Pattern Recognition Letters 139 (2020) 10–16

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Classification of gait signals into different neurodegenerative diseases using statistical analysis and recurrence quantification analysis

Pooja Prabhu[a], A.K. Karunakar[a,*], H. Anitha[b], N. Pradhan[c]

[a] Department of Computer Applications, Manipal Institute of Technology, Manipal University, Manipal 576104, India
[b] Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal University, Manipal 576104, India
[c] Department of Psychopharmacology, National Institute of Mental Health and Neurosciences, Bengaluru 560029, India

ARTICLE INFO

Article history:
Available online 5 May 2018

Keywords:
Gait
Neurological disorder
Probabilistic neural networks
Recurrence quantification analysis
Sports medicine
Support vector machine

ABSTRACT

Among all the biological signals, gait signal is one of the better features to detect movement disorders caused by a malfunction in parts of the brain and nervous system. Usually, identifying and evaluating movement disorders caused due to neurodegenerative diseases solely depends on a physicians experience. Different diseases having gait abnormalities generate a unique gait characteristic. Traditionally, Fourier analysis is used to understand the gait characteristic, thereby predicting potential diseases. Fourier analysis assumes the gait signal to be stationary, linear and noiseless which is not a reality. To overcome this, Recurrence Quantification Analysis (RQA) is used in this study to quantify gait parameters. RQA has proved to be one of the best tools for non-linear, non-stationary and short length data. It is used to quantify heart rate variability, ventricular fibrillation, wrist pulse and growth of bladder. This paper uses RQA in understanding the dynamics of human gait and the parameters obtained are used as a feature for classification using Support Vector Machine (SVM) and Probabilistic Neural Network (PNN). This study considered thirteen subjects for the classification of gait signals of patients with Neurodegenerative diseases (Amyotrophic Lateral Sclerosis, Huntington and Parkinson) and thirteen healthy control subjects using two different classification models like Support Vector Machine (SVM) and Probabilistic Neural Network (PNN). Features were extracted after statistical analysis and RQA, and Hill-climbing feature selection method was used to optimize the feature set. The accuracy deduced after binary classification using SVM and PNN ranged from 96% to 100%.

1. Introduction

A neurodegenerative disease alters the structural, biochemical or electrical activities in the nervous system. Symptoms of neurodegenerative disease include weakness in muscles, poor coordination between nervous system and joints, pain and altered level of consciousness [21]; collective effect will be altered gait pattern. Gait is the total output of primary motor cortex, basal ganglia, and the cerebellum. Neurodegenerative disease includes Alzheimers, Amyotrophic Lateral Sclerosis (ALS), Parkinson's and Huntington's disease [3]. In this study, ALS, Parkinson, and Huntington diseases are considered because all three diseases show similar symptoms like muscle weakness, slow gait and stooped posture [5]. It is difficult for a physician to distinguish the subject, just by considering the gait parameters which results in the introduction of subjective error. Thus, there is a demand for a research to find the hidden relationship between the gait signals corresponding to these three neurodegenerative diseases, which in turn helps in automating the classification of patients with neurodegenerative diseases based on the gait signals. Automatic classification eliminates the subjectivity introduced by a physician, improves efficiency and accuracy, and speeds up the process of diagnosis.

Gait involves the sequence of rhythmical and periodic pattern of left and right foot movements coming in contact with the ground [8]. Gait cycle suffers step to step fluctuations. The cause of these fluctuations is due to the presence of some underlying complex temporal structures, which are non-linear by nature. Duration covered to complete one gait cycle (stride phase) varies from individual to individual and depends on the physiological and pathological conditions of an individual. In general, gait signals exhibit properties like non-linearity, noiselessness, and non-stationarity. Gait cycle is measured by temporal and spatial parameters [8]. Temporal parameters associated with one gait cycle are recurrence in nature due to which clinicians managed to analyze the occurrence of fundamental events by recording the one cycle of gait with the

* Corresponding author:
E-mail address: karunakar.ak@manipal.edu (A.K. Karunakar).

https://doi.org/10.1016/j.patrec.2018.05.006
0167-8655/

aid of valuable tools. A biological process like human gait, which is non-linear by nature, possesses distinct properties like periodicity, deterministic behavior, and irregular cycle. The basic property of a underlying nonlinear system (can be chaotic) is that the every state of the system traverse back to same state or may traverse close to the previous state during a prescribed time, which makes gait cycle recurring in nature [7].

