Classification Of Gait Dynamics In Neurodegenerative Disease Patients Using Machine Learning Techniques

S.A.Vajiha Begum, Dr.M.Pushpa Rani

Abstract: The key objective of this paper is to classify the gait dynamics of Neurodegenerative patients using Machine Learning (ML) techniques. The studies state that every individual has a unique walking style and these patterns are termed as gait. Human gait analysis recognizes people from the way they walk, which replicate the individual's unique movement pattern. The severe walking abnormalities in human are caused by progressive brain dysfunction. Accurate diagnosis of the particular brain disorder helps to start early treatment procedures. Lack of personal analysis of the patient to identify the brain disorder needs machine learning based techniques to accurately diagnose and classify the Neurodegenerative Diseases (NDD) such as Amyotrophic Lateral Sclerosis, Parkinson's, Huntington's Diseases and the healthy control subjects. The Recurrence Quantification Analysis (RQA) and Fast Walsh-Hadamard Transform is implemented to quantify the input gait signals to extract the gait parameters. The extracted parameters are further used for classification of NDD diseases with the multi SVM and Random forest algorithm gives 89% and 91% accuracy. Hence, it is possible to classify the normal persons and persons with brain disorders by the Machine Learning techniques with the gait dynamics.

Index Terms: Gait Dynamics, Neurodegenerative Diseases (NDD), Recurrence Quantification Analysis (RQA), Fast Walsh-Hadamard Transform, Multi SVM, Random Forest, Machine Learning.

◆

1. INTRODUCTION

Millions of People are affected by Neurodegenerative diseases worldwide. The most common Neurodegenerative diseases such as Amyotrophic Lateral Sclerosis (ALS), Huntington's and Parkinson disease will reflect in abnormal and slow walking movement in humans. The Neurodegenerative disease is caused due to the loss of neurons and axons in the central nervous system which leads to serious brain disorder. The persons with brain disorders will show symptoms like tremor, slowness of walking or loss of walking, shaking in arms, hand, legs, face and jaw and muscle inflexibility [1]. The study shows that each and every individual has a different walking patterns and these patterns are termed as gait [2]. There is difference in the gait dynamics of healthy individual and the persons with neurodegenerative diseases. Mostly aged peoples are unaware and affected with Neurodegenerative diseases. The exact brain disorder is unable to diagnosis by the physicians and hence require accurate diagnosis method to start the treatment process earlier.The automatic diagnosis of particular neurodegenerative disorder with gait pattern using the machine learning technique is proposed to give efficient and accurate classification of Amyotrophic Lateral Sclerosis (ALS), Huntington's, Parkinson disease and healthy controls without human intervention [10]. The input gait signals used in this work is taken from the physionet public database. The gait signals are non-linear, non-stationary and recurrence in nature. To deal with the non-linear data, Recurrence Quantification Analysis (RQA) and Walsh-Hadamard Transform is applied to measure the input gait signals for the extraction of gait parameters. The RQA quantifies the recurrence of gait time series. The recurrence nature of gait is constructed using the Recurrence plot and features such as Recurrence Rate (RR), Determinism, Entropy and Average Diagonal Length are extracted for further classification. The Walsh-Hadamard Transform is also applied with the gait signals to gather the statistical parameters. These extracted parameters are classified with Multi SVM and Random forest classifiers.

- *S.A.Vajiha Begum, Research Scholar, Department of Computer Science, Mother Teresa Women's University, Kodaikanal. TamilNadu, India, E-mail: vajihabegum7391@gmail.com*
- *Dr M.Pushpa Rani, Professor, Department of Computer Science, Mother Teresa Women's University, Kodaikanal. TamilNadu, India, E-mail: drpushpa.mtwu@gmail.com*

