Disorder Patients with sEMG Sensor using Big Data

C. S. Indhuja[1], R. Gomathi Leela[2], S. Akshaya[3]
[1,2,3]*UG Student, Department of Computer Science and Engineering, SRM Valliammai Engineering College, Chennai, India*

***Abstract*: Neuromuscular disorders affect the nerves that control voluntary muscles and the nerves that communicate sensory information back to the brain. The peripheral nervous system, where nerve cells send messages that control muscles in order to allow movements, is affected by neurodegenerative disease. Muscles cannot be controlled by sick neurons. In this paper we propose a novel system to monitor neuromuscular diseases using EMG sensor. EMG testing has a variety of clinical and biomedical applications. The Electromyography Sensor (EMG) allows the user to measure the electrical activity of muscles. It is a non-invasive technique adopted for diagnosis of different neuromuscular disease. Through this system we can analyze the various readings from sEMG sensor during pre-defined handwriting tasks. A near accurate prediction is achieved by training analytics algorithm with historic data.**

***Keywords*: Analytics, Big Data, Healthcare, Neuromuscular disorder, sEMG sensor.**

1. Introduction

The recent inclusion and implementation of sensors in the medical field has revolutionized telemedicine. The growth in medical sensors is due to the fact that the past decade has been one of digital medicine innovation. With increasing use of digital technologies to provide improved health care, facilities have augmented electronic devices that help in collection of medical data.

The recent emergence of sensors in such large scale at affordable rates has allowed the growth of IoT at such a rapid pace. The rise of IoT has paved way to accumulation of large data sets through the data being generated from sensors. Processing, and analyzing such large stored data is becoming more difficult.

Amongst the huge repositories of data accumulated, some tend to lose it's intrinsic value if not used immediately. Such data sets are processed through edge computing. Processing these extremely large data sets and analyzing them to reveal patterns, associations and trends relating to human interactions and behavior. Big data is a large collection of structured and unstructured data obtained from various sources that inundates a business on a day-to-day basis.

Beyond the size of the data, what matters is how different organizations utilize them. Analyzing big data provides insights that help make better strategic decisions. The concept of Big data gained momentum in the early 2000s when industry analyst Dough Laney articulated the now mainstreamed definition of big data as the three V's: Volume, velocity and variety.

A. Big Data Analytics

The most prominent result obtained from processing these large data sets would be that of analytics and prediction. Big data analytics is the often complex process of examining large and varied data sets, or big data, to uncover information such as hidden patterns, unknown correlations, market trends and customer preferences that can help organizations make informed business decisions.

There are five fundamental algorithms for advanced analytics. This includes Linear regression, Logistics regression, Classification and Regression Trees, K-nearest neighbors and K-means clustering. Let us look at these in a bit detailed manner.

1) Linear Regression

One of the most basic algorithm of advanced analytics is linear regression. One can easily visualize it's working and how the input data and output data are related. It uses the relationship between two sets of continuous quantitative measures.

2) Logistic Regression

Although it sounds similar to linear regression, it actually focuses on problems involving categorization instead of quantitative forecasting. The goal is to categorize if an instance if an input variable fits within a category or not.

3) Classification and Regression trees

A decision to categorize data is used by classification and regression trees. A question related to one of the input variables makes each decision. With each question and corresponding response, the instance of data gets moved closer to being categorized in a specific way.

4) K- nearest neighbor

K-nearest neighbor is also a classification algorithm. It is known as a "lazy learner" because the training phase of the process is very limited. The learning process is composed of the training set of data being stored.

5) K-means clustering

K-means clustering focuses on creating groups of related attributes. These groups are referred to as clusters. Once these clusters are created, other instances can be evaluated against them to see where they best fit. This technique is often used as part of data exploration.

B. Big Data Tools

With the exponential growth of data, numerous types of data, i.e., structured, semi-structured, and unstructured, are producing in a large volume. Therefore, to manage these growing data in a traditional RDBMS system quite impossible. Here, we outline some of the major Big data software.