The temporal gait parameters recorded are one-dimensional. To understand the dynamics of gait, the entire topology of gait must be understood. This is facilitated by acquiring the gait parameters in all possible dimensions [7]. To fulfill this requirement, sophisticated gait recording machines must be developed, which record gait parameters in all possible dimensions. The development of gait machine with the ability to record gait parameters in all dimensions would assist the clinicians to know the extent of strain on a calf muscle in the field of sports medicine, in spatial tracking of the human body to illustrate the control process behind maintaining the equilibrium and postural imbalance, and also helps in understanding the dynamics of gait observed in subjects with myopathy and neurological disorders.

Fourier analysis on the physiological signals like the gait signals cannot be performed because of their properties like nonlinearity, non-stationarity and recurrence nature. The disadvantage of using Fourier analysis on physiological signals is that it assumes the signal to be linear and stationary, but it is adequately for long length datasets [7,19]. Hence, it is preferable to use Fourier analysis on other biological signals like electroencephalography (EEG), electromyography (EMG), and electrocardiography (ECG). In this paper, we present a technique which relaxes the concept of linearity, stationarity, and size of the gait time series, namely Recurrence Quantification Analysis (RQA). The basic concept behind RQA is to explore the recurrence nature of gait time series in the reconstructed phase space. Phase space of gait time series is reconstructed based upon embedding theorem [11]. Embedding theorem facilitates in understanding the dynamics of gait by more number of observables, which was not possible to measure earlier. Phase space reconstruction further proceeds in representing the recurrence nature of gait using Recurrence Plots (RP). RP exhibits the structures like single dots, horizontal lines, diagonal lines and vertical lines. Quantifying these structures called as RQA, yields thirteen parameters. Thirteen parameters are unique since they were deduced based on different aspects of quantification.

In this study, gait analysis is carried out on pathological and normal subjects. Pathology includes neurodegenerative diseases like Parkinsons disease, Huntingtons disease, and ALS, whose gait dynamics is studied and compared. The parameters obtained by performing cross RQA were considered as features in binary classification of pathological and normal subjects using Support Vector Machine (SVM) [20] and Probabilistic Neural Network (PNN) [18].

This paper has been organized as follows: Section 2 briefs about the previous work in gait analysis and suggests the importance of RQA in gait analysis. Section 3 elaborates methodology used in feature extraction, feature selection and classification models involved in the study. Section 4 briefs on datasets considered in the study and shows the result obtained after the classification of gait signal. Section 5 concludes the work.

2. Related works

Gait features extracted from gait signals that have been classified using several different approaches are highlighted as follows.

Yang et al. [23] used ten gait parameters to extract four features, namely, maximum signal-to-noise ratio based feature selection method, maximum signal-to noise ratio combined with minimum correlation based feature selection method, and maximum prediction power combined with minimum correlation based feature selection method and principal component analysis. Support Vector Machine (SVM) was used as a classifier. The classification accuracy ranges from 79.04% to 93.96%. The feature extraction techniques used in this work fails to extract distinct features in short length data and assumes the input signal to be noiseless. The results of binary classification show that there is a scope for improvement in the classification accuracy of particular classification problems like Parkinson vs. Huntington (79.04%) and Huntington vs. Control (84.17%).

Dutta et al. [4] extracted time and frequency domain features of the correlograms obtained by cross-correlating the gait signals with a reference. Subsequently, pre-trained Elmans Recurrent Neural Network (ERNN) was employed for automatic identification of healthy subjects and those with neurological disorders. The accuracy obtained for frequency domain features is 81.6% and for time domain feature is 87.1%. The Neural Network used in this work is processed with lot of iterations in turn were found to be time consuming. The frequency domain features are obtained from Fourier Transform which assume the gait signal to be stationary, noiseless and linear contradicting the real time nature of the signal.

Li et al. [10] conducted research using 15 individuals having Parkinsons disease, 13 individuals having ALS disease, 20 individuals having Huntingtons disease and 16 control subjects. Nine time domain features and two frequency domain features were extracted. Two classifiers were considered, namely quadratic Bayes normal classifier and SVM. This work improves the classification accuracy of Yang et al. [23]. The results of binary classification show that there is a scope for improvement in the classification accuracy of particular classification problems like Parkinson vs. Huntington (79.70%), ALS vs. Huntington (80.51%) and Huntington vs. Control (83.3%).

Xia et al. [22] used the gait data set of 16 healthy, 13 ALS, 15 Parkinson and 20 Huntingtons subjects. Nine statistical measures were extracted from gait signals and hill-climbing feature selection method was used. Leave-one-out cross-validation method was used for evaluation. The highest accuracy rate discriminating neurodegenerative patients and healthy control subjects was 96.83%.