2. LITERATURE REVIEW

Kartikay Gupta et al., [4] presented an efficient classification method and generated a new set of features with the autocorrelation and cross correlation between the gait time series. Mutual Information (MI) analysis is applied and generated 500 features for classification. A rule-based classifier technique using single decision tree classifier was implemented for classification of Huntington's Disease (HD), Parkinson's Disease (PD), Amyotrophic Lateral Sclerosis (ALS), and Neurodegenerative Diseases (NDD) with 500 features achieved classification accuracy of 88.5%, 92.3%, 96.2%, and 87.5%. The validation for HD vs control, PD vs Control, ALS vs control and NDD vs. control got accuracy of 80%, 80%, 90% and 73.33%.
P.Prabhu et al., [6] proposed a efficient method to quantify the non linear gait data using Recurrence Quantification Analysis (RQA). The parameters gained from RQA is used to determine the periodicity, complex behaviour and deterministic nature to describe the individual gait and improved the binary classification accuracy using SVM and PNN gives 96% and 100 % result. Pushpa Rani et al., [9] presented a survey result on gait pattern classification and recognition by comparing Wavelet Descriptor with ICA and Hough transform with PCA. For the classification purpose SVM and nearest neighbour classifiers are used. The gait recognition is done by cumulative match scores and results were finally shown. Pushpa Rani[5] proposed a modified version of Extreme Learning Machine called Hybrid Extreme learning Machine (HELM). The Analytical Network Process

(ANP) technique is implemented for choosing the input weights and hidden biases. The Experimental results on HELM shows classification accuracy of 99.2% and classification time of 0.38 seconds with T-Test. When comparing SVM and ELM techniques the results shows that the Hybrid Extreme learning Machine (HELM) technique for gait classification results in better accuracy compared to the existing techniques. Julius Hannink et al., [7] presents a novel approach to stride length estimation with Deep Convolutional Neural Network (DCNN) to map stride specific inertial sensor data to the resulting stride length. The data collected with shimmer 2R sensors platform which contains inertial sensors. In the stride segmentation, Mid Stance (MS) and heel-strike events are detected individually by Multidimensional Subsequence Dynamic Time Wrapping (MSDTW). Evaluation of three different strides is done by 10-fold cross validation scheme. The strides to define from mid-stance to mid-stance with the average accuracy of 0.01±5.37 cm. Wei Zeng et al., [8] implements Radial Basis Function (RBF) neural networks to classify Parkinson's disease patients and healthy patients. In the training phase the healthy controls and Parkinson's disease patients gaits patterns are approximated with Radial Basis Function and the results are stored in constant Radial Basis Function neural network (RBF). Finally the stored results are rooted in the dynamical estimators in the classification phase in the five-fold cross validation method gives 96.39% classification accuracy. Qiang Ye et al.,[12] proposed Adaptive Neuro Fuzzy Inference System (ANFIS) model with the combination of neural network adaptive technique and fuzzy logic. For learning the ANFIS model Particle Swarm Optimization (PSO) algorithm is used. The classification of ND vs. CO, ALS vs. CO, PD vs. CO and HD vs. CO groups evaluated using leave-one-out cross-validation(LOOCV) method shows accuracy of 90.63%, 93.10%, 90.22% and 94.44%.

3. METHODS

3.1. Gait Data Description:

The gait dataset was gathered through the public gait database Physionet [3] which measures the severity of Neuro Degenerative Diseases in the individual category. The Ground Reaction Force (GRF) gait data provided by Physio-Bank public database was used for the analysis and detection of gait disorder. In this database 8 distinct sensors were set in each foot of the subjects and estimated the vertical ground response constrain. The data consist of gait dynamics intervals that are taken in the real time for both male and female Neurodegenerative Diseases (NDD) patients such as 15 Parkinson's disease patients, 19 Huntington's disease patients, 13 Amyotrophic lateral sclerosis patients (ALS) and 16 healthy control subjects. The time interval of gait parameters like stance, swing, double support interval, and stride from both Left and Right foot are presented in this database. For each subject, the clinical information like age, gender, height, weight, walking speed, disease severity is specified in the physionet database. The figure 1 shows the stride intervals derived from the gait signals for both left and right foot of the person with (a) ALS Disease of age 68, (b) Huntington's Disease of age 42, (c) Parkinson's Disease of age 77 and (d) Healthy control subject of age 57. It has been exposed that the gait parameters are expressively different in each group. These variations in both the left and right foot stride intervals were considered for the classification of particular group of disease from healthy control in this study.

