Health and Technology (2019) 9:547–560
https://doi.org/10.1007/s12553-018-0274-y

ORIGINAL PAPER

Rule based classification of neurodegenerative diseases using data driven gait features

Kartikay Gupta[1] · Aayushi Khajuria[2] · Niladri Chatterjee[1] · Pradeep Joshi[3] · Deepak Joshi[2]

Received: 25 September 2018 / Accepted: 31 October 2018 / Published online: 16 November 2018

Abstract

Classification of neurodegenerative diseases (NDD) like Parkinson's disease (PD), Amyotrophic Lateral Sclerosis (ALS), and Huntington's disease (HD) is of high clinical importance. The gait analysis based classification is attractive due to its simplicity and noninvasiveness. In this paper, we propose a data driven features approach along with autocorrelation and cross correlation between gait time series to create different feature set for a sample representation. Further, a rule based classifier using Decision Tree is trained with those features to classify the neurodegenerative diseases from healthy controls. Mutual Information (MI) analysis revealed the dominance of data driven features over auto and cross correlation based features. The classifier fed with top 500 features could produce the classification accuracy of 88.5%, 92.3%, and 96.2% for HD vs. control, PD vs. Control, and ALS vs. control. Pooling all neurodegenerative samples into one as NDD class and applying current approach produced nearly 87.5% of accuracy for NDD vs. control. Finally, we validated the present approach for a challenging situation of classification of less severe patients and observed respectable accuracies of 80%, 80%, 90%, and 73.33% for HD vs. control, PD vs. Control, and ALS vs. control, and NDD vs. control, respectively. The proposed algorithm shows potential for rule based classification system in data driven features for Neurodegenerative disease classification.

Keywords Neurodegenerative disease classification · Rule based classifier · Decision tree

1 Introduction

Around 30,000 people are diagnosed with Amyotrophic Lateral Sclerosis (ALS), same with Huntington's disease (HD) and 1,000,000 with Parkinson disease (PD) each year in the United States [1]. The prevalence rate of the spectrum of neurological disorders in India has a mean of 2394 per 100,000 populations, providing a rough estimate of over 30 million people with neurological disorders (excluding neuro-infections and traumatic injuries) [2]. A progressive Neurodegenerative disorder like ALS affects the nerve cells in the brain and the spinal cord. The ability of the brain to initiate and control the muscle movements is lost with the degeneration of motor neurons [3]. The primary effect of Parkinson's disease is on dopaminergic (the dopamine-producing neurons) in the substantia nigra part of the brain. Bradykinesia, gait and balance problems, rigidity of limbs, tremor are some of the symptoms of Parkinson's disease [4]. Identification of factors that contribute to mobility and gait impairments due to neurological impairments such as Parkinson disease has been done [5] which shows balance is the most relevant factor for the same. Thus exercise interventions focusing on balance may be best able to impact gait and mobility in Parkinson's disease. HD is caused due to the expansion of CAG trinucleotide in Huntingtin's gene, which causes polyglutamine repeat in the huntingtin protein. It is a protein misfolding movement disorder in basal ganglia which includes

✉ Deepak Joshi
joshid@cbme.iitd.ac.in

Kartikay Gupta
kartikay000555000@gmail.com

Aayushi Khajuria
Aayushi.Khajuria@cbme.iitd.ac.in

Niladri Chatterjee
niladri@maths.iitd.ac.in

Pradeep Joshi
joshipradeep_2004@yahoo.com

[1] Department of Mathematics, Indian Institute of Technology Delhi, New Delhi, India

[2] Center for Biomedical Engineering and Department of Biomedical Engineering, Indian Institute of Technology Delhi and All India Institute of Medical Sciences (AIIMS), New Delhi, India

[3] School of Business, Quantum University, Roorkee, India

chorea, tremor, motor restlessness and myoclonus thereby causing gait impairments [6]. The symptoms of PD, ALS, and HD, collectively called as neurodegenerative diseases (NDD), are not specific with easily overlooked nature. It creates a significant overlap to each other among NDDs; leading to a misdiagnosis. As per the report of The Michael J. Fox Foundation for Parkinson's research, up to 25% of Parkinson's disease diagnoses are incorrect. Such misdiagnosis put those patients on wrong drugs and delays the correct treatment. Neuroimaging tests, Computer Tomography or Magnetic Resonance Imaging, sometimes along with blood and urine tests, are the state of the art for NDDs diagnosis at present. However, a recent study [7] concluded that conventional Magnetic Resonance Imaging (MRI) was not found to be a reliable diagnostic tool for ALS with a sensitivity and specificity of 48% and 76%, respectively. In addition, MRI and CT are expensive, time-consuming, and require specific skills. Hence, there is a need of alternate diagnosis method which is quick, low-cost, and can be easily operated without specific skills. With this motivation, in the present approach we utilized the gait variable measurements to explore the feasibility of NDD diagnosis from walking pattern of the individuals.

Previous works The movement disorders due to NDDs decline the ability of a person to walk properly and lead to a disturbed gait cycle. The analysis of gait parameters affected due to such diseases has applications in explaining neural component of locomotion and developing an automated noninvasive classification methodology. Among all gait parameters, the spatiotemporal variables of gait cycles utilize the simplified instrumentation and hence suitable for real time low resource settings [8]. The effectiveness of backward walking on spatio-temporal gait variables has been reviewed [9] where it is concluded that backward training could improve participants spatio-temporal gait characteristics and is potentially useful in neurological rehabilitation. Utilization of few foot switches provides information of stance, swing, and double support intervals from both lower limbs [10, 11]. Due to binary nature of switches, this information is directly available, with minimal computation, in the form of time series for processing and does not require any extensive pre-processing method before analysis. A sensor network is also proposed that allows to capture knee-ankle data in children while they walk for the purpose of gender classification [12]. This makes it attractive for utilizing in neurodegenerative disease classification also. Various features and classification methods have been reported earlier which utilized the stance, swing, and double support interval time series to classify neurodegenerative diseases. Time-domain characterization of these gait intervals followed by pattern recognition techniques has shown respectable accuracy [13–15]. Introducing a signal turn count (STC) feature along with other time domain feature improved the classification accuracy to 90.32% [16]. STC being a deemed parameter of frequency [17] could have introduced frequency representation of gait intervals and thus provided remarkable classification accuracy. Motivated by this, our recent work [18] utilized wavelet transformation based time-frequency representation of gait interval and achieved similar accuracy by using less input information i.e. only one gait interval time series. Improvement in the accuracy up to 100% was observed after pooling the wavelet features from all the gait interval time series. Similar results are reported from other researchers where wavelet transform based coherence and entropy have been proved useful for the classification of the control and the NDD patients [19]. Various other feature extraction methods such as maximum signal-to-noise ratio (MSNR), maximum signal-to-noise ratio combined with minimum correlation (MSNR & MC), maximum prediction power combined with minimum correlation (MPP & MC) and principal component analysis (PCA) provided a remarkable classification accuracy for the classification of patients with neurological disorders against controls [20]. Also approaches like deterministic learning theory, empirical mode decomposition, phase synchronization and conditional entropy have been shown of great potential in the categorization of controls and NDD patients [14, 21, 22].