RQA is used by many researchers to quantify heart rate variability, ventricular fibrillation, wrist pulse and growth of bladder [11]. In recent advances, RQA is extended in gait analysis. The work by Negahban et al. [13] on the unilateral anterior cruciate ligament injury subjects explored by using RQA showed that double postural task show more regularity in comparison to single postural task. Labini et al. [9] used RQA in understanding the complexity of the head, trunk, and pelvis during locomotion in normal and subjects with unilateral vestibular hypofunction. The work by Riva et al. [17] explored that RQA can be used in testing the stability of an individual by recording the angular velocities and acceleration during the gait cycle using tri-axial sensors placed at the trunk and leg.

Prabhu and Pradhan [15] used gait data of 13 subjects each from Control, ALS, Parkinson and Huntington. Eight parameters of RQA were extracted from gait data associated with stride, stance and swing phase. These parameters were considered as features for multiclass classification using PNN. The accuracy corresponding to stance, swing and stride phase stride, swing and stance phase is 54.16%, 37.5% and 50% respectively. This suggested that there is a scope to reduce the computational complexity by limiting the parameters of RQA in the stride phase alone, because stride phase is a combination of 60% stance phase and 40% swing phase. The effort required acquiring the gait data during stance and swing phase is immense, due to fact that these two phases lapse for short duration.

The motivation for this work was that the classification of features extracted from gait signal using ERNN and cross-validation method failed to yield a good accuracy rate when classifying normal and pathological gait. The feature extraction method used in previous works relaxes the real-time constraints like the linearity of gait signal, non-stationary and noiselessness nature. RQA used in understanding the nature of gait signals proved that it can be used to quantify gait with various disorders without relaxing the real-time constraints. From [15] it is understood that RQA is one of better techniques to extract a feature from gait signal. This work is an exhaustive extension of [15] and considers the statistically significant parameters of RQA as features in improving the accuracy of binary classification problems as listed in [10,22,23].

3. Optimization of gait feature set for classification

Gait parameters can be measured in terms of temporal and spatial aspects. Temporal gait parameters are non-linear, nonstationary and noisy by nature. The non-linear nature is due to the fact that gait is coordinated by the synchronous activities of the brain [8]. Gait signals exhibit chaotic behavior owing to the fact that these signals are non-linear. Chaotic systems are deterministic [12]. According to Henry Poincar e, a pioneer who introduced the concept of recurrences [14], stated that In this case, neglecting some exceptional trajectories, the occurrence of which is infinitely improbable, it can be shown, that the system recurs infinitely many times as close as one wishes to its initial state. This can be interpreted as, any chaotic system exhibiting indefinite motion of trajectories, can recur. Based on this fact, one can infer that the chaotic nature of gait is recurrent in nature. The recurrent nature of gait can be explored by Recurrence Quantification Analysis.

The entire topology of gait can be explored using Taken's time delay embedding theorem [19]. The time series or gait signal now changes from one dimension to higher embedded dimension. The analysis which incorporates this criterion is Recurrence Quantification Analysis (RQA). RQA mainly concentrates on quantifying the geometrical patterns [16] that are observed in the form of small-scale textures and are graphed as Recurrence Plots (RP) [7]. By quantifying the small-scale structures (single dots, vertical lines and horizontal lines, and diagonal lines), eight parameters of RQA deduced were recurrence rate, determinism, average diagonal line, maximal diagonal line, entropy, trapping time, laminarity and maximal vertical line [11,15]. As RQA relies on recurrence nature of gait signal, it does not demand any assumptions as that of Fourier analysis.

The concept of using RQA in extracting features of gait signal has already been proposed by Prabhu and Pradhan [15]. The same concept has been applied in this work by reducing the number of parameters of RQA. These parameters are considered as features for the classification models like SVM and PNN. This work considers the statistical measures derived from gait parameters (left and right leg stride interval) along with the parameters derived from RQA. The contribution of the present work is to improve the performance of classification problems listed in [22] by using parameters of RQA and statistical measures.

Fig. 1 shows a block diagram of the proposed system for the binary classification problems as identified by Li et al. [10], Xia et al. [22], Yang et al. [23]. The proposed system depicts that the gait signal is classified under Class 1 or Class 2. Features from gait signals are extracted in two stages. Firstly, gait signals (left leg and right leg stride interval) undergo RQA in extracting eight parameters of RQA, now considered as features. Subsequently, six statistical measures of gait signals are considered as features. The optimal feature set is generated in feature selection. The resulting feature set is used in classifying gait signals using classification models like Support Vector Machine (SVM) and Probabilistic Neural Networks (PNN).