1) Apache Hadoop

It is one of the most prominent tools. This permits reliable distributed processing of large volume of data in a dataset across clusters of computers.

2) Quoble

Quoble is the cloud-native data platform which develops machine learning model at an enterprise scale. The vision of this tool is to focus on data activation.

3) HPCC

LexisNexis Risk Solution develops HPCC. This open source tool provides a single platform, single architecture for data processing. It is easy to learn, update, and program. Additionally, easy to integrate data and manage clusters.

4) Cassandra

Apache Cassandra is a big data tool which will you provide scalability and high availability as well as excellent performance. This tool is a free, open source, NoSQL distributed database management system.

5) MongoDB

This Database Management tool, MongoDB, is a cross-platform document database that provides some facilities for querying and indexing such as high performance, high availability, and scalability.

C. EMG Sensors

There are two types of EMG sensors; namely, surface EMG and intramuscular EMG. Muscle function is assessed using sEMG by recording muscle activity from the surface above the muscle on the skin. However, they tend to provide minimal level of assessment of muscle activity. Two or more electrodes are used in sEMG to record the potential difference between two separate electrodes.

Using a variety of different types of recording electrodes intramuscular EMG can be performed. A comparison and an overview of different EMG methods for myopathic evaluation are presented in [1]. In this system we suggest the use of sEMG sensor to monitor the muscle activity.

D. Neuromuscular disorder

Progressive neuromuscular diseases causes reduced physical activity which negatively impacts quality of life and health outcomes. Chronic respiratory failure occurs as a consequence of neuromuscular disease in the long-term.

Studies show that long-term noninvasive mechanical ventilation (NIV) improves symptoms, gas exchange, quality of life and survival. NIV improved these parameters in muscular dystrophies and also in patients with amyotrophic lateral sclerosis without severe bulbar dysfunction. [12]. Your muscles tend to become weak and waste away due to neuromuscular disorders. Symptoms such as spasms, twitching, and pain may also be prevalent.

We have selected neuropathy and myopathy for our study from a large number of neuromuscular disorders, as their consistency of clinical appearance is good. [2]

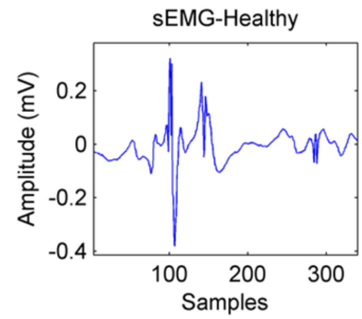

Fig. 1. sEMG reading of healthy controls

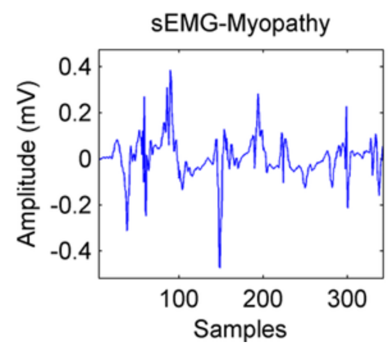

Fig. 2. sEMG reading of patients with myopathy

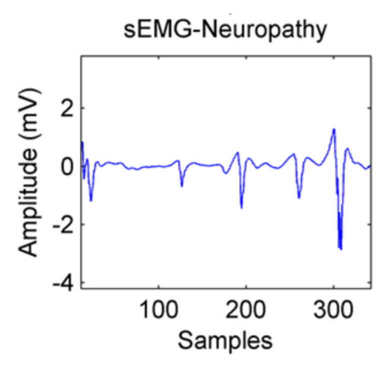

Fig. 3. sEMG reading of patients with neuropathy

Surface EMG data obtained from four different datasets comprised of forty subjects were analyzed in this study. These datasets were collected independently at different institutes and are publicly available [8]-[11].

In patients with respiratory muscle weakness caused by neuromuscular disorders severe nocturnal hypoxemia may occur. Negative pressure ventilators causes upper airway obstructive apneas which are only partially effective in these patients.