RBC (Rule Based Classifier) has advantage of being interpretable and are "white-box" model in contrast to other available classifiers. While interpretability has been confusing and underspecified in many ways earlier [23], in our present work, we refer interpretability in the context that they are basically the "white boxes" in the sense that the acquired knowledge can be expressed in a readable form like if-else compared to just some matrices or mathematical representation, while other classifiers like KNN, SVM, Neural networks are generally "black boxes" that is we cannot read the acquired knowledge in a comprehensible way. The importance of interpretability lies in ability of users to understand the model. Having an interpretable model like Decision Tree reveals new hidden pattern and serves as a positive feedback to the user/expert. For example, if we develop an interpretable model to predict the severity of disease and if the user/expert is a clinician he can bring the expert knowledge domain to correlate the particular feature, critical for classification as adjudged by Decision Tree, to clinical symptoms based on individual patient's history and condition. In summary, interpretability facilitates more generalized model to be handled by user/expert. Even though a good predictor would certainly be useful in practice, making a model that reveals the reasons why the outcome was wrong in specific cases would be much more meaningful and would enable the experts to design better model in the future. Also the rule based classifier needs low computational resources for implementation in hardware and the gait variable measurement used in the present study are easily and quickly measurable, therefore, one of the application of present work would be to develop a portable system to be used as wearable system for patient to observe any abnormality in gait pattern at initial stages and timely consult a physician.

We selected decision tree (DT) as a RBC representation for our work. The decision trees are easy to use, free of ambiguity and robust even in the presence of missing values [24, 25].

The decision tree was trained by three types of features as following – 1) Autocorrelation based features 2) Data Driven Features and 3) Correlation between time series. Autocorrelation in time provides an explicit estimation of frequency and hence indicate some information of frequency content in the signal without actual transformation in frequency domain [26] while saving the computation resources. Data driven features are the human observations that brings qualitative approaches combined with quantitative approach. Human observer can highlight the essential, clinically meaningful parts, thereby providing the quantitative approaches with a more relevant subset of the available data. Therefore, data driven features are important as they use the visual information captured by the expert and have been shown useful in representing various bio signals. For example, feature extraction has been performed using data driven methods from night sleep PSG (Polysomnography) recordings for sleep/wake stage classification [27]. Finally, correlation feature between gait interval time series is used to introduce the bilateral coordination during walking. It has been shown recently [22] that considering the coordinated locomotor pattern between both legs showed impressive NDD classification accuracy. We first generate large number of features using all these three types of features, then do feature selection using mutual information (MI) and finally train a decision tree classifier. Finally, we validated the present approach for a challenging situation of classification of less severe patients for a more realistic and meaningful clinical applications.

2 Materials and methods

2.1 Gait database description

The gait database used is freely available on the web page of Physionet[1] [28]. The record contains the gait parameter intervals that are taken in the real time for control ($n = 16$; 2 males and 14 females) and NDD (Parkinson's disease - $n = 15$; 10 males and 5 females, Huntington's disease - $n = 19$; 6 males and 13 females, Amyotrophic lateral sclerosis - $n = 13$; 10 males and 3 females) patients. This database reports time interval of gait parameters (stance, swing, double support, and stride) from both legs. In the experiment [29], each subject was requested to walk at his or her normal pace along a straight hallway of 77 m in length for 5 min without stopping (unless he or she had to turn at the end of the hallway) on level ground. Force signals from ultrathin force sensitive switches inside each subject's shoes were recorded with a sampling frequency of 300 Hz. These force signals were used to determine stance, swing, stride, and double support phase interval. The database also quantifies the severity of NDD in the respective category. A Hohn and Yahr score is provided which gives the severity of the Parkinson's disease and varies from 1.5 to 4. A total functional capacity measure for Huntington's disease is also provided which varies from 1 to 12. For the patients suffering from Amyotrophic Lateral Sclerosis this database gives the severity since the diagnosis of the disease. As the dataset is imbalanced the present study utilizes random under sampling to balance class distribution by randomly eliminating majority class examples. Two scenarios were considered while balancing the dataset – 1) Control vs. Parkinson, Control vs. HD, Control vs. ALS and 2) Control vs. NDD. In scenario#1, the minimum of 13 subjects in each category were available to balance the dataset i.e. uniform distribution. If we choose $n = 13$ we will have equal number of observations from each category for scenario #1. Therefore a new dataset was derived with n = 13 from each category. Now in this new dataset for scenario #2 i.e. Control vs. NDD 13 subjects in control are available and to match the equal number of patients in NDD group 13 patients were needed from NDD. However, NDD consist of three classes and it is not possible to take equal observations from all three classes and make it 13 – the number 13 is not completely divisible by 3. Therefore 12 was the preferred choice as taking $n = 12$ for NDD makes equal observations (4 patients) from each three category namely Parkinson's disease, Huntington's disease, and ALS disease thus making the dataset balanced for Control vs. NDD classification. Table 1 shown below represent the summary of demographics of various groups.

To validate our proposed approach we used the dataset of less severe patients ($n = 5$) in each category (PD, HD and ALS). Classifying less severe patients who are in the early stages of the disease would be more challenging and will have a rich clinical applications with assisting the clinicians for better diagnosis. Table 2 shown below represent the summary of demographics of various groups belonging to less severe patients.

2.2 Processing of the data

In order to neglect the startup effects we removed the data of first 20 s. As described in the Physionet database that the significantly different strides were detected when the patients have to turn around the end of hallway space of walking. These strides were considered as outliers and were identified as the data point with the value three standard deviations greater or less than the median value [16]. These outliers were replaced with the median value of the corresponding time series because simply just excluding the outliers from the analysis would firstly shrink the data points of the time series and secondly decrease the variance in the data and cause a bias based on under or overestimation [30]. The median value is a measure of central tendency and offers the advantage of being

[1] Online available at http://www.physionet.org/physiobank/database/gaitndd/

Table 1 Summary of Demographics and severity measures of different groups

S.NO.	GROUP	AGE (yrs) (mean ± SD)	HEIGHT (meters) (mean ± SD)	WEIGHT (kg) (mean ± SD)	Gait speed(m/s) (mean ± SD)	Severity Measures (mean ± SD)
1.	Control (*n* = 13)	42.23 ± 18.89 (range, 22–74)	1.83 ± 0.083	69.46 ± 10.44	1.37 ± 0.171	NA
2.	Parkinson's Disease Patients (n = 13)	68.46 ± 9.53 (range,53–80)	1.88 ± 0.15	76.0 ± 16.98	0.97 ± 0.203	3 ± 0.73 (H & Y score)
3.	Huntington's Disease Patients (n = 13)	47.53 ± 11.11 (range,33–71)	1.85 ± 0.094	71.61 ± 15.53	1.24 ± 0.32	7.2 ± 3.46 (Total functional capacity measure)
4.	ALS Disease Patients (n = 13)	55.61 ± 12.82 (range,36–70)	1.74 ± 0.09	77.11 ± 21.14	1.05 ± 0.21	18.30 ± 17.81 (Time in months since the diagnosis)

very insensitive to the presence of outliers [31]. That is why the outliers which are with the value three standard deviations greater or less than the median value are replaced with the median value of the corresponding time series. Figure 1 shows the seven time intervals for a representative sample from each group.

