Classification of Electromyogram Using Vertical Visibility Algorithm with Support Vector Machine

Patcharin Artameeyanant[a], Sivarit Sultornsanee[b] Kosin Chamnongthai[a] and Kohji Higuchi[c]
[a] Department of Electronic and Telecommunication Engineering
Faculty of Engineering
King Mongkut's University of Technology Thonburi
126 Pracha Uthit Rd., Bang Mod, Thung Khru, Bangkok 10140, Thailand
koyka_patcharin@hotmail.com, kosin.cha@kmutt.ac.th
[b] School of Business
University of the Thai Chamber of Commerce
126/1 Vibhavadi Rd., Dindang, Bangkok 10400, Thailand
sivarit_sul@utcc.ac.th
[c]Department of Electronic Engineering
The University of Electro-Communications
1-5-1 Chofu-ga-oka, Chofu, Tokyo 182-8585, Japan
higuchi@ee.uec.ac.jp

Abstract— **Analyzing the electromyogram is an important issue on diagnosis of neuromuscular diseases. The classification of electromyogram signal plays a significant role in this issue. Since the characteristic of the signals is complex and non-stationary, so the complex network is an appropriate tool in extracting feature of the signal. In this paper we propose a novel feature extraction technique based on transforming the signal to complex network via vertical visibility algorithm. Characteristic on the measurements of community structure and distance property are examined. The pattern on the relationship of nodes in the network is investigated. Support vector machine was employed for classification. The proposed method can classify the signals into 3 cases, i.e., healthy, myopathy, and neuropathy, with remarkable experimental results.**

Index Terms—EMG Signal, Complex Network, Vertical Visibility Algorithm, Community Structure.

I. INTRODUCTION

The neuromuscular system composes of the nervous system and the muscular system. The neuromuscular disorders which usually affect spinal cord, nerves, or muscles can be considered into muscular (myopathy) and neuronal (neuropathy) disorders. The primary symptom of myopathy is muscle weakness due to dysfunction of muscle fibres. Neuropathy affects the nervous system which transmits information to the nerves and spinal cord. The damage ranges from loss of strength to amputation due to neuron or muscle death.

Electromyogram (EMG) records electrical activity which is the tool in used for diagnosing patients with neuromuscular disorders. The analysis of EMG signal is generally carried out by highly trained neurologists, which is short in demand. The development of the automated diagnostic systems based on EMG readings is crucial. Time, expense, and life could be saved. In the past, many researches had been conducted as follow. Hu et al. (2005) [1] proposed wavelet packet transform to transform the EMG signal into several sub-bands. The features were extracted from relative wavelet packet energy (RWPE) on these sub-bands. Compared with the conventional method, the classified result is improved greatly. Subasi et al. (2006) [2] used an autoregression (AR) model of EMG signals as an input to the classifier. The feed-forward error-back propagation artificial neural networks (FEBANN) and wavelet neural networks (WNN) based classifiers were employed. The accuracy yields 90.7% for WNN and 88% for FEBANN. Sultornsanee et al. (2011) [3] introduced recurrence quantification analysis and support vector machines in analyzing EMG signals. The overall accuracy yields 98.28%. Subasi (2013) [4] decomposed EMG signals into frequency sub-bands using discrete wavelet transform (DWT) and classified with PSO-SVM model which is the hybrid between particle swarm optimization (PSO) and support vector machine (SVM). The total classification accuracy yields around 97.41%.

An EMG signal is highly non-linear and non-stationary signal. The non-linear signal can be transformed to complex networks utilizing visibility algorithm. Thus, EMG signal can be manipulated as linear system by means of complex network. Lacasa et al. (2007) [5] proposed the visibility algorithm to convert time series signal to a graph. The constructed graph inherited several properties of the series in its structure. Luque et al. (2009) [6] employed the horizontal visibility algorithm, a geometrically simpler and analytically solvable version of the visibility algorithm. The exact results were obtained on measurement mechanics of the degree distribution, the clustering coefficient, and the mean path length. The algorithm discriminated randomness in the time series and also reduced the complexity of the network. Campanharo et al. (2011) [7] studied duality between time series and networks and proposed a map of different time series results in the networks with distinct topological properties.