Recurrence Plot (RP) visualizes spatial and temporal correlations [1] between gait parameter associated with the right leg and left

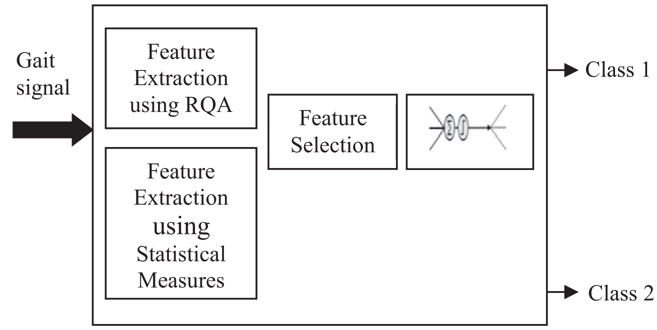

Fig. 1. Block diagram of proposed system.

leg, defined by the matrix,

$$R_{i,j}(\epsilon) = H(\epsilon - \|x_i - y_j\|), \tag{1}$$

where ϵ is the threshold which predefined by the user (usually 1), x_i and y_j are the gait parameters (stride interval) associated with the right and left leg respectively and **H** is the Heaviside function. Line of identity is a diagonal line that appears in a plot when i=j.

Recurrence plot generates the small-scale patterns also called as textures, which can be observed on closer inspection. Textures generated by RP include single isolated dots, and vertical and horizontal lines, and diagonal lines. In practical applications, Qualitative analysis (or Recurrence Plot) alone may not be good choice because of difficulty in witnessing the small-scale patterns by visual inspection. Inorder to overcome this difficulty, a new analysis called as Recurrence Quantification Analysis (RQA) [2] was introduced by Zbilut and Webber. Unlike Recurrence plot, which was qualitative analysis, RQA is a quantitative analysis, as the name suggests it quantifies the small-scale structures observed in RP [15], as mentioned earlier as well.

The parameters produced from RQA are recurrence rate, determinism, average diagonal line, maximal diagonal line, the entropy of diagonal line, laminarity, trapping time, maximal vertical line, recurrence time of first type, recurrence time of second type, recurrence time entropy, clustering coefficient and transitivity. The choice of features solely depends on intent of the work and this work intends to explore the dynamic nature of gait signals. From [15] it can be observed that the eight parameters of RQA that contribute in understanding the dynamic characteristics of gait signal like periodicity, deterministic nature and complex behavior describe the individuals gait distinctly. The eight parameters are listed as follows.

Recurrence Rate (*RR*)

RR quantifies the percentage of recurrence points or quantifies the number of 1's in the RP except those included in line of identity. This implies that periodic signals show higher RR. Periodicity is more in gait disordered subjects than in normal subjects. The value of RR will be comparatively less in normal subjects. Recurrence rate is calculated as in (2).

$$RR(\epsilon) = \frac{1}{N^2}\sum_{i,j=1}^{N} R_{i,j}(\epsilon) \tag{2}$$

Determinism (*DET*)

DET measures the ratio of recurrence points forming the diagonal lines of minimum of length l_{min}. Longer diagonal lines are witnessed in periodic signals whereas chaotic signals produce shorter diagonal lines. Thus, long diagonal lines can be visualized in gait impaired subjects than in normal subjects. Determinism is calculated using (3).

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{N^2 * RR} \tag{3}$$

Where, l is the length of a diagonal lines which appear parallel to the line of identity and $P(l)$ is the histogram of such diagonal lines.

Average diagonal line (*L*)

Average diagonal line is obtained by averaging all consecutive 1's in RP which are parallel to line of identity. This parameter measures the average time while the trajectories associated with two series were close to each other. The average diagonal line is estimated using (4).

$$L = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=l_{min}}^{N} P(l)} \tag{4}$$

Maximal diagonal line (*L_{max}*)

L_{max} determines the length of diagonal line which is longest when compared to other diagonal lines which are parallel to line of identity. Less stable signals or chaotic signals show shorter L_{max}. Thus, the gait pattern witnessed in normal subject which is chaotic in nature show smaller L_{max} value whereas periodic signals produce long diagonal lines [25], so for ALS or gait impaired subjects L_{max} must be large. The maximal diagonal line is estimated using (5).

$$L_{max} = max(\{l_i; i = 1, 2, 3 \ldots N_l\}) \tag{5}$$

where, N_l is the number of diagonal lines which appear parallel to line of identity.