2.3 Feature extraction

The gait signals used in the present study consists of seven gait intervals in form of time series given in the database which are left and right stride interval, left and right swing interval, left and right stance interval and double support interval. From the data

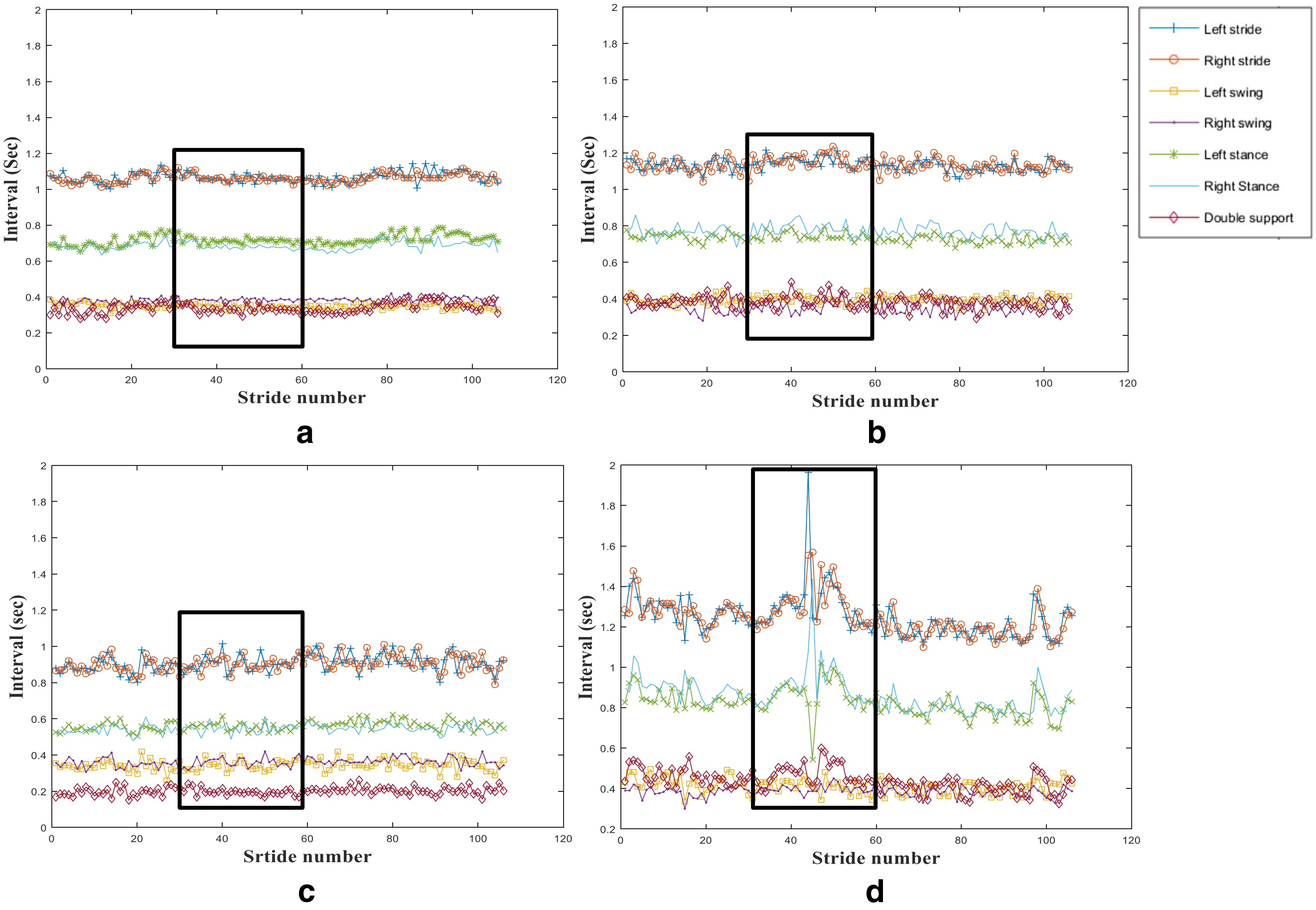


Fig. 1 Time series plots for one representative sample from each class. [**a** - Control, **b** – Parkinson's Disease, **c** - Huntington's Disease, **d** –ALS]. The black box highlights the standard deviation of subsequences from 30 to 60 (x-axis) for the 'Control' group which is quite low as compared to other groups. This difference becomes more prominent in the left and right swing interval time series

set, highlighted in the box, please see Fig. 1, it can be easily visualized that the standard deviation along certain dimensions in controls is remarkably low from NDD. Thus, while generating features it was made sure that all such features would be included in the analysis. Overall, 7546 features were extracted from each of the samples to aid in classification, which is mentioned below:

Auto-correlation based features As mentioned earlier the auto-correlation based features provide an explicit indication of frequency contents in the signal. Following autocorrelation analysis was performed and further features were selected as below:

a) Pearson correlation at different lag values between the elements of the time series was calculated. The lag considered was from 0 to 100. Here, the length of time series is 106. Thus, this gives us 101 features for each dimension. Formally, auto-correlation is defined as:

$$\rho(h) = \sum_{i=1}^{N-h}\left(y_i - \overline{y}\right) \times \left(y_{I+h} - \overline{y}_h\right) \Big/ \sqrt{\sum_{i=1}^{N-h}\left(y_i - \overline{y}\right)^2 \times \sum_{j=h}^{N} y_j - \overline{y}_h\Big)^2} \qquad (1)$$

Where:

- $y_i$is the value of the time series at time 'i'
- h is the lag
- N is the total number of time stamps in time series

$$\overline{y} = \sum_{i=1}^{N-h} y_i/(N-h) \qquad (2)$$

- $\overline{y}$ is the mean of time series from 1 to N-h

$$\overline{y}_h = \sum_{i=h}^{N} y_i/(N-h) \qquad (3)$$

- $\overline{y}_h$ is the mean of time series from h to N

b) Each of this time series was first differenced to obtain another time series of length 105. Then, similarly as in Eq. 1, Pearson correlations at different lag values were found. This gave additional 101 features.

Hence, from autocorrelation based features we get 1414 features. This can be explained as:

$$\left[7^{*}\ 101^{*}\ 2 = 1414\right]$$

Where 7 is the number of times series for each subject

101 are the different time lags considered for different time series.
2 is the type of time series that is first is the original time series and second is the differentiated version of the original time series.

Data driven features Data driven features were observed heuristically in the time series as an observer. As mentioned earlier, visual observation suggested the remarkable difference between controls and NDD at different time windows. Following features were calculated accordingly:

a) Mean and Standard Deviation of the time series was calculated and added to the features list.
b) A window size was determined say 'w' and then moving average and moving standard deviation were calculated at each point i.e. average and standard deviation of all sets of continuous 'w' points were added to the feature sets, shown in Fig. 2. This gives 2*(107 – w) features for each

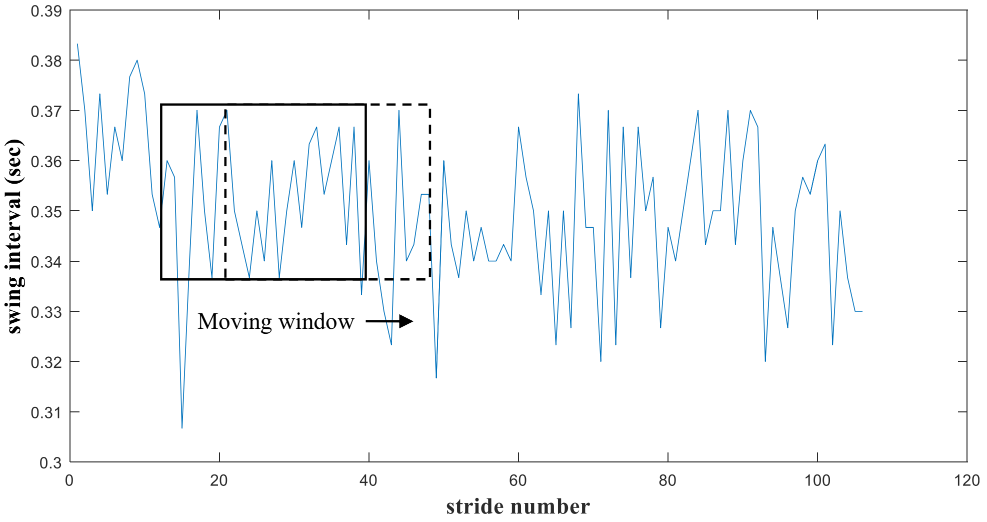


Fig. 2 Statistical features were extracted from moving windows of different sizes. Mean and Standard deviation for the data points inside the window frames were calculated and added to the features pool