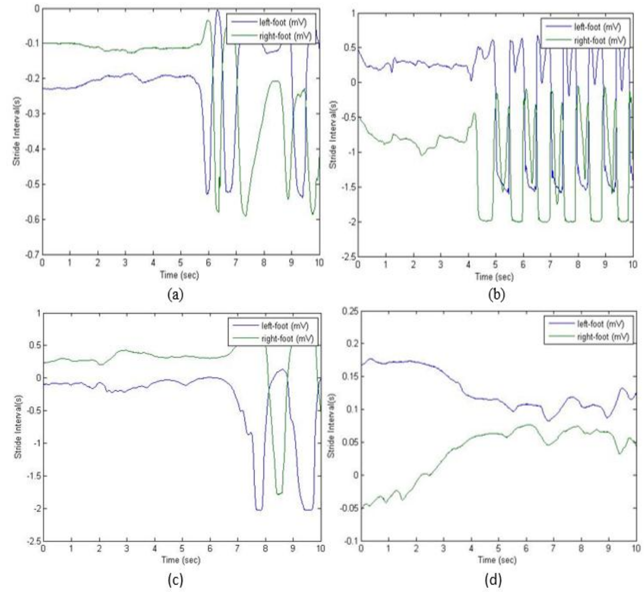

Fig 1. The stride intervals derived from the gait dynamics for both left and right foot of the persons with (a) ALS Disease, (b) Huntington's Disease, (c) Parkinson's Disease and (d)Healthy control subject.

3.2. Clinical Gait Analysis Phases:

Clinical gait analysis techniques are implemented to recognize the brain disorders by gait dynamics of the patients. The figure 2 shows the block diagram of the proposed system.

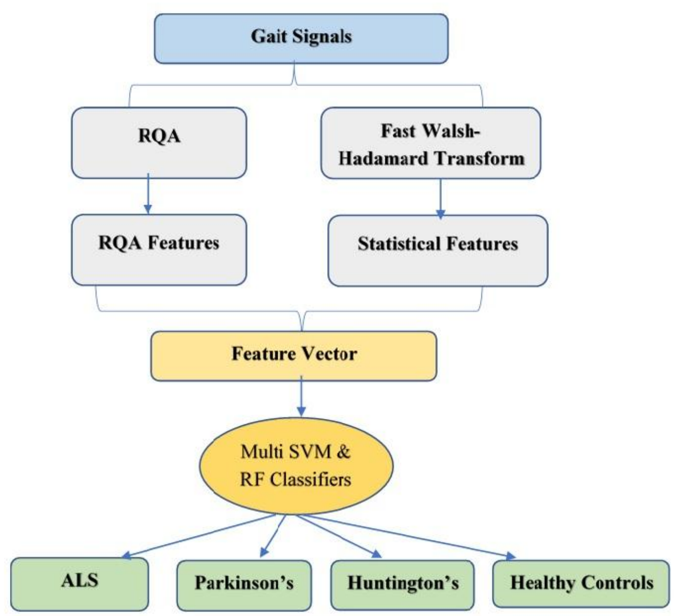

Fig 2. Block diagram of the proposed system

3.3. Feature Extraction:

The gait signals of both left and right foot stride intervals are considered for feature extraction. Here Recurrence Quantification Analysis (RQA) and Walsh-Hadamard Transform is implemented to quantify the input gait signals to extract the gait parameters.

A. Recurrence Quantification Analysis (RQA):

The gait signals are non-linear in nature which is synchronized by the corresponding activities of the brain. Recurrence Quantification Analysis (RQA) is good in handling the non-linear dynamically changing data [6]. The RQA was utilised based on small scale structures of gait dynamics by quantifying the recurrence plots (RP). The Recurrence Plot (RP) pictures spatial and temporal correlations between features associated with both the left foot and right foot, which shown in matrix form,

$$RP_{i,j} = H(\in - |x_i - y_i|) \qquad (1)$$

Where RP is the recurrence plot matrix, which is the difference of left foot stride interval and right foot stride interval, ∈ is the threshold value and H is Heaviside function. The RQA features are extracted from the RP. The features extracted from the RQA were Recurrence Rate (RR), Determinism (DET), Average Diagonal Line (AL) and Entropy (ENT). These features are further used for the classification of Neurodegenerative diseases using Multi SVM and Random Forest classifiers. The RQA features are as follows:

(a) Recurrence Rate (RR):