Entropy (*ENT*)

ENT measures the complexity of the signal by determining the probability of occurrence of the all diagonal lines whose length is above l_{min}. Since, periodic signals produce diagonal lines of equal length; the entropy measured would be small, thus less complex. Gait patterns of normal subjects show diagonals of variable length, thus they are complex, so *ENT* must be small. The reason for complexity was that a normal subjects show various degrees of freedom. In case of ALS, *ENT* will be large, since recurrence is more, complexity is less. Shannon entropy for distribution of diagonal line is calculated as in (6).

$$ENT = -\sum_{l=l_{min}}^{N} p(l) ln(p(l)) \tag{6}$$

Where, *p(l)* is the probability of occurrence of a diagonal line of length l in the RP, and can be computed as $p(l) = P(l)/k$, where k is the sum of all the diagonal lines of length l.

Laminarity (*LAM*)

LAM measures the ratio of recurrence points forming the vertical lines of minimum length v_{min}. Laminarity is calculated using (7).

$$LAM = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)} \tag{7}$$

Where, $P(v)$ is the histogram of vertical lines v.

Trapping Time (*TT*)

TT is obtained by averaging all vertical line lengths above the threshold. In case of Huntington, *TT* value is high when compared to other class, since they show sudden movement (jerky) followed resting tremors. Trapping time is calculated using (8).

$$TT = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=1}^{N} P(v)} \tag{8}$$

Maximal vertical line (*V_{max}*)

Maximal vertical line is given by,

$$V_{max} = max(\{v_i; i = 1, 2, \ldots, N_v\}) \tag{9}$$

Where, N_v is the number of vertical lines.

The concrete idea behind the choice of parameters is that these parameters should help in distinguishing the dynamics of gait among control, ALS, Parkinson and Huntington subjects. Parameters like laminarity and the average diagonal line have been eliminated in this work. Laminar state is observed only in Huntington gait but this work aims in using the parameters that show significance in control, ALS, Parkinson and Huntington subjects. Processes with chaotic behavior lead to very short diagonals, whereas periodic processes lead to longer diagonal lines [12]. Gait signal of normal or neurodegenerative disease subjects show periodic and chaotic nature at different intervals. Average diagonal line adds all these diagonal lines in the recurrence plot and gives an overall response; it fails to give in-depth information of gait cycle that would help to distinguish between a normal and neurodegenerative disease subjects.

Experimentally the choice of six parameters is done by performing Mann-Whitney test. Mann-Whitney test is performed to select the statistically significant feature of control and Neurodegenerative group [22]. The hypothesis is that the feature whose p-value is less than 0.05 is considered to be statistically significant. *RR, DET*, L_{max}, *ENT, TT* and V_{max} have p-value less than 0.05 and therefore, they are statistically significant. These statistically significant features are only considered for classification.

3.1. Statistical features of the gait

The embedded features within the raw gait signal are not considered by Prabhu and Pradhan [15]. This work considers parameters of RQA along with statistical features (*Minimum(Min), Maximum(Max), Mean, Standard Deviation(Std), Skewness*, and *Kurtosis*) of raw gait signal for improving the classification accuracy. Statistical measures are calculated from the two gait parameters (left leg stride interval and right leg stride interval) as shown in Table 1.

Further, Mann-Whitney test is performed to retain the statistically significant features of control and Neurodegenerative group as used by Xia et al. [22]. The hypothesis is that the feature whose p-value is less than 0.05 is considered to be statistically significant. It is found that the *Max, Mean* and *Std* has p-value less than 0.05 and hence are statistically significant. The statistically significant features are considered for classification.

3.2. Feature selection

Nine statistically significant features (six parameters of RQA and three statistical measures) are further condensed to optimal feature set using feature selection. Hill climbing feature selection method [24] is employed in this work. Hill climbing feature selection method was chosen to avoid the computational burden to find the optimal feature set by performing the complete search for different feature combinations. Feature selection method performs

Table 1
Six statistical measures obtained from left and right leg stride interval.

Class	Statistical measures					
	Min	Max	Mean	Std	Skew	Kurtosis
ALS	1.03	55.15	1.4034	1.5206	24.9921	722.2584
Control	0.11	2.66	1.1018	0.994	2.621	20.612
Huntington	0.6367	9.9567	1.1482	0.2876	10.2854	273.9508
Parkinson	0.5067	18.5133	1.1366	0.4633	28.8645	973.8158

Table 2
Data points corresponding to thirteen subjects of ALS, Control, Huntington and Parkinson.