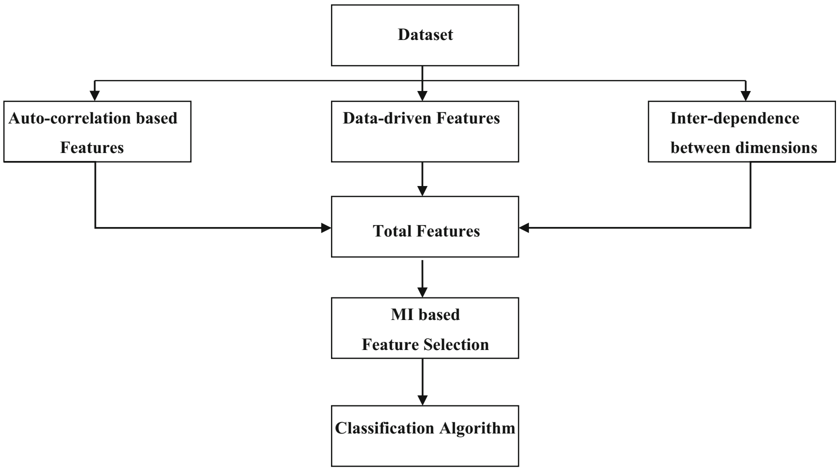

Dataset
Auto-correlation based Features
Data-driven Features
Inter-dependence between dimensions
Total Features
MI based Feature Selection
Classification Algorithm

Fig. 3 Flow chart representing features, their selection methodology and classification

time series and window size ‘w’ in the experiment. The window sizes used were 5, 10, 20, 30, and 40 timestamps.

Hence, from data driven based features we get 6020 features. This can be explained as:

$$\left[7^{*}5^{*}2=70\right]$$

Where 7 is the number of times series for each subject

2 is the type of time series that is first is the original time series and second is the differentiated version of the original time series.

And 5 (7 *2 = 14 features representing mean and standard deviation along complete separate time series. The same features were added 5 times corresponding to each window.)

As Windows = [5, 10, 20, 30, 40].

Suppose ‘w’ is 5, then first window would be from 1 to 5, second from 2 to 6, third from 3 to 7 and finally from 102 to 106. Thus, in total we have 106–4 windows. This can be written as:

$106-4=106-(5-1)=107-5$

Thus, (107-w) features were calculated for each time series.

Therefore, Sum (7*(107 - w)*2) = 6020 # sum over w, where w is the window size.

Inter-dependence between time series This feature finds inter and intra limb coordination in and among various gait intervals. Following features were extracted under this category:

a) Correlation between each of the seven dimensions was determined and added to the features list. This gave 21 additional features. Formally, correlation is defined as:

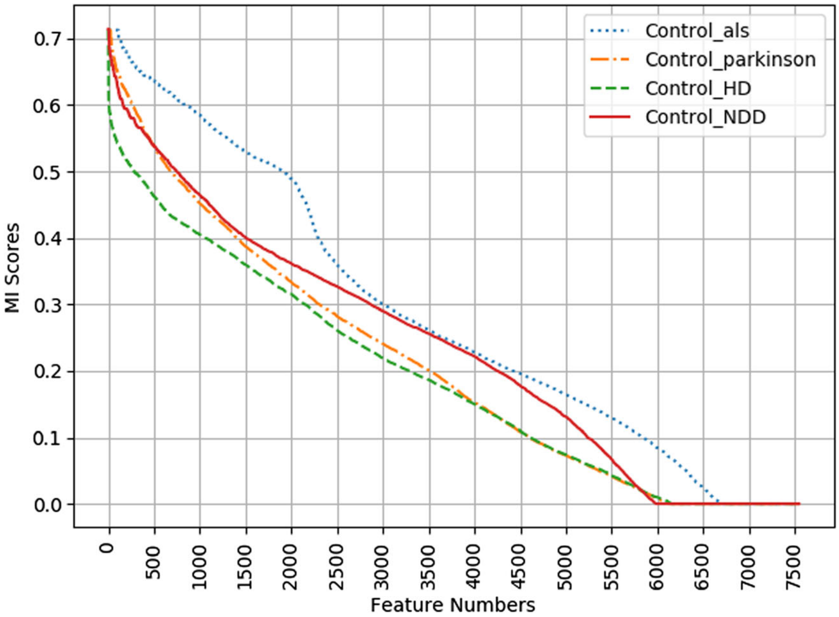


Fig. 4 MI score between features and class labels for different binary classification tasks. 500 top features were selected for constructing classification tree in all the tasks

$$cor(X,Y) = \sum_{i=1}^{N} \left(y_i - \overline{y}\right) \times \left(x_i - \overline{x}\right) / \sqrt{\sum_{i=1}^{N} \left(y_i - \overline{y}\right)^2 \times \sum_{j=1}^{N} \left(x_j - \overline{x}\right)^2} \quad (4)$$

Where:

- $N$is the total number of time stamps in time series.
- y_i, x_iIs the value of the time series X and Yat time 'i'.

$$\overline{y} = \sum_{i=1}^{N} y_i/(N) \quad (5)$$

- $\overline{y}$ is the mean of time series Y

$$\overline{x} = \sum_{i=1}^{N} x_i/(N) \quad (6)$$

- $\overline{x}$ is the mean of the time series X

b) Each of the time series was first differenced to obtain another set of seven time series. Then, the correlation between each of the time series was used as a feature. This also added another 21 features.

This gives additional 42 features which can be explained as:

$$[21 + 21 = 42]$$

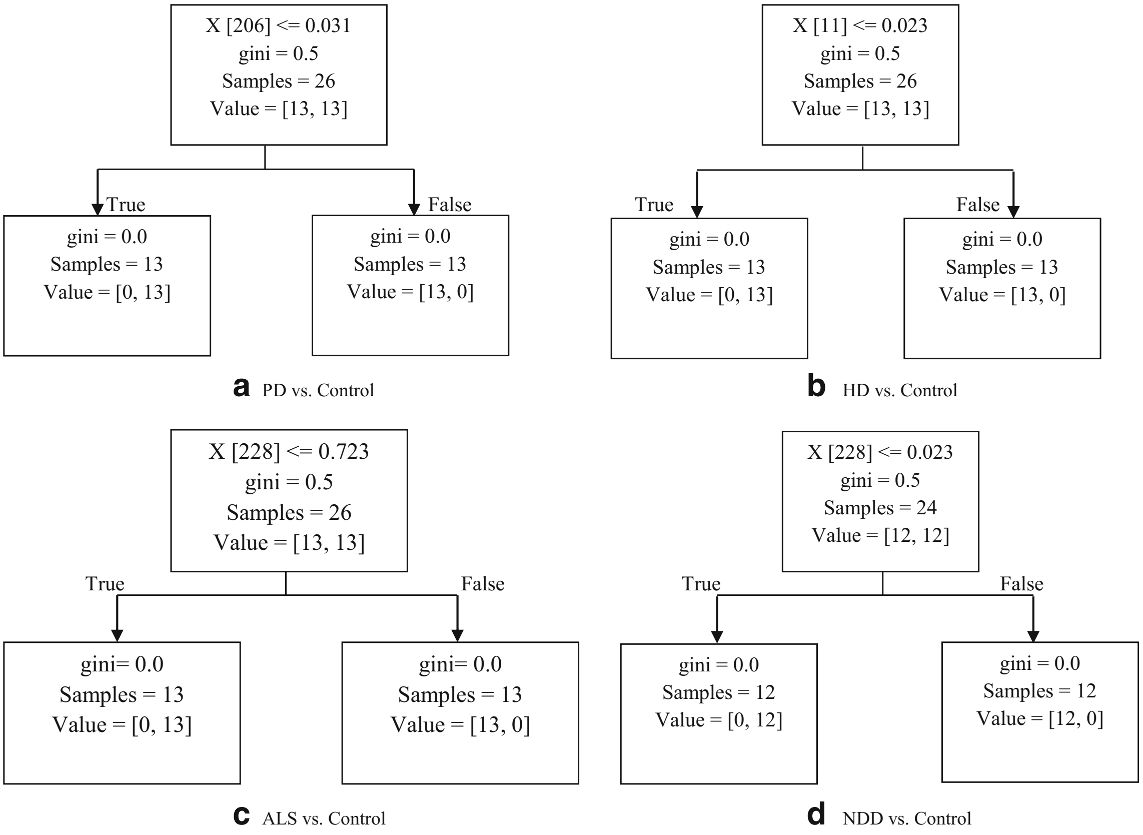
Here 21 is the correlation between the two type of time series that is first is the original time series and second is the differentiated version of the original time series.