In this paper, we introduce a novel feature extraction method of EMG signal based on measurement mechanics of vertical visibility algorithm employing support vector machine for classification. The proposed work overcomes the non-linearity of the signal and the computational complexity of the data. EMG signals are transformed to complex network utilizing horizontal visibility algorithm. The characteristic of the signals are preserved, regardless differences on strength and energy of the signal. This is so because the algorithm utilizes the relative amplitude of the signal to create complex network. The adjacency matrix which is a computational-capable form in a linear environment of the complex network is then employed for measurement mechanics calculation. The feature vectors are formed from these measurement mechanics for classification via support vector machine. The proposed method shows outstanding results with high overall accuracy.

II. RELATED METERIALS AND PROPOSED METHOD

The following sections are the related materials in used and the proposed method.

A. Data Set

Data were collected with a Medelec Synergy N2 EMG Monitoring System (Oxford Instruments Medical, Old Woking, United Kingdom). A 25mm concentric needle electrode was placed into the tibialis anterior muscle of each subject. The patient was then asked to dorsiflex the foot gently against resistance. The needle electrode was repositioned until motor unit potentials with a rapid rise time were identified. Data were then collected for several seconds, at which point the patient was asked to relax and the needle removed.

The data were recorded at 50 KHz and then downsampled to 4 KHz. During the recording process two analog filters were used: a 20 Hz high-pass filter and a 5K Hz low-pass filter. The data have 3 groups, Healthy, Myopathy, and Neuropathy that show in Fig.1.

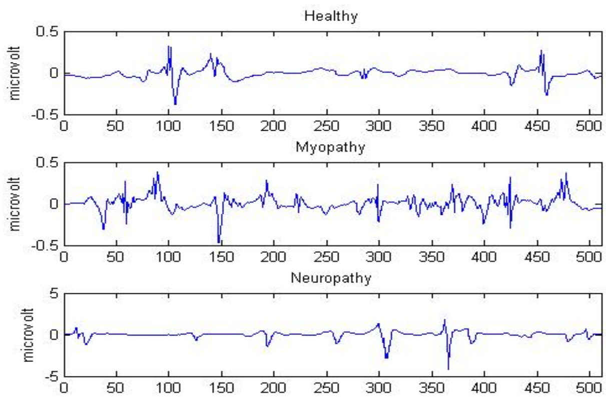


Fig. 1 Sample of 512 points recorded from the EMG signals of Healthy, Myopathy, and Neuropathy

B. Complex Network and vertical visibility graph

A complex network involves methods developed in a field of mathematics referred to as the graph-theory. Mathematical structures called graphs are used to model pairwise connections between components of a network. A complex network is thus represented as graph G = (N, L), where N denotes a set of components and L is a set of connections among them. The components are called nodes (vertices) and the connections among them are called links (edges). We only consider unweighted and undirected graphs, meaning that there is no distinction how two components are connected, or in terms of the graph theory, how a link may be directed from one node to another. In addition, a graph cannot contain self-loops or connections beginning and ending at the same component. In this way, the complex network as a system of nontrivial interconnected components is modeled as a graph G = (N, L) with respective N = |N| nodes and L = |L| links. The complex network can be described as a graph of N nodes connected by L links and can represented by an adjacency matrix. The adjacency matrix A of a simple graph is a matrix with elements A_{ij} which are defined by (Newman, 2010) [8]:

$$A_{ij} = \begin{cases} 1 & x_i, x_j > x_n \text{ for all n where } i < n < j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

the adjacency matrix of an undirected network is symmetric, i.e. $A_{ij} = A_{ji}$.