Number of Subjects	Data Points			
	ALS	Control	Huntington	Parkinson
1	194	259	310	245
2	242	241	225	277
3	215	255	232	230
4	135	267	268	222
5	205	250	263	263
6	176	249	263	269
7	159	260	190	226
8	232	261	256	203
9	212	275	255	222
10	246	198	220	288
11	229	269	217	230
12	122	244	258	247
13	183	251	167	251

step-by-step search by considering one feature after the other. The set of features that gives better accuracy is considered to be the optimal feature set.

3.3. Classification models

Classification of normal and neurodegenerative gait is performed by SVM [20] and PNN [18]. Nine statistically significant features were used in all the six classification tasks. Leave-one-out-cross-validation (LOOCV) is used to calculate the performance outcome like accuracy, sensitivity and specificity. Grid search using different value pairs of sigma $\in$ {0.01, 0.1, 0.5, 1, 1.5} was calculated to find the optimal value of sigma for SVM using radial basis function(rbf) kernel, whereas for PNN the sigma value found in the range 0.2-2 is most suitable.

4. Results and discussion

4.1. Datasets under study

In this study, Neuro-Degenerative Disease Gait Dynamic (NDDGD) database is used, which was published by Physionet [6]. For this study, temporal gait parameters of thirteen subjects each for three neurodegenerative disease (ALS, Huntingtons and Parkinsons disease) and thirteen healthy control subjects were obtained from the database. Data points of each subject pertaining to ALS, Control, Huntington and Parkinson disease is tabulated in Table 2.

4.2. Pre-processing of gait signal

Raw gait signal is composed of spurious peaks and noise outliers that cannot be directly used in RQA. Spurious peaks occur when subject lifts the leg for a long period of time while recording gait parameters. The occurrence of spurious peak was observed to be random, because of which there was no scope to use band-pass filter. Thus, Prabhu and Pradhan [15] designed a new filtering technique, which filters out the spurious peak. The new filtering technique checks the segment of gait signal that has no spurious peak, and the average of the segment is calculated, which is further considered as a threshold for remaining segment of gait signal [15]. This work adopts the same filtering technique used by Prabhu and Pradhan [15].

4.3. Analysis and quantification of gait signals using RQA

The mean values of the parameters corresponding to RQA associated with stride phase is tabulated by Prabhu and Pradhan [15]. The values tabulated corresponds to ALS, control, Huntington and Parkinson subjects. From the tabulation in [15], it can be observed that parameters related to RQA like *RR* and *DET* and L_{max} are minimal in control subjects, which infers that gait signals of control subject show less periodic nature when compared to ALS, Huntington and Parkinson subjects. The reason behind less periodicity is that the segmental links of control subjects involved in walking show high degrees of freedom. The other parameters like entropy is minimal for control subjects which infers that gait patterns of healthy subject is complex when compared to gait patterns of ALS, Huntington and Parkinson subjects. After referring to the tabulation, it is known that parameters corresponding to RQA are distinct and can help in classifying the subjects.

4.4. Binary classification using SVM and PNN

For the purpose of classification, we considered 6 binary classification problems viz., ALS vs. Control group (AC), PD vs. Control group (PC), HD vs. Control group (HC), ALS vs. PD (AP), ALS vs. HD (AH) and PD vs. HD (PH).

Binary classification problems are classified using two classification models, SVM (rbf and linear kernel) and PNN. Initially, the six statistically significant parameters of RQA were used as the feature for classification. LOOCV was used and performance outcome like accuracy, sensitivity, and specificity was calculated and tabulated in Table 3. Table 3 shows that the performance of SVM (rbf and linear kernel) is poor whereas the performance of PNN is comparatively better.

Performance outcome of classification problems was found to be improved by considering additional features like statistical measures. Instead of considering twelve features (six parameters of RQA and six statistical measures) for classification, the statistically significant features that show the significant difference between normal and neurodegenerative group were used. Inorder to raise the performance of classifier by reducing the computational burden, the optimal feature set was generated using feature selection method. LOOCV was used in measuring the performance outcome for SVM and PNN.

Table 4 tabulates the accuracy associated with 6 binary classification problems using optimal feature set. The performance of SVM classifier is improved after considering the statistical measures. From Table 4 it can be noted that most of the features are the parameters drawn after RQA. The accuracy of the entire classification group is 100%, except for that of AC. The reason is that the parameters drawn from RQA and statistical measures help in increasing the similarity index so that the classification performance is good.

Table 3
Classification result of SVM and PNN across six binary classification using six parameters of RQA.

Classification Task	Accuracy (%)		Sensitivity (%)		Specificity (%)	
	SVM(rbf)	PNN	SVM(rbf)	PNN	SVM(rbf)	PNN
AC	69	94.74	76	94	69	90
PC	61	94.44	72	91	65	90
HC	80	100	86	100	72	100
AP	69	88.89	76	0.92	69	90
AH	61	94.44	72	100	66	90
PH	73	93.33	81	90	76	88

Table 4
Classification result of SVM and PNN across six binary classification using optimal feature set.