Hence, total features were:

$$\mathrm{Sum}\ ([42, 1414, 6020, 70]) = 7546$$

2.4 Feature selection

Mutual information (MI) between each of the features and the class label was determined. Then, those features were retained which had higher MI. High MI value depicts less randomness between the values of the two sets. Low MI shows that the values of the two sets are mostly independent. Hence one variable cannot be used to predict the other variable, if MI value is low.



a PD vs. Control

b HD vs. Control

c ALS vs. Control

d NDD vs. Control

Fig. 5 Decision Tree classifiers obtained for the 4 binary classification tasks. The features used in these trees are described in Table 5. X [n] denotes the statistic value of feature 'n', gini indicates the gini value of all samples within the box, samples indicate the number of training samples reaching that box in the decision tree, and value given by [x, y] denotes the 'x' diseased sample and 'y' healthy/control sample reaching that box. Tables 6, 7, and 8

Formally, MI between two random variables is defined as:

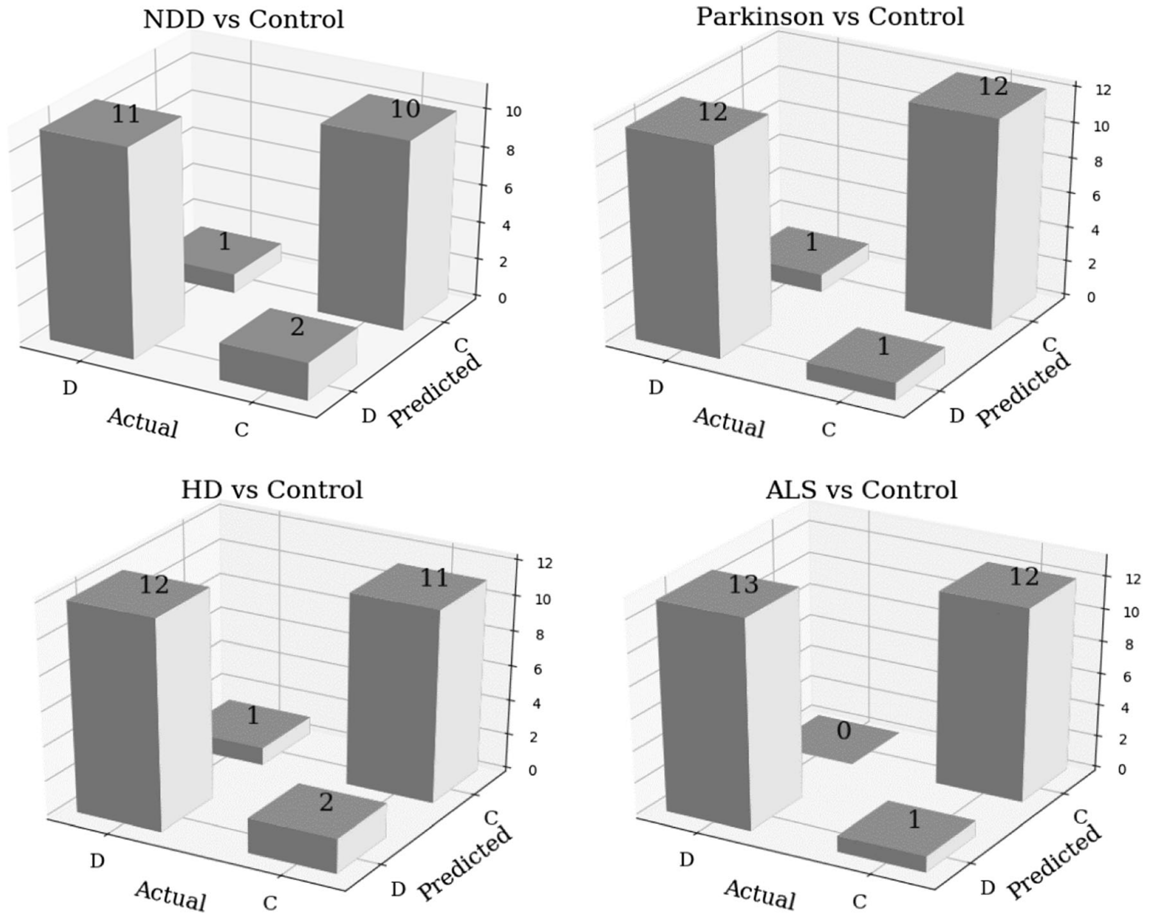
$$MI(X;Y) = \iint p(x,y) \times \log\left(\frac{p(x,y)}{p(x)p(y)}\right) dxdy \qquad (7)$$

Where *p(x, y)* is the joint probability density function of X and Y, and *p(x)* and *p(y)* are the marginal probability density functions of X and Y respectively. In the present study X is a feature and Y is associated label class. In case of discrete valued random variable integral is replaced by summation and probability density function is replaced by probability mass function in Eq. 7. Here, the label class is a discrete valued random variable. High MI value indicates high chances of predicting that class correctly using the features, hence high MI value features was used for further analysis.

The shortlisted features were used for constructing classification tree. Features with low MI score are not used for classification as it may lead to construction of poor trees. These features may get included in the decision rules at the lower branches and thus lead to construction of poor rules. Thus, only top 500 features were used for constructing classification tree. Figure 3 shown below represents the flowchart of the methodology employed in the present approach.

2.5 Classification and evaluation

Decision Tree Classifier, as implemented with the name ‘Decision Tree Classifier’ in the scikit-learn module of python (version 0.19.0) [32], was used for classification. The aforementioned implementation was used with default values for training the decision tree classifier. Still, different runs of the same algorithm may produce slightly different results due to randomness inherent in the algorithm. The algorithm randomly selects a feature from the pool of all features, without taking into consideration any specific order, and then selects the best split point in that feature. This way, it goes through all the list of features. So, if two split points are equally good, then the order in which they are found becomes important. Hence, this brings slight randomness in the algorithm. Decision Tree Classifier is chosen because the classifier needs to be trained on 500 features and the training samples are very few. This classifier itself does feature selection to find the best splits. Also, the decision tree provides interpretability (set of rules) to the overall classification system. In contrast, SVM and ANN classifiers would simply over fit on such a training data set where number of features far exceed the number of available training samples. Also, they do not provide any set of


NDD vs Control

Parkinson vs Control

HD vs Control

ALS vs Control

Fig. 6 Confusion matrices for all the category of binary classification

Table 2 Summary of Demographics and severity measures for less severe patients ($n = 5$)