The effectiveness of positive pressure ventilation was examined through a nose mask in preventing nocturnal hypoxemia and was compared with negative pressure systems. It can be reasoned that nasal positive pressure will provide stability for the upper airway.

Handwriting requires precision, coordination, abilities to execute and planning cognitively. Significant changes in handwriting performance are a prominent features in patients with neuromuscular disease. A special role in the context of neurodegenerative disease assessment can be covered by handwriting.

Cerebral cortex, basal ganglia, and cerebellum are involved in learning and performing handwriting [3]—complex activity entailing cognitive, kinaesthetic, and perceptual-motor components.

Several advances have already been made in the offline (static) domain [4] but, nowadays, online (dynamic) systems can be adopted. The main advantage of online acquisition devices is their ability to acquire kinematics (dynamics) of the writing process which are lost in offline systems.

2. Data Acquisition

There are different factors that must be considered: Gathering participants, selection of data acquisition device and choosing the right handwriting task.

A. *Gathering participants*

Participants must include a set of patients in various stages of disease. They should also be monitored during their medication hours and when they are off. Also we must include a set of healthy patients to differentiate readings from healthy data set and affected patients.

B. *Selection of data acquisition device*

A wide set of devices for data acquisition is available. In some situations the use of an electronic pen on a digital screen could be unusual or unfamiliar to patients, so writing with an ink pen on paper fixed to the tablet may be an option [5], [3].

C. *Choosing the right handwriting task*

A complex feedback system is involved in the writing process. It implicates the participation of several cognitive motor processes. The tasks can be categorized into three groups: Simple writing and drawing tasks and Complex tasks. Simple drawing tasks include Spirals, meanders, straight lines. Simple writing tasks are nonsense words containing one or more character repetitions. Complex functional writing tasks involve tasks such as copying details of a bank check into the right fields.

D. *Historical Data set*

To analyze and predict patterns obtained from big data, we need to train the algorithm with existing data sets. These are known as historical data set. Data sets are of various types. They could be Image data, Text data, Sound data, Signal data, Physical data, Biological data etc., Such training data sets can be obtained from labs, EHR or various data set sources from the internet.

3. Preprocessing and Feature extraction

Standard signal processing algorithms are used to enhance the raw data acquired by the device. These processes include Filtering, noise reduction and smoothing. However, in this domain, it would lead to the loss of information related to the time spent by each participant in performing a specific task (that is a discriminative feature). Given this consideration, it is quite usual to not adopt preprocessing steps (e.g., [3]). In the process of feature extraction, two categories are considered; functional and parameter features namely.

A. *Functional features*

The most common function features are: position in terms of (x,y) coordinates, time stamp, button status, pressure, azimuth, altitude, displacement, velocity, and acceleration. Some of these features are directly conveyed by the acquisition device, whereas others are numerically derived.

B. *Parameter features*

In this case, the trait is characterized as a vector of elements, each one representative of the value of a feature. Parameter features are obtained by means of transformations on the function features. This also includes data set from sEMG sensors while performing the handwriting tasks.

Many of the above-reported parameters have been normalized, based on the total time duration of the task or stroke. Finally, in order to reduce data dimensionality and to select the most discriminating features, well-known feature selection schema have been adopted, such as the Mann-Whitney U-test [6] and the Relief algorithm [6], [7].

4. Analytics

Upon completion of pre-processing the obtained data, we can further move to the next step of performing analysis. In order to perform analytics we must choose the most appropriate algorithm.

For this purpose, the most suited algorithm would be K-means clustering. However, for the algorithm to cluster the data obtained in real-time, it must have training for the same. Algorithms are trained with the help of various historic data.

Historic data is fed into the system. This data is then clustered based on the plotting of data points. We can then tag the various clusters. To further train the system, a certain percentage of dataset is withheld to check the accuracy of the system.

The output of this withheld data set is then verified and corrections are made for the same. This way we can ensure a prediction system with at most accuracy.