Classification Task	Accuracy (%)		Sensitivity (%)		Specificity (%)		Feature Set
	SVM(rbf)	PNN	SVM(rbf)	PNN	SVM(rbf)	PNN	
AC	96.15	96.15	100	94	94	92	L_{max}, ENT, TT, V_{max}, Mean
PC	100	96.15	100	92	100	90	DET, L_{max}, Max
HC	100	96.92	100	100	100	96	RR, DET, Std
AP	100	100	100	100	100	100	TT, V_{max}, Max
AH	100	96.15	100	94	100	92	V_{max}, Max, Mean
PH	100	96.92	100	100	100	96	ENT, TT, V_{max}, Max

Table 5
Comparison of classification accuracy of current work with previous work using SVM.

Classification Task	Yang *et al.*(2009)	Li *et al.*(2014)	Xia *et al.*(2015)	Current work
AC	93.9	93.86	96.5	96.15
PC	86.43	85.89	100	100
HC	84.17	85.32	100	100
AP	85.47	85.09	96.43	100
AH	86.52	84.78	96.88	100
PH	79.04	79.48	91.18	100

Previous work of Xia et al. [22] showed an accuracy of 96.43%, 96.88% and 91.18% for the classification task AH and ph respectively using SVM. Table 5 shows that the current work improved the classification performance of classification task AP, AH, and ph to 100%. For the classification task PC and HC, current work and work carried out by Xia et al. [22] resulted in approximately same classification accuracy of 100%. For the classification task AC, the current work and the work by Xia et al. [22] show the same accuracy of 96%. The previous work considered features like Lempel-Ziv complexity, fuzzy entropy, and Teager-Kaiser energy to measure the complexity of gait signals. The gait signals were recorded for short durations of about 5 min, and to measure the complexity measures the gait signals should have been of long data length. The present study shows improved accuracy because RQA works well irrespective of data length. The other reason could be that RQA is able to draw the hidden relationship of gait signal (for example: periodic or chaotic nature) without assuming the signal to be stationary, linear and noiseless and thus extract the set of features.

5. Conclusion

Applying RQA on gait signals help in understanding dynamics of gait like periodicity and randomness. RQA does not rely on assumptions like non-linearity, non-stationarity, and noiselessness, and works well on short length gait time series. RQA obtained by embedding dimension i.e. gait signals acquired in one dimension will help in understanding the gait dynamics in all dimensions. This is because RQA relies on Taken's theorem. The main objectives of the work include understanding the dynamics of human gait, quantitatively analyzing the gait pattern of normal and pathological subjects and improving the accuracy of binary classification problems. The accuracy of binary classification problems like AP, AH and ph have improved by considering the parameters of RQA. The features are classified used SVM and PNN but of these SVM gives the better result.

Acknowledgments

The authors wish to thank Manipal Academy of Higher Education (MAHE) for supporting to accomplish the work.

References

[1] F. Censi, V. Barbaro, P. Bartolini, G. Calcagnini, S. Cerutti, Non-linear dynamics of atrial rate during atrial fibrillation assessed by recurrence plot analysis, in: Computers in Cardiology 1998, 1998, pp. 197–200, doi:10.1109/CIC.1998.731767.
[2] F. Censi, G. Calcagnini, S. Cerutti, Recurrence plot analysis of the coupling between respiration and cardiovascular variability signals, in: Computers in Cardiology 1997, 1997, pp. 211–214, doi:10.1109/CIC.1997.647868.
[3] P.-H. Chen, R.-L. Wang, D.-J. Liou, J.-S. Shaw, Gait disorders in parkinson's disease: assessment and management, Int. J. Gerontol. 7 (4) (2013) 189–193, doi:10.1016/j.ijge.2013.03.005.
[4] S. Dutta, A. Chatterjee, S. Munshi, Hybrid correlation-neural network synergy for gait signal classification, in: Advances in Heuristic Signal Processing and Applications, Springer Berlin Heidelberg, 2013, pp. 263–285.
[5] Dutton, M. Dutton, Physical Therapist Assistant Exam Review Guide, First, Jones & Bartlett Publishers, 2011.
[6] R. Hervás, S. Lee, C. Nugent, J. Bravo, Ubiquitous Computing and Ambient Intelligence: Personalisation and User Adapted Services: 8th International Conference, UCAmI 2014, Belfast, UK, December 2-5, 2014, Proceedings, in: Lecture Notes in Computer Science, Springer International Publishing, 2014.
[7] S. Hou, L. Li, R. Bo, W. Wang, T. Wang, Feature generation using recurrence quantification analysis with application to fault classification, in: System Science, Engineering Design and Manufacturing Informatization (ICSEM), 2011 International Conference on, 2, 2011, pp. 43–46, doi:10.1109/ICSSEM.2011.6081324.
[8] C. Kirtley, Clinical Gait Analysis: Theory and Practice, Elsevier, 2006.
[9] F.S. Labini, A. Meli, Y.P. Ivanenko, D. Tufarelli, Recurrence quantification analysis of gait in normal and hypovestibular subjects, Gait Posture 35 (1) (2012) 48–55, doi:10.1016/j.gaitpost.2011.08.004.
[10] Z. Li, W. Chen, J. Wang, J. Liu, An automatic recognition system for patients with movement disorders based on wearable sensors, in: 2014 9th IEEE