S.NO.	GROUP	AGE (yrs) (mean ± SD)	HEIGHT (meters) (mean ± SD)	WEIGHT (kg) (mean ± SD)	Gait speed(m/s) (mean ± SD)	Severity Measures (mean ± SD)
1.	Control	33.2 ± 14.99 (range,22–52)	1.87 ± 0.07	71 ± 7.07	1.50 ± 0.047	NA
2.	Parkinson's Disease patients	64 ± 15.11 (range, 44–80)	1.86 ± 0.12	78.4 ± 14.53	1.13 ± 0.151	1.8 ± 0.27 (H & Y score)
3.	Huntington's Disease patients	41.8 ± 7.36 (range, 36,54)	1.86 ± 0.08	80.4 ± 15.5	1.44 ± 0.311	10.6 ± 1.67 (Total functional capacity measure)
4.	ALS patients	51.8 ± 14.58 (range, 36–68)	1.78 ± 0.07	86.54 ± 12.05	1.16 ± 0.2	5.1 ± 3.17 (Time in months since the diagnosis)

rules for interpretability like the Decision Tree Classifier. K nearest neighbor classifier would also become inefficient due to the curse of dimensionality [33]. Leave One Out cross validation scheme was used for evaluating the classifier. In each iteration of the validation scheme, the following steps were done. One sample was removed from the complete data set for testing. The rest of the samples were used for feature selection and training the classifier. Then, the prediction of the classifier on the test sample was noted. The performance of the classifier was evaluated using sensitivity, specificity and accuracy and was calculated as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \times 100 \tag{8}$$

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \times 100 \tag{9}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{FN} + \text{TN} + \text{FP})} \times 100 \tag{10}$$

Where TP is true positive, TN is true negative, FP is false positive and FN is false negative value for the evaluation of classification performance. A confusion matrix was also constructed for further evaluation of classifier.

Table 3 ANOVA Test for demographics differences across groups

S. No	Demographic variables	F- ratio	p value
1.	Age	9.205	0.000
2.	Height	3.771	0.016
3.	Gait Speed	7.864	0.000
4.	Weight	0.647	0.589

3 Results

A one-way analysis of variance (ANOVA) test was conducted to observe any differences in demographics across groups in Table 1. The results show that there was a significant difference across the groups in all demographic variables, except weight, as shown with the corresponding p value in Table 3 and 3.

Post hoc analysis showed a significant difference ($\alpha = 0.05$) across the groups with respect to age, height and gait speed. The findings are shown in Table 4. As shown in Table 4, control group was significantly different from ALS in gait speed and from Parkinson in age and gait speed.

The hyper-parameter 500 for feature selection was chosen by cross-validation on 4 values (250, 500, 1000, and 1900) in one of the classification task. The cross-validation results did not have much difference in terms of accuracy. Selection of 500 features ensured the MI values of nearly 0.5 or more for all the classes and was most obvious choice due to optimality between number of features to be selected and high MI values. In further analysis of the top 500 features, it was found that all 500 features were data-driven features. Figure 4 shown below the MI score between features and the class labels for different binary classification tasks.

Table 4 Post Hoc test for multiple comparisons across the groups

S. No.	Demographic variables	Differences between the groups	p value
1.	Age	Control and Parkinson's disease	0.000
		Huntington's disease and Parkinson's disease	0.001
2.	Height	ALS and Parkinson's disease	0.014
3.	Gait Speed	Control and Parkinson's disease	0.000
		Control and ALS	0.005
		Huntington's disease and Parkinson's disease	0.024

Table 5 The description of various features employed in the decision trees

Classification Tree	Feature Number	Feature Description
NDD vs. Control	228	Standard deviation of subsequence from 12 to 51 in right stance interval time series
Parkinson's Disease vs. Control	206	Standard deviation of subsequence from 16 to 45 in right stride interval time series
Huntington's Disease vs. Control	11	Standard deviation of subsequence from 2 to 41 in right stance interval time series
ALS Disease vs. Control	228	Mean of subsequence from 41 to 70 in left stance interval time series

Figure 5 gives decision trees for each of the binary classification tasks. The decrease in gini value from higher node to a lower node of a tree denotes the strength of the split. Higher decrease indicates better splits for a given tree. Gini impurity for a set of items with J classes (1, 2, 3 … J), and p_i denoting the fraction of items labeled with class i, is given by:

$$gini = 1-\sum_{i=1}^{J} {p_i}^2 \tag{11}$$

Figure 5 (A) depicts a decision tree which classifies Parkinson's disease versus control. If the standard deviation of subsequence from 16 to 45 in right stride interval time series is less than or equal to 0.031, then the subject is classified as control. If this statistic is greater than 0.031, then the subject is classified as Parkinson's disease patient. Similarly, all other trees can be deciphered. A summary of features in all trees of Figs. 5 and 6 is given in Table 5.

Table 6 Description of various features employed in the decision trees for less severe patients

Classification Tree	Feature Number	Feature Description
NDD vs. Control	3	Standard deviation of subsequence from 34 to 53 in right stance interval time series
NDD vs. Control	310	Standard deviation of subsequence from 73 to 77 in left stance interval time series
Parkinson's disease vs. Control	338	Standard deviation of subsequence from 38 to 67 in left stance interval time series
Huntington's Disease vs. Control	338	Standard deviation of subsequence from 59 to 68 in right stance interval time series
ALS Disease vs. Control	338	Mean of subsequence from 76 to 105 in right stride interval time series

4 Discussion

Previous studies have demonstrated importance of statistical [16], frequency [18], and bilateral limb coordination features [22] in NDD classification. We utilized a combination of these approaches keeping in mind the human visual observation of the data. This combined approach of features led to a high dimensional data of greater than 7000 which was reduced to 500 based on mutual information criterion. MI based analysis revealed the dominance of data driven features over other auto and cross correlation based features. Among top ranking 500 features with MI value nearly 0.5 or greater all were data driven features. Utilizing these data driven features in current approach produced better accuracy than previously reported accuracies in any category of binary classification. For example, the classification of Parkinson's disease was achieved with 90.32% in previous work [16] using time domain and STC features, however current work improved the accuracy up to 92.3% by utilizing data driven features. Similarly, previous work [13] reported the classification accuracy of 82.8% while classifying ALS using the mean of the left-foot stride interval and the modified Kullback-Leibler divergence (MKLD). Our work shows superiority of data driven features by classifying ALS with 96.2%. It has to be noted that the classifiers are different in both the previous studies compared to our work which may account for differences in classification accuracies. Present study used Decision Tree classification compared to SVM in previous studies as Decision Tree is more interpretable than SVM. However, for NDD vs. control classification present approach underperform comparative to some previously reported accuracies [34]. We attribute the higher accuracy in previous work [34] to the unbalanced dataset used. The previous work used unbalanced dataset for the classification (20 patients with HD, 13 patients with ALS, 15 patients with PD and 16 healthy controls) and for NDD vs. control (48 patients with NDD and 16 healthy controls) which might have led to over fitting and biased classification accuracy. However, in the present work, in order to develop an unbiased and not an over fit classifier the number of subjects in each NDD category was compromised to have a balanced dataset ($n = 13$ in each category) and for NDD vs. control (12 patients with NDD and 12 healthy controls). Further, pooling Huntington's disease in NDD might have deteriorated the classification accuracies by narrowing down the classification margin - the classification accuracy for Huntington's disease vs. control was lower with 88.5%. Contrary to classifiers like Support Vector Machine (SVM) and others in previous studies, Decision tree classifiers do features selection itself by choosing the best splitting point amongst all features to create

Table 7 Sensitivity, Specificity and Accuracy (all in %) values of the classifier for the classification

NDD vs. Control			Parkinson's Disease vs. Control			Huntington's Disease vs. Control			ALS Disease vs. Control		
Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
84.6	90.9	87.5	92.3	92.3	92.3	85.7	91.7	88.5	92.8	100	96.2

split between the data and the final classifier is based on only very few features and thus minimizing over fitting. A Leave one out cross validation (LOOCV) method was performed further to avoid the redundant features in an attempt to minimize over fitting with the small dataset in the present study.

A detailed comparison of present work with previous reported work is shown in Table 9.