The recurrence rate measures the density of recurrence points in a recurrence plot (RP). The higher the RR value indicates the gait disorder and lower the RR value indicates the healthy control subjects. The Recurrence Rate is defined as in (2)

$$RR = {}^{1}/_{N^2} \sum_{i,j=1}^{N} RP_{i,j}(\in) \qquad (2)$$

(b) Determinism (DET):

The determinism measures the percentage of RP which forms the diagonal lines of minimum length L_{Min}. The Determinism is measures using (3)

$$DET = \frac{\sum_{L=L_{Min}}^{N} Lp(L)}{\sum_{L=1}^{N} Lp(L)} \qquad (3)$$

Where L is the length of the diagonal lines and p(L) is the histogram of that diagonal lines. The longer diagonal lines will indicate the gait disorder and shorter will indicates the healthy control subjects.

(c) Average Diagonal Line (AL) :

The Average diagonal line indicates the average time in RP which were parallel to each other. The Average Diagonal Line is defined as

$$AL = \frac{\sum_{L=L_{Min}}^{N} Lp(L)}{\sum_{L=Min}^{N} p(L)} \qquad (4)$$

(D) Entropy (ENT):

Entropy is the measure of Recurrence Points (RP) occurrence which form the diagonal line whose length is more than L_{Min}. The entropy measure is small for the healthy control subjects and the entropy measure is large for the gait disorder. The recurrence is more in the gait disorder patients. The Entropy is measured as

$$ENT = -\sum_{L=L_{Min}}^{N} \left(\frac{P(L)}{k}\right) Ln\left(\frac{P(L)}{k}\right) \qquad (5)$$

Where, p(L) is computed as p(L)/k which shows the recurrence of diagonal lines in RP, where k is the summation of length (L) of all diagonal lines.

B. Statistical Features:

In this study the Fast Walsh-Hadamard Transform is implemented to extract the statistical features like Energy, Standard Deviation, Mean, Variance and Co Variance from the time series of left and right strides intervals. The Fast Walsh-Hadamard Transform (FWHT) is efficient to compute Walsh-Hadamard Transform (WHT), equivalent to a multidimensional Discrete Fourier Transform of matrix of order 2. The Hadamard Transform is defined as

$$H_n = {}^{1}/_{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix} \qquad (6)$$

(a) Energy

The Energy (E) of the gait series stride signal x(s) from each subject is computed as

$$E = \int_{-\infty}^{\infty} |x(s)|^2 \, dt \qquad (7)$$

(b) Mean

The Mean is defined here as the average value of the gait signal. The mean is calculated as the sum of signal values X_i divided by the total number of subjects S.

$$\mu = \frac{\sum_{i=1}^{S} X_i}{S} \qquad (8)$$

(c) Standard Deviation

The standard deviation is calculated with the measure of variability and reliability of subjects. The signal values is represented as X_i, μ is the mean, S is the total number of subjects, and the standard deviation (σ) is measured as

$$\sigma = \sqrt{\frac{\sum_{i=1}^{S} (X_i - \mu)^2}{S-1}} \qquad (9)$$

(d) Variance

The Variance is the squared measure of standard deviation. The sum of square of each term X_i minus the mean μ, divided by the total number of subjects S. The variance is calculated as

$$Var = \frac{\sum (X_i - \mu)^2}{S-1} \qquad (10)$$

(e) Co Variance

The Co Variance calculates the data points from the average value in a dataset. The covariance between two random variables X and Y can be calculated with the mean $\bar{x}$ and $\bar{y}$ by the formula shown as

$$\mathrm{Cov}(x, y) = \frac{\sum_{i-1}^{S}(X_i - \bar{x})(Y_i - \bar{y})}{S-1} \qquad (11)$$

Where, $\bar{x}$ and $\bar{y}$ are the means of X and Y respectively.

The Recurrence Quantification Analysis (RQA) features like Recurrence Rate (RR), Determinism (DET), Average Diagonal Line (AL) and Entropy (ENT), the Statistical features extracted with Walsh-Hadamard Transform like Energy, Standard Deviation, Mean, Variance and Co Variance were considered for the classification purpose. These nine features are termed as Feature vectors, which were further classified using the Machine Learning classifiers such as Multi SVM and Random Forest classifiers.