Vertical Visibility Graph : A vertical visibility algorithm is a simplification of visibility graphs and is always a sub-graph of its associated visibility graphs. The vertical visibility graph is obtained from the mapping of a time series into a graph according to the following criterion. Let $\{x_i\}$, i = 1, 2, 3, …, N be a time series of N data. Each datum of the series is assigned to be a node in a visibility graph. Two nodes i and j in the graph are connected if the following geometrical criterion is fulfilled. A straight line that was drawn in the time series joining x_i and x_j without x_n higher than this line for all n where i < n < j.and . $A_{ij} = A_{ji}$. See Fig. 2 for an illustration.

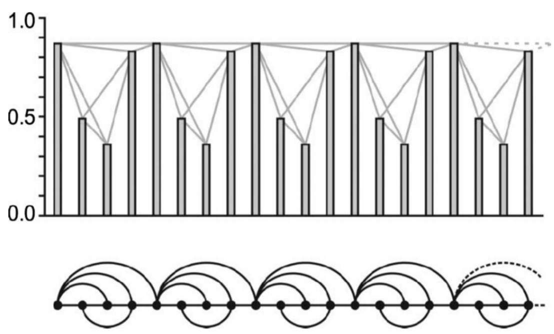


Fig. 2 Example of vertical visibility algorithm

C. Statistical Mechanics

The statistical mechanics, which is the measurement of complex network, can be obtained from manipulation of the adjacency matrix. In our work, we chose measurements based on the statistical mechanics of an intermediate level. The considered measurements are briefly reviewed as follow.

Average Degrees (AD) The average degree indicates the average number of links connecting a node in a network to the other nodes. The average degree of the node in a graph is defined as 2 times the number of links divided by the number of nodes (Barabasi, 2012)[9].

$$AD = \frac{2L}{N} \tag{2}$$

where L is the number of links and represents the total number of interactions between nodes.

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i \tag{3}$$

where k_i is the degree of the i^{th} node in the network.
N is the size of the network

Average Clustering Coefficient (ACC) : The average clustering coefficient represents the relationship between the nodes in a complex network. The degree of clustering of the whole network is captured by the average clustering coefficient, corresponding to an average of C_i over all nodes i=1,…, N (Barabasi, 2012) [9].

$$ACC = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{4}$$

Where C_i is the local cluster coefficient $C_i = \frac{2L_i}{k_i(k_i-1)}$

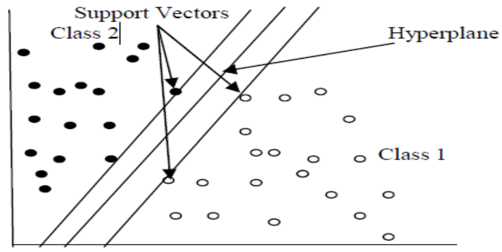
Transitivity (T) : Transitivity is a measure of the presence of a heightened number of triangles in the network – sets of three vertices, each of which is connected to the others. T is the fraction of triples that have their third edge filled in to complete the triangle (Newman, 2003) [8].

$$T = \frac{3 \times \text{number of triangles in network}}{\text{number of connected triples of verticed}} \tag{5}$$

where a "connected triple" is a single vertex with edges connecting to an unordered pair of vertices.

D. Classification

SVM is a supervised machine learning paradigm capable of solving linear and non-linear classification and regression problems. Due to its accuracy and capability of handling a great number of predictors, it has been widely used in EMG signal classification. SVM is a supervised learning method that generates input-output mapping functions from a set of labeled training data. The goal of SVM is to produce a model (based on the training data) which classifies the test data. A classification task usually involves separating data into training and testing sets. SVM can classify data separated by non-linear and linear boundaries. SVM maps input vectors to a higher dimensional space where a maximal separating hyperplane is constructed. Two parallel hyperplanes are constructed on each side of the hyperplane that maximizes the distance between the two parallel hyperplanes. An assumption is made that the larger the margin or distance between these parallel hyperplanes the better the generalization error of the



classifier. Fig. 3 shows a simple kind of linear SVM classifier.