It is interesting to see that the visual pattern were dominating and successfully transferred to decision tree similar to previous research [27]. Recently, it has also been shown that the expert knowledge improve automatic probabilistic classification of gait joint motion patterns in children with cerebral palsy [35]. In this paper we have followed the similar mechanism and visual input from the researchers, as shown in Fig. 1, that in control subjects the standard deviation of data points in all gait variables from 30 to 60 is highly different from NDD patients were embedded in features list which proved efficient as all dominating features selected by Decision tree were data driven features. Though data driven features provided superior accuracy but it led to very high dimensional feature space. Mutual information (MI) was utilized for data reduction in the present work. An MI approach was preferred for dimensionality reduction over Principal Component Analysis (PCA) because PCA loses physical interpretation after linearly transformation of original variables, while MI retains the physical interpretation which suits to Rule based classification. The previous study [21] reported that the random forest classifier has the best average performance amongst all classifiers. Random forest classifier is the ensemble of decision tree classifiers. Also, random forest classifiers are not as interpretable as single decision tree classifier. Hence, classification is done using Decision Tree classifier. Previously, Random forest has shown lower classification accuracy than SVM for the same database [34], however in the present work we could reach to the at par accuracy using hybrid approach features and decision tree.

With the encouraging results in Table 8 for classification in less severe patients, we believe the present work has potential to serve both – 1) the physician and 2) the patient. The work has a potential to be translated for the benefit of physicians and the patient. In one hand, the physician can use the rule based classifier to identify the stage of the patient and thus decide the diagnosis, while on the other hand the system is portable to be used as wearable system for patient and observe any abnormality in gait pattern at initial stages and timely consult a physician. Despite of overlapping walking speed at early stages of NDDs, as shown in Table 2, the rule based classifier provides respectable accuracy in classification showing the impact of approach in the present work. Thus with more number of participants, the present approach has enough scope to improve the classification accuracy in less severe participants.

The primary motivation to adapt the present methodology for real time implementation is of two fold. One was to use the minimal computational resources, for example implementable on any smart phone, so that the overall system is portable, wearable and can be used as home-based solutions. The present approach practically needs only a switch-based insole and a smart phone for implementing the proposed system. With this home-based solution we target the elderly population who are unable to visit clinics on a regular basis. Second, a real time implementation will facilitate a quick diagnosis thus saving the time of physicians and the patients both.

Present study has some limitations too. We utilized only one classifier but this attributes to the choice of rule based classification system. Rule based classification is favorable for real time system and a wearable sensors integrated shoe in the author's laboratory is under process for real time implementation and will be reported in future publications. Dataset unbalancing problem in the present study case was addressed by random under-sampling method however some other balancing dataset solution can be adopted in future to improve the accuracy further. Future work will primarily involve the real-time implementation of current approach. The presented work studied the online data available and therefore a real-time study is required to investigate for robust evaluation of the proposed approach. In addition, it will be interesting to see in future that if the methods can be helpful for home-based diagnosis of neurodegenerative disease patients who are at

Table 8 Sensitivity, Specificity and Accuracy (all in %) values of the classifier for the classification of less severe patients

NDD vs. Control			Parkinson's Disease vs. Control			Huntington's Disease vs. Control			ALS Disease vs. Control		
Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
73.3	73.3	73.3	100	71.4	80.0	80.0	80.0	80.0	100	83.3	90.0

Table 9 Comparisons of current approach to previously reported approaches

S.No	AUTHORS	NO. OF MEASURED VARIABLES	FEATURE EXTRACTION METHODS	NO. OF FEATURES	SELECTED FEATURES	CLASSIFIER	DATASET USED	ACCURACY (%)
1.	Wei Zeng, Cong Wang, 2015.	1	Radial basis function neural networks	4	Swing and stance interval of left and right foot.	Deterministic learning theory	Same*	Control vs. NDD = 93.75 Control vs. PD = 87.1 Control vs. HD = 83.33 Control vs. ALS = 89.66
2.	Mohammad Reza Daliri, 2012	7	Genetic algorithm for feature selection	28 (4 from each time series)	Minimum, maximum, average and standard deviation	Support vector machine (SVM)	Same*	Control vs. NDD = 90.63 Control vs. PD = 89.33 Control vs. HD = 90.28 Control vs. ALS = 96.79
3.	Peng Ren et al., 2017	5	Empirical mode decomposition	5 (First five IMFs)	IMFs (Kendall's coefficient of concordance and the ratio of energy change)	Na¨ıve Bayes, support vector machine, random forest, multilayer perceptron, and simple logistic regression	Same*	Average of all classifiers [AUC values] Control vs. PD = 0.901 Control vs. HD = 0.881 Control vs. ALS = 0.898
4.	Peng Ren et al., 2016	5	Phase synchronization and conditional entropy	_	_	Multilayer perceptron, random forest, and Naive Bayes classifier	Same*	[AUC values] Control vs. PD =0.928[MLP], 0.910[RF],0.898[NB] Control vs. HD = 0.910[MLP], 0.959[RF],0.920[NB] Control vs. ALS = 0.824[MLP], 0.789[RF],0.750[NB]
5.	E. Baratin et al., 2015	All gait data	Discrete Wavelet Transform [7 levels]	2 [from all levels]	Coherence and Entropy	SVM (support vector machine)	Same*	Control vs. NDD = 80.4 Control vs. PD = 87.1 Control vs. HD = 86.1 Control vs. ALS = 86.2
6.	Yi Xia,Qingwei Gao, Qiang Ye, 2015	5	Statistical methods	45 (9 from each time series)	Mean, STD, Max, Mini, Skewness, Kurtosis, Lempel-Ziv complexity, Fuzzy entropy, Teager-Kaiser Energy Feature.	Support vector machine (SVM), random forest (RF), multilayer perceptron neural-networks (MLP) and K-nearest neighbor (KNN),	Same*	Control vs. NDD = 96.8 Control vs. PD = 100 Control vs. HD = 100 Control vs. ALS = 96.55 ALS vs. PD = 96.43 ALS vs. HD = 96.88 HD vs. PD = 91.18
7.	Mingjing Yang et al., 2009	10	Maximum signal-to-noise ratio, Maximum signal-to-noise ratio combined with minimum correlation, Maximum prediction power combined with minimum	_	_	SVM (Support vector machine – radial basis function)	Same* (force signals)	Control vs. ALS + PD + HD = 86.85 Control vs. PD = 86.43 Control vs. HD = 84.17 Control vs. ALS =93.96

Table 9 (continued)

S.No	AUTHORS	NO. OF MEASURED VARIABLES	FEATURE EXTRACTION METHODS	NO. OF FEATURES	SELECTED FEATURES	CLASSIFIER	DATASET USED	ACCURACY (%)
			Correlation, Principal component analysis.					
8.	Tomohiro Shirakawa et al. [36]	8	Triaxial acceleration sensor	–	–	Cluster analysis and Principal component analysis	Different	–
9.	Current method	7	Mutual Information between each of the features selected	–	Auto-correlation based features, Data-driven Features, Inter-dependence between dimensions	Decision Tree Classifier	Same*	Control vs. NDD = 87.5 Control vs. PD = 92.3 Control vs. HD = 88.5 Control vs. ALS =96.2

*Refers to the data set given by Physionet website (Gait dynamics in neurodegenerative database, 2017)

early stage. Such an application will be of tremendous use in improving the quality of life among these patients.

5 Conclusion

The present work generated a new set of features from the gait signals, which were more effective in doing classification than a recently published study. Simplistic feature selection was then done, and finally a single decision tree classifier was trained to do classification. This method achieved accuracy of 88.5, 92.3, 96.2, and 87.5 (all in percentage value) while classifying controls from HD, PD, ALS, and NDD respectively. The method was also effective while classifying the less severe patients and thus proposing the impact of the work for meaningful clinical application. The decision tree provided a set of rules for classification, which added interpretability to the classifier. This research work clearly demonstrated the importance of the features generated, which were never used before in the prior studies. These features should be taken into account while doing further studies on this problem.