3.4. Classification:

The RQA features are extracted from the stride intervals difference of both left and right foot are gathered. The statistical features are extracted with the FWHT from the individual left and right stride intervals. Finally, the extracted features are undergone for classification process with efficient Machine Learning algorithms. The Multi SVM and Random Forest (RF) Classifiers were implemented with these extracted features to classify gait disorders such as Amyotrophic Lateral Sclerosis, Parkinson's, Huntington's Diseases and the healthy control subjects. SVM is a binary classification model and Multi SVM is multiclass classification model. Here Multiclass SVM Classification method is implemented with the Radial basis function (rbf) kernel, to identify and classify each particular NDD disease and control subjects with the trained feature vectors. The Random Forest Classifier is also implemented here to classify the feature vectors. The RF classifier creates a set of decision tree by randomly selecting the features from trained feature vectors. The method of finding the root node and splitting the feature nodes will run randomly. After the Forest of decision tree is created, then calculates the votes for each predicted target. The Leave-one-out-cross-validation (LOOCV) is implemented here to measure the performance of the proposed system. LOOCV randomly select single subject for validation and take the rest data for training purpose. This training process is repeated multiple times by selecting different data for validation and finally performance is measured. The performance measures like accuracy, sensitivity, specificity, Precision, Recall and Fscore are measured for the Multi SVM and Random Forest Classifiers for the classification of NDD disorders and healthy controls.

4. RESULT AND DISCUSSION

In the proposed study the gait dynamics are collected from the physionet database [3] for NDDs such as ALS, Huntington's, Parkinson's diseases and healthy control subjects. In this work the gait stride interval difference for both left and right foot in all subjects is selected. The figure 3 shows the Left stride intervals and Right stride intervals of all subjects. The figure 4 shows the Left stride intervals intermediate difference and Right stride intervals intermediate difference of all subjects. The variations in the stride intervals must be considered for the classification of particular group of disease and healthy control in this study. Here Recurrence Quantification Analysis (RQA) and Fast Walsh-Hadamard Transform is implemented to extract the gait features for the classification. The Multi SVM and Random Forest (RF) Classifiers were implemented with these extracted features to classify Neurodegenerative disorders like Amyotrophic Lateral Sclerosis, Parkinson's, Huntington's Diseases and the healthy control subjects.

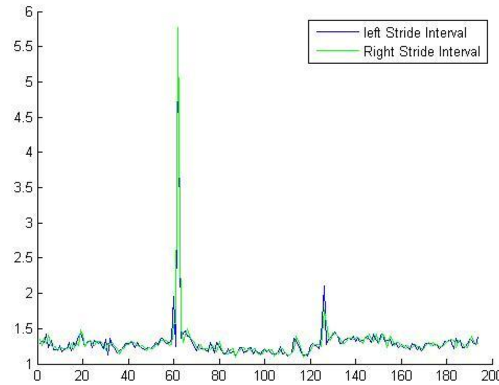

***Fig 3**. The Left and Right Stride intervals*

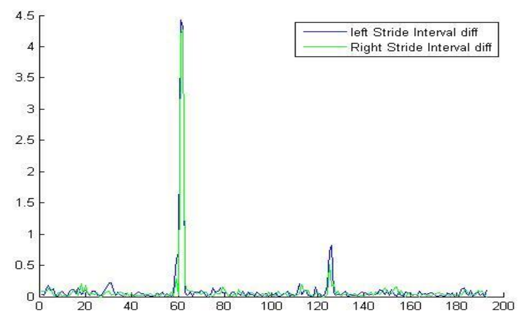

***Fig 4.** The Left and Right Stride interval differences*

The performances of the Multi SVM and Random Forest classifiers are evaluated in terms of sensitivity, specificity, accuracy, precision, Recall and Fscore for the classification of each NDDs and healthy controls. The Recall and sensitivity measures are similar. These measures are calculated by comparing the original test output and the predicted output by LOOCV method. The confusion matrix gives the number of true positives (TP), false positives (FP), false negatives (FN) and true negatives (TN) for both classifiers. These Performance measurements are calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Fscore = 2 * \frac{Precision * Recall}{Precision + Recall}$$

The classification of neurodegenerative diseases and healthy controls with the Multi SVM and Random Forest algorithms gives the accuracy of 89% and 91.4% result. The performance analysis of these classifiers are presented in the table 1. The Proposed system shows Random forest classifier with the RQA and Walsh-Hadamard transform features gives better result compared to Multi SVM technique.