Fig. 3 Simple kind of Linear SVM

E. Proposed Method

Since the EMG signal is the signal that is transmitted from the muscle. It is the deterministic signal of which frequencies are in a low range of a frequency spectrum. The chaos due to abnormal electrical activities from neuromuscular symptoms causes this deterministic signal to have an alteration. This alteration has some patterns in common which can be extracted via the vertical visibility algorithm. We should see a variation of the pattern clearly since this algorithm uses comparison criteria for the relative amplitude of the signal. The obtained result is a connection network which can transform to an adjacency matrix for computing characteristic properties of the signal. Thus, this algorithm can circumvent problems on the signal strength and energy as well. So, we propose the following system as shown in Fig. 4, which is the block diagram of the proposed method.

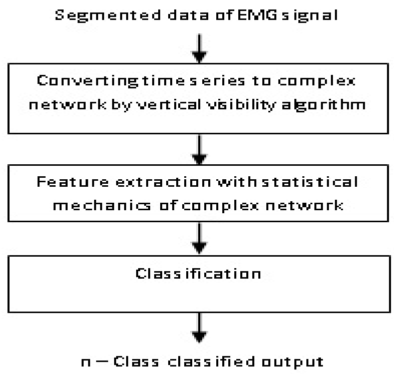


Fig.4 Block diagram of the proposed method

First, we processed the segmented 512 data point to obtain an adjacency binary undirected connection matrix by means of the vertical visibility algorithm via complex network. Second, the statistical mechanics was calculated from this adjacency matrix to form feature vector. Finally, the feature vector was classified with support vector machine. Fig.4 is the block diagram of the proposed method.

III. EXPERIMENTAL RESULTS

The result of statistical mechanics that was calculated from the adjacency matrix of the vertical visibility algorithm is in part A. The next part is the classified result that obtained from SVM classifier.

A. Feature Vector

Boxplots were generated to compare the dataset of the 250 epochs on each signal. Fig.5, 6, and 7 are the boxplot of average degree, average cluster coefficient, and transitivity, respectively. These statistical mechanics were employed to set as feature vector for classification.