Acknowledgements We thank Prof. Vinay Goyal (Professor, Department of Neurology, Neuroscience Centre, AIIMS, New Delhi) for guiding us with his valuable suggestions while preparing the manuscript.

Funding This study was not funded by any of the agency.

Compliance with ethical standards

Conflicts of interests All the authors have no conflicts of interests.

Research involving human participants/animals This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Disease Statistics, OHSU Brain Institute. http://www.ohsu.edu/xd/health/services/brain/in-community/brain-awareness/brain-health/disease statistics.cfm Retrieved on November 7, 2017.
2. Gourie Devi M. Epidemiology of neurological disorders in India: review of background, prevalence and incidence of epilepsy, stroke, Parkinson's disease and tremors. Neurol India. 2014;62(6):588–98.
3. Wu Y, Krishnan S. Computer-aided analysis of gait rhythm fluctuations in amyotrophic lateral sclerosis. Med Biol Eng Comput. 2009;47:1165–71.
4. Hausdroff JM, Cudkowicz ME, Firtion R, Wei JY, Goldberger AL. Gait variability and basal ganglia disorders: stride-to-stride variations of gait cycle timing in Parkinson's disease and Huntington's disease. *Mov Disord*. 1998;13(3):428–37.
5. Christofoletti G, McNeely ME, Campbell MC, Duncan RP, Earhart GM. Investigation of factors impacting mobility and gait in

Parkinson disease. *Hum Mov Sci*. 2016;49:308–14. https://doi.org/10.1016/j.humov.2016.08.007.

6. Pyo SJ, Kim H, Kim S, Park Y-M, Kim M-J, et al. Quantitative gait analysis in patients with Huntington's disease. J Move Disorders. 2017;10(3):140–4.
7. Gupta A, Nguyen TB, Chakraborty S, Bourque PR. Accuracy of conventional MRI in ALS. Can J Neurol Sci. 2014;41:53–7.
8. Wahid F, Begg R, Hass CJ, Halgamuge S, Ackland DC. Classification of Parkinson's disease gait using spatial temporal gait features. IEEE J Biomed Health Inform. 2015:2168–94. https://doi.org/10.1109/JBHI.2015.2450232.
9. Wang J, Yuan W, An R. Effectiveness of backward walking training on spatial-temporal gait characteristics: a systematic review and meta-analysis. *Hum Mov Sci*. 2018;60:57–71. https://doi.org/10.1016/j.humov.2018.05.007.
10. Hausdorff JM. ZviLadin, Jeanne Y.Wei. Footswitch system for measurement of the temporal parameters of gait. J Biomech. 1995;28(3):347–51.
11. Barker S, Craik R, Freedman W, Herrmann N, Hillstrom H. Accuracy, reliability, and validity of a spatiotemporal gait analysis system. Med Eng Phys. 2006;28:460–7.
12. Monrraga Bernardino F, Sánchez-DelaCruz E, Ruíz M. Knee-Ankle Sensor for Gait Characterization: Gender Identification Case. Intelligent Computing Systems, Communications in Computer and Information Science, Springer, 2018, 820.
13. Wua Y, Shib L. Analysis of altered gait cycle duration in amyotrophic lateral sclerosis based on nonparametric probability density function estimation. Med Eng Phys. 2011;33:347–55.
14. Zeng W, Wang C. Classification of neurodegenerative diseases using gait dynamics via deterministic learning. Inf Sci. 2015;317:246–58.
15. Daliri MR. Automatic diagnosis of neurodegenerative diseases using gait dynamics. Measurement. 2012;45:1729–34.
16. Wu Y, Krishnan S. Statistical analysis of gait rhythm in patients with Parkinson's disease. IEEE Trans Neu Syst Rehab Eng. 2010;18(2):150–8.
17. W. Van Drongelen. Signal Processing for Neuroscientists: An Introduction to the Analysis of Physiological Signals, Academic Press, 2006.
18. Joshi D, Khajuria A, Joshi P. An automatic non-invasive method for Parkinson's disease classification. Comput Methods Prog Biomed, Elsevier. 2017;145:135–45.
19. Baratin E, Sugavaneswaran L, Umapathy K, Ioana C, Krishnan S. Wavelet-based characterization of gait signal for neurological abnormalities. Gait Posture, Elsevier. 2015;41:634–9.
20. Yang M, Zheng H, Wang H, Mclean S. Feature Selection and Construction for the Discrimination of Neurodegenerative Diseases Based on Gait Analysis. *Pervasive Computing Technologies for Healthcare,* 3rd International Conference IEEE, London, 2009.
21. Ren P, Tang S, Fang F, Luo L, Xu L, Bringas-Vega ML, et al. Gait rhythm fluctuation analysis for neurodegenerative diseases by empirical mode decomposition. IEEE Trans Biomed Eng. 2017;64(1):52–60.
22. Ren P, Zhao W, Zhao Z, Bringas ML, Valdes-Sosa PA, Kendrick KM. Analysis of gait rhythm fluctuations for neurodegenerative diseases by phase synchronization and conditional entropy. IEEE Trans Neural Syst Rehab Eng. 2016;24(2):291–9.
23. Lipton ZC. The Mythos of Model Interpretability. ArXiv e-prints, 2016.
24. Tanner L, Schreiber M, Jenny GH. Low et al. decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. PLoS Negl Trop Dis. 2008;2(3):10.1371/journal.pntd.0000196.
25. Nukala BT, Nakano T, Rodriguez A, et al. Real-time classification of patients with balance disorders vs. Normal subjects using a low-cost small wireless wearable gait sensor. Biosensors. 2016, 6(4):58–80. https://doi.org/10.3390/bios6040058.
26. Tu Y-Q, Shen Y-L. Phase correction autocorrelation-based frequency estimation method for sinusoidal signal. *Sign Proc, Elsevier*. 2017;130:183–9.
27. Zoubek L, Charbonnier S, Lesecq S, Buguet A, Chapotot F. Feature selection for sleep/wake stages classification using data driven methods. *Biomed Sign Proc Control, Elsevier*. 2007;2(3):171–9.
28. Gait dynamics in neurodegenerative database, Physionet. http://www.physionet.org/physiobank/database/gaitndd/, retrieved on 15 September 2017.
29. 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.
30. Kwak SK, Kim JH. Statistical data preparation: management of missing values and outliers. Korean J Anesthesiol. 2017;70(4):407–11.
31. Leys C, Ley C, Klein O, et al. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J Exp Soc Psychol/ Elsevier. 2013;19(4):764–6.
32. Pedregosa, et al. Scikit-learn: machine learning in Python. JMLR. 2011;12:2825–30.
33. K-nearest neighbor's algorithm. https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm. Retrieved on 11 November 2017.
34. Xia Y, Gao Q, Ye Q. Classification of gait rhythm signals between patients with neurodegenerative diseases and normal subjects: experiments with statistical features and different classification models. Biomed Sign Proc Control, Elsevier. 2015;18:254–62.
35. De Laet T, Papageorgiou E, Nieuwenhuys A, Desloovere K. Does expert knowledge improve automatic probabilistic classification of gait joint motion patterns in children with cerebral palsy? PLoS One. 2017;12(6):10.1371/journal.pone.0178378.
36. Shirakawa T, Sugiyama N, Sato H, Sakurai K, Sato E. Gait analysis and machine learning classification on healthy subjects in normal walking. Int J Parallel, Emerg Distrib Syst, Taylor and Francis. 2015;32(2):185–94.