Table 1
Performance Measures of Classifiers in Classification

Performance Metrics	Random Forest	Multi SVM
Accuracy	91.4%	89%
Sensitivity	82.5%	78.8%
Specificity	94.3%	92.6%
Precision	81.9%	78.9%
Fscore	81.9%	78.2%

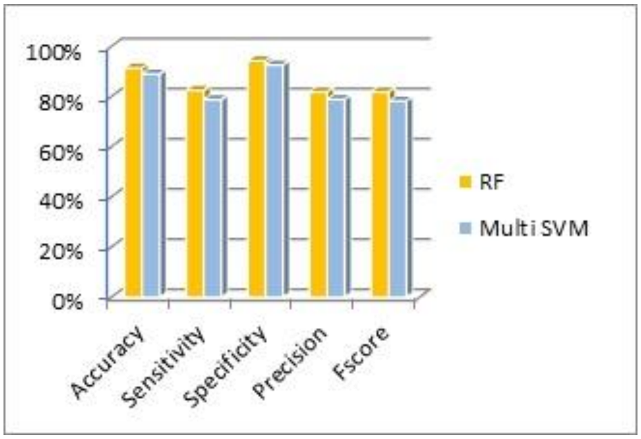

Fig 5. *Performance Analysis for Classification with RF and Multi SVM*

5. CONCLUSION

The present work extracted the differentiation of left and right foot stride intervals from the gait dynamics with the Recurrence Quantification Analysis (RQA) and statistical features from left and right stride intervals with Fast Walsh-Hadamard Transform which were effective in the classification of each NDD diseases and healthy controls. The efficient classifiers like Multi SVM and Random Forest classifiers were involved in the classification task and achieved overall accuracy of 91.4% for Random forest and 89% for Multi SVM. Hence, it is possible to classify the healthy persons and persons with brain disorders by the Machine Learning techniques with the gait dynamics. In future all other gait parameter will be taken in feature extraction state in this present study.

REFERENCES

[1] Sandra Amor, Fabiola Puentes, David Baker, Paul van der Valk,"Inflammation in neurodegenerative diseases",Immunology. 2010 Feb; 129(2): 154–169.
[2] T. Carletti, D. Fanelli, A. Guarino, A new route to noninvasive diagnosis in neurodegenerative diseases?, Neuroscience Letters 394 (2006) 252–255
[3] www.physionet.org/physiobank/database/gaitpdb/
[4] Kartikay Gupta,et al. “Rule based classification of neurodegenerative diseases using data driven gait features” Springer, 2018.
[5] M.Pushpa Rani, "Abnormal GAIT classification using hybrid ELM," 2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE), Toronto, ON, 2014, pp. 1-8.
[6] P.Prabhu et al., “Classification of Gait Signals into Different Neurodegenerative Diseases using Statistical Analysis and Recurrence Quantification Analysis”, Pattern Recognition Letters, Elsevier, 2018
[7] J.Hannink et al., Mobile Stride Length Estimation with Deep Convolutional Neural Networks, IEEE Journal Of Biomedical And Health Informatics, 2017.
[8] Wei Zeng et al., “Parkinson’s disease classification using gait dynamics via deterministic learning”, Elsevier,2016.
[9] M.Pushparani & D.Sasikala, 2012. “A Survey of Gait Recognition Approaches Using PCA & ICA”, Mother Teresa Women’s University, Kodaikanal, India, Global Journal of Computer Science and Technology Volume 12 Issue 10 Version 1.0.
[10] S.A.Vajiha Begum & Dr.M.Pushpa Rani, “Classification of Brain Disorder using Gait Patterns”, International Journal of Scientific Research in Computer Science Applications and Management Studies, Volume 8, Issue 1, January 2019.
[11] Hausdorff JM, Lertratanakul A, Cudkowicz ME, et al. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. J Appl Physiol. 2000;88:2045–53.
[12] Qiang Ye et al., “Classification of Gait Patterns in Patients with Neurodegenerative Disease using Adaptive Neuro-Fuzzy Inference System”, Hindawi, Computational and Mathematical Methods in Medicine, Volume 2018.