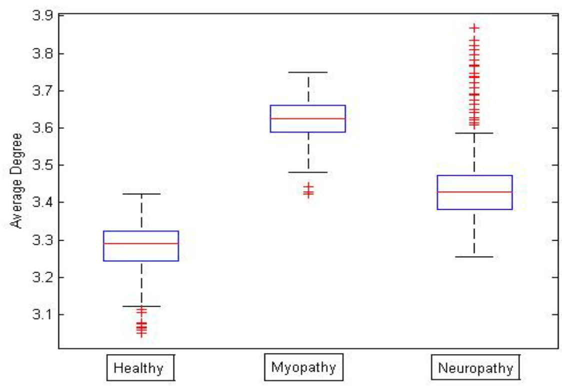


Fig. 5 Boxplot of Average Degree

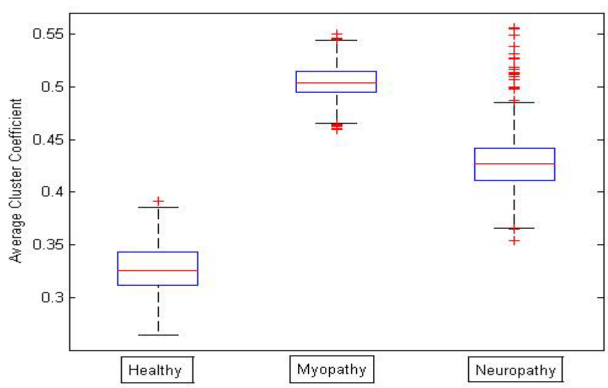


Fig. 6 Boxplot of Average Cluster Coefficient

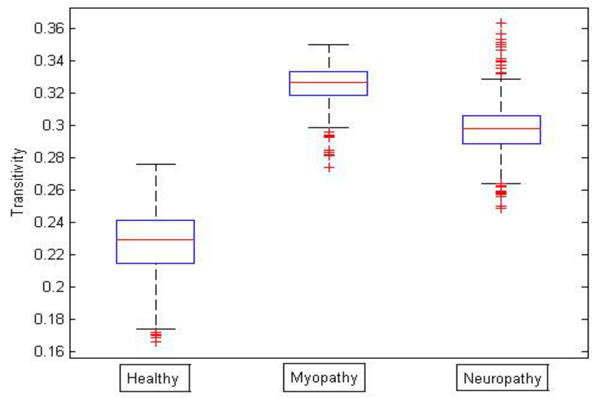


Fig. 7 Boxplot of Transitivity

B. Classification Results

On classification, we classified EMG signal into 3 categories: Healthy, Myopathy, and Neuropathy.

The performance of the classifier is evaluated by statistical measurements of sensitivity, specificity, and classification accuracy. The measure definitions are as follows :

Sensitivity (SEN):

$$SEN = TP / (TP + FN)$$

Specificity (SPE):

$$SPE = TN / (TN + FP)$$

Classification accuracy (CA):

$$CA = (TP + TN) / (TP + TN + FN + FP)$$

Where the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are used to evaluate performance of classifiers.

Table 1 is the classification result on statistical measurements as specified above on the EMG signals:

Table 1
The Values of the statistical measurement in the classification.

EMG data sets	Statistical measurement		
	Sensitivity	Specificity	Classification accuracy
Healthy	0.984	0.994	0.9907
Myopathy	0.984	0.994	0.9907
Neuropathy	0.976	0.998	0.9907

VI. DISCUSSION

As we see from the boxplot of average degree, average cluster coefficient, and transitivity in fig. 5-7, respectively, there are distinct region of separation on these three signals which make it easy to utilize as feature vector for classification. SVM was employed as a classifier. The classification result is on table 1. All statistical measurements are in an outstanding figure as expected.

V. CONCLUSION

Vertical visibility algorithm with statistical mechanics can be used as a tool to extract feature of the signal. This is a linear graph method employed to work with the deterministic signal. So we can say that this is a new tool in analyzing the signal in linear environment. The data size in used is smaller than the convention method a lot, this make it more applicable in practical. But some critical caution must be considered on the frequency spectrum since the sampling rate is still governed by Nyquist.

REFERENCES

[1] X. Hu, Z. Wang and X. Ren, " Classification of surface EMG signal using relative wavelet packet energy." Computer Methods and Programs in Biomedicine 17, pp.189-195, 2005.

[2] K. Sabri, "Classification of EMG signals using Neuro-Fuzzy system and diagnosis of neuromuscular diseases." Journal of Medical System, 34, pp.321-329,2010.

[3] S. Sultornsanee, I. Zeid and S. Kamarthi, "Classification of Electromyogram using Recurrence Quantification" Procedia Computer Science 6, pp. 375-378, 2011.

[4] disorders." Computer in Biology and Medicine 43, pp. 576-586, 2013.

[5] Lacasa, L., Luque, B., Ballesteros, F., Luque, J. &Nuno, J., C. "From time series to complex networks: The visibility graph." PNAS, 105(13), pp. 4972-4975, 2007.

[6] Luque, B., Lacasa, L., Ballesteros, F., &Luque, J. "Horisontal visibility graphs: Exact results for random time series." Physical Review E, 80, pp.1-11, 2009.

[7] Campanharo, A., S., L., O., Sirer, M., I., Malmgren, R., D., Ramos, F., M. &Amaral, L., A., N. "Duality between time series and networks". PLoS ONE, 6 (8), pp. 1-12, 2011.

[8] Newman, M., E., J. "Networks an introduction." Oxford University Press,1-771, 2010.

[9] Barabasi, A., L. Network science.E-book version, 2012.