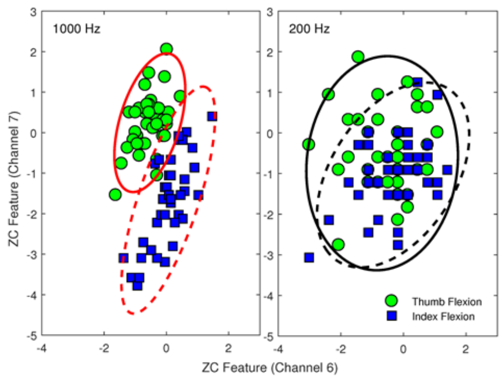

Fig. 4. K-means clustering of sEMG data of Patients performing Handwriting task.

5. Conclusion and Future Scope

Neuromuscular disorders are acquired or inherited (genetic) conditions that affect some part of the neuromuscular system. They tend to be progressive in nature. Hence they result in muscle weakness and fatigue. There is no cure for most neuromuscular disorders, but some can be effectively managed and treated.

Current interventions include; drug therapy, referrals to appropriate specialists and patient and family education and counselling. Introducing telemedicine system to monitor could prove to be path breaking. The research provided in this paper sets a foreground for mechanism and data procurement for the same.

References

[1] C. D. Katsis, Y. Goletsis, A. Likas, D. I. Fotiadis, and I. Sarmas, “A novel method for automated EMG decomposition and MUAP classification,” Artificial Intelligence in Medicine, vol. 37, pp. 55-64, 2006.

[2] Classification of EMG Signals for Assessment of Neuromuscular Disorders Anjana, International Journal of Electronics and Electrical Engineering Vol. 2, No. 3, September 2014.

[3] P. Werner, S. Rosenblum, G. Bar-On, J. Heinik, A. Korczyn, "Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment", J. Gerontology Ser. B Psychological Sci. Social Sci., vol. 61, no. 4, pp. 228-236, 2006.

[4] N. Zhi, B.K. Jaeger, A. Gouldstone, R. Sipahi, S. Frank, "Toward monitoring Parkinson's though analysis of static handwriting samples: A quantitative analytical framework", IEEE J. Biomed. Health Informat., vol. 21, no. 2, pp. 488-495, Mar. 2017.

[5] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, I. Schlesinger, "Handwriting as an objective tool for Parkinson's disease diagnosis", J. Neurol., vol. 260, no. 9, pp. 2357-2361, 2013.

[6] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, M. Faundez-Zanuy, "Decision support framework for Parkinson's disease based on novel handwriting markers", IEEE Trans. Neural Syst. Rehabil. Eng., vol. 23, no. 3, pp. 508-516, May 2015.

[7] P. Zham, S. Arjunan, S. Raghav, D. K. Kumar, "Efficacy of guided spiral drawing in the classification of Parkinson's disease", IEEE J. Biomed. Health Informat.

[8] Khushaba, R.N.; Takruri, M.; Miro, J.V.; Kodagoda, S. Towards limb position invariant myoelectric pattern recognition using time-dependent spectral features. Neural Netw. 2014, 55, 42–58.

[9] Khushaba, R.N.; Al-Timemy, A.; Kodagoda, S.; Nazarpour, K. Combined influence of forearm orientation and muscular contraction on EMG pattern recognition. Expert Syst. Appl. 2016, 61, 154–161.

[10] Al-Timemy, A.H.; Khushaba, R.N.; Bugmann, G.; Escudero, J. Improving the Performance Against Force Variation of EMG Controlled Multifunctional Upper-Limb Prostheses for Transradial Amputees. IEEE Trans. Neural Syst. Rehabil. Eng. 2016, 24, 650–661.

[11] Pizzolato, S.; Tagliapietra, L.; Cognolato, M.; Reggiani, M.; Müller, H.; Atzori, M. Comparison of six electromyography acquisition setups on hand movement classification tasks. PLoS ONE 2017, 12.

[12] Chronic respiratory care for neuromuscular diseases in adults N. Ambrosino, N. Carpenè, M. Gherardi European Respiratory Journal 2009 34: 444-451.