Conference on Industrial Electronics and Applications, 2014, pp. 1948–1953, doi:10.1109/ICIEA.2014.6931487.

[11] N. Marwan, Current Developments of Concepts Based on Recurrence Plots and Their Applications, Univ. of Potsdam, Physics Dept., Germany, 2003 Ph.D. dissertation.

[12] N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems, Phys. Rep. 438 (5) (2007) 237–329, doi:10.1016/j.physrep.2006.11.001.

[13] H. Negahban, M. Salavati, M. Mazaheri, M.A. Sanjari, M.R. Hadian, M. Parnianpour, Non-linear dynamical features of center of pressure extracted by recurrence quantification analysis in people with unilateral anterior cruciate ligament injury, Gait Posture 31 (4) (2010) 450–455, doi:10.1016/j.gaitpost.2010.01.020.

[14] H. Poincaré, On the problem eme of the three bodies and equations of dynamics, Acta Mathematica 13 (1890).

[15] P. Prabhu, N. Pradhan, Recurrence quantification analysis of human gait in neurological movement disorders, Int. J. Eng. Res. Technol. 5 (3) (2016) 447–452, doi:10.17577/IJERTV5IS030693.

[16] A.F. Rabbi, M.K. Jaiswal, S. Lei, R. Fazel-Rezai, Changes in dynamical characteristics of epileptic eeg in rats using recurrence quantification analysis, in: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 2562–2565, doi:10.1109/IEMBS.2011.6090708.

[17] F. Riva, M.C. Bisi, R. Stagni, Gait variability and stability measures: minimum number of strides and within-session reliability, Comput. Biol. Med. 50 (2014), doi:10.1016/j.compbiomed.2014.04.001.

[18] D.F. Specht, Probabilistic neural networks, Neural Networks 3 (1) (1990) 109–118.

[19] P.I. Terrill, S.J. Wilson, S. Suresh, D.M. Cooper, C. Dakin, Attractor structure discriminates sleep states: recurrence plot analysis applied to infant breathing patterns, IEEE Trans. Biomed. Eng. 57 (5) (2010) 1108–1116, doi:10.1109/TBME.2009.2038362.

[20] V. Vapnik, The nature of statistical learning theory, Springer Science & Business Media, 2013.

[21] C. Vaughan, B. Davis, J. O'Connor, Dynamics of human gait, Dynamics of Human Gait, 2, Human Kinetics Publishers, 1992.

[22] Y. Xia, Q. Gao, Q. Ye, Classification of gait rhythm signals between patients with neuro-degenerative diseases and normal subjects: experiments with statistical features and different classification models, Biomed. Signal Process. Control 18 (Complete) (2015) 254–262, doi:10.1016/j.bspc.2015.02.002.

[23] M. Yang, H. Zheng, H. Wang, S. McClean, Feature selection and construction for the discrimination of neurodegenerative diseases based on gait analysis, in: 2009 3rd International Conference on Pervasive Computing Technologies for Healthcare, 2009, pp. 1–7, doi:10.4108/ICST.PERVASIVEHEALTH2009.6053.

[24] M. Yang, H. Zheng, H. Wang, S. McClean, J. Hall, N. Harris, A machine learning approach to assessing gait patterns for complex regional pain syndrome, Med. Eng. Phys. 34 (6) (2012) 740–746, doi:10.1016/j.medengphy.2011.09.018.

[25] C. Yuan, X. Zhu, G. Liu, M. Lei, Classification of the surface emg signal using rqa based representations, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008, pp. 2106–2111, doi:10.1109/IJCNN.2008.4634087.